Assessing the benefits and challenges of Clinical Quality Language (CQL) – a HL7 specification for representing logic criteria for clinical quality measures and clinical decision support

Raja A. Cholan
Master's of Science Thesis Proposal
Summer 2018
Department of Medical Informatics and Clinical Epidemiology
Oregon Health & Science University

Abstract: U.S. health care costs continue to rise, while the quality of care delivered is worse than other nations; this has led to several decades of efforts around improving the quality and value of U.S healthcare. Historically, electronic clinical quality measures (eCQMs) have had logic encoded in the Quality Data Model (QDM). Clinical Quality Language (CQL) is a new HL7 standard for eCQM logic-expression replacing the QDM for 2019 eCQMs. Our objective was to assess interpretability of human-readable CQL-representations of eCQMs compared to QDM-representations, using a qualitative approach with a purposive sample of eCQM experts. We found that individuals may have differing preferences for models of logic-expression that delineate eCQM criteria, namely those around readability, temporal relations, logical operators, inclusion of encounter-types, and description of initial patient populations. While there are some potentially beneficial aspects of CQL specifications (e.g., temporal relations, readability, abstraction) the rapid shift from HQMF/QDM to CQL presents a paradigm shift in which further training materials, community forums, and examples may be required around CQL.

Introduction / Motivation

Efforts to redesign and transform the health care system in the United States are at the forefront of political and public discourse. Despite spending a disproportionate amount in current and projected annual health care costs compared to other nations, Americans – 40% of whom are affected by an ongoing, chronic disease – are more likely to die from preventable conditions than residents of other high-income countries. (1, 2) Notorious challenges plague the quality and value of health care delivered in U.S. including: high costs, difficulty accessing primary care, disparities in care because of patients' social determinants of health, and inefficient administration. (3) Potentially exacerbating these challenges is the fact that, historically, our health care system has operated under a fee-for-service model in which each service or unit of care is funded individually. These volume-based payments may influence clinicians to emphasize quantity over quality of services provided, as the volume of services delivered drives the amount of payment reimbursements received. Therefore, there have been a number of recent efforts to enhance the value of health care in U.S., where value is defined as the quality of care provided divided by its cost. Following a long line of previous programs, the Centers for Medicare and Medicaid Services' (CMS) Quality Payment Program (QPP) will pay clinicians based on the quality of care provided to their patients at the population-level beginning in 2019. (4, 5) This model, by definition, requires the ability to quantify the quality of care. Therefore, QPP and other value-based payment programs require hospitals and clinicians to measure and report electronic Clinical Quality Measures (eCQMs).

eCQMs intend to measure healthcare quality by tracking various evidence-based elements of structure, process, and outcomes using clinical data recorded in Electronic Health

Record systems (EHRs). These electronic measures have become a fundamental part of quality payment programs, such as QPP's Merit-based Incentive Payment System (MIPS), and are used to provide performance-based incentive payments. Historically, clinical quality measures have been around for decades, preceding the widespread adoption of EHRs. Early efforts to measure quality centered on manual chart abstractions, but as EHRs were incentivized, the quality measurement ecosystem rapidly evolved to develop standardized, electronic clinical quality measures, or eCQMs. These "official" eCQMs, endorsed by the Centers for Medicare and Medicaid Services (CMS) or the National Quality Forum, have standardized measure logic encoded in the "Health Quality Measure Format" (HQMF), a Health Level Seven International (HL7) representation of eCQM criteria. HQMF CQMs include machine-readable specifications encoded in Extensible Markup Language (XML), as well as human-readable descriptions of the measure logic and components – including pseudocode intended to represent the logic criteria for the measure. The specifications also include definitions for value sets, or structured sets of codes and terms to define clinical concepts of interest. (6) The HQMF for an eCQM (CMS154), assessing the appropriate treatment for children with upper respiratory infection (URI), for example, provides computable and human-readable definitions of the measure's denominator criteria or the subset of patients who the measure is intended to address (e.g., patients aged 3 months – 18 years of age with URI); those who meet the measure's numerator criteria or patients for whom a clinical action or therapy should occur (e.g., patients not prescribed an antibiotic on or 3 days after the URI episode); and those who are excluded from the measure's calculation due to risk of harm by the intended therapy (e.g., patients already taking antibiotics in the 30 days prior to the encounter or patients who had a competing diagnosis for respiratory conditions). The pseudocode in the HQMF (see Figure 1 for an example) provides an algorithm that represents the relationships between patients and clinical concepts using a standardized model from the Quality Data Model (QDM). QDM is an informational model that defines relationships between clinical concepts (e.g., diagnoses, encounters) using standards-based terminologies required in the calculation of an eCQM, and is intended to facilitate standardized quality performance measurement.

Despite the fact that HQMF/QDM is intended to serve as a standard representation for eCQM criteria and logic, questions have arisen around the impact of measure interpretability on the implementation of these measures and, as a result, the accuracy of the resulting

performance.(7) Pragmatically, an eCQM is implemented by manually mapping the right clinical data specified in the HQMF/QDM from the right sources in the EHR to generate the eCQM performance score. Many clinics may not have the skills to interpret the encoded HQMFs, or the technical capacity to implement custom queries for generating eCQM performance data. Therefore, they might increasingly rely on the use of Health Information Technology that is certified to calculate these measures, or a recognized quality measure calculation registry for eCQM reporting. (8, 9) Regardless of the data source used to calculate eCQM data – whether it is a certified EHR, registry system, or custom query – eCQM implementations should reflect the most up-to-date standards and specific criteria encoded in the eCQM's logic defined by the official HQMFs.

Unfortunately, repeated updates to to eCQMs, including their HQMF/QDM specifications, have led to wide discrepancies in version uptake and implementation. For the 71 official eCQMs, regular updates occur at least annually, and sometimes two to three times per year. These updates to eCQM definitions may result in changes to measure logic or to the official sets of included and excluded codes in the standards-based terminologies (i.e. value set vocabularies). In previous work, we found that clinics often lag behind in implementing the most updated, and accurate, versions of official eCQM as outlined by the HQMF files. When older and newer versions of eCQMs were implemented against the same clinical data, we found changes in measurement of quality of up to 5% difference is overall performance score, and up to 28% difference in the number of patients included in a measure's denominator. (10) Similarly, in other work, we showed that implementations of the same eCQM using distinct value set specifications also led to variations in the calculated prevalence of patients at risk for key conditions, and in some cases led to variations in CQM performance percentages.(11)

This variation in eCQM calculation caused by implementation differences is problematic. Because eCQMs intend to promote evidence-based clinical processes, variations in eCQM data caused by having inaccurate or antiquated implementations may impact the ability of clinicians to assess care and improve quality. Jean-Jacques et al. showed that health information technology-supported quality improvement (QI) initiatives can decrease disparities for some chronic disease management and preventive measures;(12) and as technical assistance providers, we observe data-driven QI efforts that rely heavily on patient-level data generated by eCQM reports. If clinicians rely on inaccurate implementations of eCQMs, then they may have lists with

patients intended to be excluded from a measure, and may therefore, target inappropriate patients for therapies, such as recommending aspirin use for someone at high-risk for a fatal bleeding event. Furthermore, their lists will not include the newly added patients who may need certain therapies to improve outcomes. Having accurate eCQM data may translate into potential lives saved, and avoidable harms. Furthermore, the comparability of eCQM performance scores between healthcare organizations is negatively impacted by the vast variation in eCQM implementations. Programs like QPP rely on standardized implementations of measures to be able to benchmark scores effectively, and administer value-based payments accordingly. In the current ecosystem, eCQM data that is collected by the QPP may not be implemented in a standardized way, and therefore the ability to increase value, and enhance population health, may be hindered.

Despite the fact that eCQM logic is maintained by standardized HQMF representations, we currently see vast variation in eCQM implementations, due to the various networks, communication channels, set of transformations, and different organizations of clinical data that currently exist.(13) While some of variability in eCQM implementations can only be addressed through automated (or partially automated) mapping of clinical concepts, improvements to the interpretability of eCQM logic specifications would also serve to eliminate part of the variability. An emerging Health Level Seven (HL7) International standard that might help with electronic processing of eCQM logic is the Clinical Quality Language (CQL), a new specification that focuses on a common model for representing expression logic for CQMs and Clinical Decision Support. According to CMS' eCQI Resource Center, CQL will be used in all HQMFs in the future, will replace the QDM, and is intended to reduce the burden on implementers for consuming measure artifacts.(14) CQL representations of eCQMs will replace the QDM pseudocode historically published in HQMF files; it aims to provide a human-readable, conceptual-level language to define eCQMs and clinical decision support independent of specific data models, such as the QDM or FHIR (Fast Healthcare Interoperability Resource).

CMS is rapidly rolling out the CQL standard in its eCQMs for the 2019 reporting year. While the technical feasibility of translating eCQMs into CQL has been demonstrated, (15) there is a lack of literature evaluating CQL, particularly the interpretability of the human-readable pseudocode components.

Therefore, in this work, our objective is to assess interpretability of human-readable CQL-representations of eCQMs compared to human-readable HQMF/QDM-representations of CQMs. Our primary study question was: what are the benefits and challenges of CQL-representations of clinical quality measure logic?

Background

Historical context for quality measurement

Efforts around improving healthcare quality date back to the 19th century, when U.S. deaths became publicly reported, which sparked social, medical, and epidemiological interest about healthcare gaps, and the unnecessary waste of life. In the early 20th century, the state of medical education became a hot topic, especially after the publication of The Flexner Report, which emphasized the need for science-based foundation of medical training, and adherence to biomedical guidelines. Starting in the 1920s, U.S. business and industry widely adopted quality improvement methodologies aimed to reduce waste. These methodologies trickled down to healthcare formally in 1951 by the formation of The Joint Commission on Accreditation of Hospitals (now The Joint Commission), which required hospitals to report federal quality metrics. In 1966, The Donabedian Model was published, which defined a structure-process-outcomes approach for quality measurement, and was transformative in providing the framework for modern healthcare quality measures. The Healthcare Effectiveness Data and Information Set was developed in the early 1990s to compare the quality of various health plans. The Institute of Medicine published two influential reports that highlighted medical errors, and emphasized quality improvement as a healthcare priority in 1999 and 2000, due to growing concerns about the insufficiency of health care services. (16)Shortly, thereafter, CMS began financially penalizing hospitals that did not report data to The Joint Commission starting in 2005, and the Physician Quality Reporting System (PQRS) was created in 2006. PQRS publically reports eCQM data, and provides a reimbursement bonus at the physician-level. Since 2015, PQRS became no longer voluntary – providers eligible to report are subject to penalties for not participating.(16)PQRS has now been rolled into CMS' current Quality Payment Program which determines clinicians and hospitals reimbursements based on quality performance.

Early efforts in the development of standardized quality measures suffered from a number of challenges. The aforementioned quality initiatives took place when little electronic data was available, and typically required hospitals and clinicians to conduct manual chart

reviews on a sample of patient records. The abstracted data would then be entered into submission tools and reporting forms – a process that often required human intervention and interpretation of the chart abstraction documentation provided by CMS. The measures as defined were also not always supported by evidence directly related to outcomes, and therefore did not provide much benefits on population health.(7) As measures were updated over time to reflect evidence-based guidelines to be more clinical and temporally actionable, this required new data collection and implementation techniques, which led to challenges around consistency, accuracy, and lack of comparability between sites.

Starting in 2009, the Health Information Technology for Economic and Clinical Health (HITECH) Act and Meaningful Use program were operationalized, and outlined plans for the adoption of Electronic Health Records. By 2020, adoption of Electronic Health Record (EHR) systems in the U.S. is expected to approach 100%, due in part to EHR incentive programs created by the HITECH Act and Meaningful Use program(17).(18, 19) This growth of EHRs has provided large clinical datasets for secondary use, increasing the ability to implement eCQMs and increase population health. However, there are commonly understood limitations in our ability to extract and compare data from the high number of disparate EHR systems across the health care system. These include issues of data quality around conformance, completeness, plausibility, verification, and validation – all of which apply to eCQM implementations. (20) To combat these limitations, the Meaningful Use and EHR incentive programs sought to increase the adoption of certified EHR products according to certification criteria outlined by the Office of the National Coordinator for Health Information Technology. (21) Among the certification criteria is the requirement to map data to standardized clinical data standards used in the QDM and HQMF of eCQMs. These clinical data standards consist of structured sets of codes and terms organized in hierarchies to represent clinical concepts – including diagnoses, procedures, medications, administrative data, and laboratory results. Historically, there have been substantial efforts to standardize medical information for clinical data (e.g., SNOMED CT, LOINC, RxNorm). Despite having these standards, there are challenges in the application of controlled medical terminologies to clinical care; representing clinical concepts with codes and terms is complex. Cimino and others have described the challenges of concept orientation, redundancy, completeness, correctness, currency, and granularity, for example, when designing and implementing re-usable medical terminologies.(22) With EHRs being nearly ubiquitous there are

a number efforts by Health IT vendors and health care organizations that aim to effectively manage standards-based terminologies centered on (1) improving measurement, as the quality of care delivered can tell us about population health, and (2) improving accuracy of eCQM implementations and clinical decision support (CDS) tools to act on evidence-based guidelines.

Historical context of pseudocode expression-logic

While there have been advancements in implementing clinical data standards intending to provide rules to allow for the exchange, integration, and management of electronic clinical information, there is still room to improve portability of eCQM queries and CDS implementations across sites. Different health care institutions may increasingly have their clinical data encoded according to standards-based terminologies, but each site will still require human intervention and hand-crafted implementations of eCQMs and computerized CDS including patient safety alerts, and health maintenance reminders intended to improve eCQM performance scores and population health. One implementation of an eCQM at Oregon Health and Science University, for instance, is not completely transferrable to another institution; analysts at each organization will have to modify the underlying query which might be in Structured Query Language (SQL) for instance so that it is computable against their respective clinical database structure. Furthermore, all eCQM concepts are not recorded in the same way in different EHR implementations; what may be structured in one system may be free text in another. While HQMF, QDM, and CQL are the more recent efforts to improve standard representations of eCQM and CDS expression logic, the efforts date back to the mid 1970's when early implementers of computer-based clinical information systems were first recognizing the value of computer-based decision support into their designs.

Notably, Clem McDonald's work on the Regenstrief Medical Record System exemplifies important early work in pseudocode logic expression. McDonald realized that the number of CDS reminders and alerts would quickly increase, so rather than hand-crafting each rule into computer code with programmers, he created one of the first CDS rule languages called CARE (23). The CARE language allowed clinical experts and those without programming expertise to structure if-then logic of alerts using a flexible scripting language that could be interpreted by programmers to implement against the patient record system. As computerized EHRs continued to spread to other academic medical centers in the years that followed, it became clear that a standard way to replicate the expression logic of CARE-style if-then decision rules would be

needed. In the late 1980s, informaticists at Columbia led an important initiative to standardize CDS scripting language and created the Arden Syntax or Medical Logic Modules, the goal of which was to encode if-then-else rules in a standard format that could be computed against different EHR systems, regardless of the location or specific vendor. Arden Syntax logic modules were novel in that they consisted of standard sections called 'categories', and each category contained several 'slots'. For instance, the 'logic' slot contained the actual clinical logic of a rule, and the 'action' slot defined the message that the rule would display to the clinician-user. Modern EHRs often still use this framework even if the full Arden Syntax is not used: when a clinician's workflow reaches a trigger point, then a rule in the system is fired, and evaluates the clinical logic attached to the trigger point. (23)

By the mid 1990's, CDS rules using the Arden Syntax began to spread to numerous commercial systems, however dissemination was limited in that rules written in one facility would not run against any other system. While the 'logic' slot contained machine-executable if-then-else code, there was also a 'curly brace' part of the syntax that only contained a human-readable textual description of the database process and actions necessary for the rule to access clinical data in the EHR. This required human-interpretation and hand-crafting at each specific site and this challenge was referred to as the "curly brace problem". (23) This challenge to achieve portability across environments has persisted throughout the 1990s to 2000s to the current day. These challenges were only worsened when guideline-based techniques were introduced attempting to separate clinical problems into a series of linked clinical decisions. There were some notable efforts such as Guideline Interchange Format (GLIF), PROforma, SAGE, and GEM, which aimed to incorporate a guideline's logic into the executable part of decision logic, however these languages suffered due to a lack of maturity of standards to integrate the guideline engines into EHRs directly.(23)

In 1998, HL7 found the "curly brace" problem to be unsolvable by the Arden approach and began efforts to create expression logic standards based on HL7 Version 3 Reference Information Model (RIM).(24) One attempt was the GELLO Expression Language, which in theory was supposed to access and manipulate clinical data by common clinical entities; however RIM was not proven to be a very practical representation of the complexity of real-world data. Only a small number of vendors were successfully able to implement RIM-based EHRs and therefore the vision of GELLO and HL7 V3 efforts remained unproven. In 2013, HL7 replaced

GELLO with ECA ("Event, Condition, Action"), an expression in XML data structures intended to abstract the representation of expression logic. Shortly thereafter, the standards community realized the benefit in aligning CDS logic expression with those of eCQMs as the goals of CDS rules are often used to prompt clinicians to achieve improved clinical quality outcomes, and therefore HL7 defined QUICK – the Quality Improvement and Clinical Knowledge model. (25) QUICK and ECA have now been wrapped up into CQL which attempts to capture lessons learned from Arden, GELLO, and ECA. The goal of CQL moving forward is to use emerging Application Programming Interfaces like the Fast Healthcare Interoperability Resource (FHIR) as way to allow for more direct access to clinical data that does not require the overhead of RIM mapping. (15) The potential for FHIR and CQL in CDS and eCQM implementations remains to be seen, however, the community is optimistic.

Clinical Quality Language (CQL)

The CQL standard is a draft model by HL7 and is scheduled for Spring 2018 release with implementations in calendar year 2019.(14) Therefore, the majority of background information about CQL centered on HL7/ONC resources, documentation on the eCQI resource center, and github repos containing draft CQL documents. CQL is designed to cover three levels of representations: (1) "a conceptual level" containing human-readable and precise logic; (2) "a logic-level" in which the Expression Logic Model is used for machine-rendering of XML-CQL logic; (3) "a physical level" in which the different implementation environments will be leveraged to execute the ELM and translate from the ELM to the target environment (e.g., SPARL query language, structured query language).(26)

In April 2017, CMS published a PDF describing the current, near-term, and long-term goal of the CQM calculation environment.(27) The slides describe the current calculation architecture, in which CQM implementers typically manually translate HQMF/QDM representations into a target environment (e.g., DBMS). Soon, the CQL standard will replace the HQMF/QDM format. While there will still be manual translation of CQL into CQM implementations in the near-term future, there is a non-trivial goal to transition to an automatic translator to use Expression Logic Model (ELM) which would decrease the burden caused by customization and hand-crafted installations of eCQM calculation engines. Longer-term, the hope is that EHRs will adopt these data model standards to directly consume CQL/ELM files and translate them into CDS and CQM implementations, thereby reducing the burden on

implementers. There may be promise in this vision for the future. Jiang et al. at Mayo(15) recently developed a semantic web-based framework for executing CQL using FHIR and demonstrated technical feasibility by successfully deploying a prototype which translated CQL to SPARL queries by leveraging open-source FHIR-based Resource Description Framework (RDF) transformation tools.

CQL is a pseudocode framework for logic-expression; Smed et al.(28)describe algorithms using pseudocode format following guidelines set by Cormen et al. (2001) and Jensen et al. (1985): (28)

> *"Presenting the algorithms in pseudo-code has also a pedagogic rationale. There are two opposing views on how a new algorithm should be taught: First, the teacher describes the overall behaviour of the algorithm (i.e. its substructures and their relations), which often requires explanations in a natural language. Second, to guide on how to proceed with the implementation, the teacher describes the important details of the algorithm, which calls for a light formalism that can be easily converted to a programming code. The teacher's task is to find a balance between these two approaches. To support both approaches, the pseudo-code formalism with simple data and control abstractions allows the teacher to explain the topics in a natural language when necessary."*

Methods and Materials

Figure 1. CQL-representations vs HQMF/QDM-representations for logic-expression for CMS154 Appropriate Treatment for Children with Upper Respiratory Infection (URI)

a) HQMF/QDM

```
Population criteria

    Initial Patient Population =
        AND: "Patient Characteristic Birthdate: birth date" >= 3 month(s) starts before start of "Measurement Period"
        AND: "Patient Characteristic Birthdate: birth date" <= 18 year(s) starts before start of "Measurement Period"
        AND:
            OR: "Occurrence A of Diagnosis, Active: Upper Respiratory Infection" starts during
                OR: "Encounter, Performed: Office Visit"
                OR: "Encounter, Performed: Emergency Department Visit"
                OR: "Encounter, Performed: Preventive Care– Initial Office Visit, 0 to 17"
                OR: "Encounter, Performed: Preventive Care – Established Office Visit, 0 to 17"
                OR: "Encounter, Performed: Hospital Observation Care – Initial"
                OR: "Encounter, Performed: Face-to-Face Interaction"
                during "Measurement Period"
            OR:
                OR: "Encounter, Performed: Office Visit"
                OR: "Encounter, Performed: Emergency Department Visit"
                OR: "Encounter, Performed: Preventive Care– Initial Office Visit, 0 to 17"
                OR: "Encounter, Performed: Preventive Care – Established Office Visit, 0 to 17"
                OR: "Encounter, Performed: Hospital Observation Care – Initial"
                OR: "Encounter, Performed: Face-to-Face Interaction"
                during ("Occurrence A of Diagnosis, Active: Upper Respiratory Infection" during "Measurement Period")
    Denominator =
        AND: "Initial Patient Population"
    Denominator Exclusions =
        AND:
            OR: "Diagnosis, Active: Competing Conditions for Respiratory Conditions" <= 3 day(s) starts after start of "Occurrence A of Diagnosis, Active: Upper Respiratory Infection"
            OR: "Medication, Active: Antibiotic Medications" <= 30 day(s) starts before or during "Occurrence A of Diagnosis, Active: Upper Respiratory Infection"
    Numerator =
        AND NOT: "Medication, Order: Antibiotic Medications" <= 3 day(s) starts after start of "Occurrence A of Diagnosis, Active: Upper Respiratory Infection"
    Denominator Exceptions =
        None

Data criteria (QDM Data Elements)

    "Diagnosis, Active: Competing Conditions for Respiratory Conditions" using "Competing Conditions for Respiratory Conditions Grouping Value Set (2.16.840.1.113883.3.464.1003.102.12.1017)"
    "Diagnosis, Active: Upper Respiratory Infection" using "Upper Respiratory Infection Grouping Value Set (2.16.840.1.113883.3.464.1003.102.12.1022)"
    "Encounter, Performed: Emergency Department Visit" using "Emergency Department Visit Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1010)"
    "Encounter, Performed: Face-to-Face Interaction" using "Face-to-Face Interaction Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1048)"
    "Encounter, Performed: Hospital Observation Care – Initial" using "Hospital Observation Care – Initial Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1002)"
    "Encounter, Performed: Office Visit" using "Office Visit Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1001)"
    "Encounter, Performed: Preventive Care – Established Office Visit, 0 to 17" using "Preventive Care – Established Office Visit, 0 to 17 Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1024)"
    "Encounter, Performed: Preventive Care– Initial Office Visit, 0 to 17" using "Preventive Care– Initial Office Visit, 0 to 17 Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1022)"
    "Medication, Active: Antibiotic Medications" using "Antibiotic Medications Grouping Value Set (2.16.840.1.113883.3.464.1003.196.12.1001)"
    "Medication, Order: Antibiotic Medications" using "Antibiotic Medications Grouping Value Set (2.16.840.1.113883.3.464.1003.196.12.1001)"
    "Patient Characteristic Birthdate: birth date" using "birth date LOINC Value Set (2.16.840.1.113883.3.560.100.4)"

Reporting Stratification

    None

Supplemental Data Elements

    "Patient Characteristic Ethnicity: Ethnicity" using "Ethnicity CDC Value Set (2.16.840.1.114222.4.11.837)"
    "Patient Characteristic Payer: Payer" using "Payer Source of Payment Typology Value Set (2.16.840.1.114222.4.11.3591)"
    "Patient Characteristic Race: Race" using "Race CDC Value Set (2.16.840.1.114222.4.11.836)"
    "Patient Characteristic Sex: ONC Administrative Sex" using "ONC Administrative Sex Administrative Sex Value Set (2.16.840.1.113762.1.4.1)"
```

b) CQL

```
using QUICK

valueset "Competing Conditions for Respiratory Conditions": '2.16.840.1.113883.3.464.1003.102.12.1017'
valueset "Upper Respiratory Infection": '2.16.840.1.113883.3.464.1003.102.12.1022'
valueset "Emergency Department Visit": '2.16.840.1.113883.3.464.1003.101.12.1010'
valueset "Face-to-Face Interaction": '2.16.840.1.113883.3.464.1003.101.12.1048'
valueset "Hospital Observation Care - Initial": '2.16.840.1.113883.3.464.1003.101.12.1002'
valueset "Office Visit": '2.16.840.1.113883.3.464.1003.101.12.1001'
valueset "Preventive Care - Established Office Visit, 0 to 17": '2.16.840.1.113883.3.464.1003.101.12.1024'
valueset "Preventive Care- Initial Office Visit, 0 to 17": '2.16.840.1.113883.3.464.1003.101.12.1022'
valueset "Antibiotic Medications": '2.16.840.1.113883.3.464.1003.196.12.1001'

parameter MeasurementPeriod default Interval[DateTime(2013, 1, 1, 0, 0, 0, 0), DateTime(2014, 1, 1, 0, 0, 0, 0))

context Patient

define "InDemographic":
    AgeInMonthsAt(start of MeasurementPeriod) >= 3
        and AgeInYearsAt(start of MeasurementPeriod) <= 18

define "ValidEncounters":
        ["Encounter": "Office Visit"]
                union ["Encounter": "Emergency Department Visit"]
                union ["Encounter": "Preventive Care- Initial Office Visit, 0 to 17"]
                union ["Encounter": "Preventive Care - Established Office Visit, 0 to 17"]
                union ["Encounter": "Hospital Observation Care - Initial"]
                union ["Encounter": "Face-to-Face Interaction"]

define "EncountersDuringMeasurementPeriod":
        "ValidEncounters" E where E."period" during MeasurementPeriod

define "URIDiagnosis":
        ["Condition": "Upper Respiratory Infection"]

define "ValidDiagnosis":
        "URIDiagnosis" U
                where exists ("EncountersDuringMeasurementPeriod" E where U."onsetDateTime" during E."period")
                        or exists ("ValidEncounters" E where Interval[U."onsetDateTime", U."abatementDate"] includes E."period")

define "InitialPopulation":
        "ValidDiagnosis" V
                where "InDemographic"

define "Denominator":
        true

define "Numerator":
        "ValidDiagnosis" V where not exists (["MedicationPrescription": "Antibiotic Medications"] M where M."dateWritten" occurs 3 days or less after V."onsetDateTime")

context Population

define "MeasureScore": (Count("Numerator") / Count("Denominator")) * 100
```

Study Design

0. Overview

We used a qualitative design with sample of thirteen eCQM developers, authors, and implementers. Semi-structured interviews based on components from the PRECEDE-PROCEED model(29) were conducted with purposively selected participants to identify key information about each expert's background, experience with eCQMs, and their insights about Predisposing, Reinforcing and Enabling Constructs in Educational Diagnosis and Evaluation (PRECEDE), and Policy, Regulatory, and Organizational Constructs in Educational and Environmental Development (PROCEED) – in the eCQM environment. PRECEDE involves assessing community factors by determining the social problems and needs of a given population, the determinants of an identified problem, as well as the behavioral and environmental determinants that predispose, reinforce, and enable certain behaviors.(29) PROCEED involves the identification of outcomes and implementation by assessing availability for resources, whether certain programs are reaching intended populations, and evaluating behaviors based on incidence of negative/positive behaviors.(29)

1. Measure Selection

We chose CMS 154, an official, standardized eCQM from CMS which has logic representations in both CQL and the HQMF/QDM format (see Figure 1). Experts were asked to review CQL-based and HQMF/QDM-based logic-representations for CMS154v4: 'Appropriate Treatment for Children with Upper Respiratory Infection'. This measure was chosen because it is an official, standardized eCQM, and contains multiple relationships between the clinical concepts, including various encounter-types and complex temporal facets. Six subjects were randomized to be shown the CQL-representation of the measure first and the HQMF/QDM version second, and seven were randomized to be shown the HQMF/QDM-representation first.

2. Identification and Recruitment of experts in CQM logic expression and implementation

Thirteen experts were identified and recruited to participate in semi-structured interviews. Experts were contacted from HL7's Clinical Quality Information working group, CQL's github repository, and developers in CMS' Measure Authoring Tool. The CQM expert population of interest has the following profile:

- Technically trained in software and/or medical informatics
- Experienced with implementing and/or developing CQMs
- Familiar with HQMF/QDM CQM specifications

We used a purposive sampling approach.(27) Given the technical expertise required to provide meaningful insights, purposive sampling was appropriate because we intended not to randomly select units from the general population, but instead wanted participants from the population of interest outlined above. This technically trained group, with specific experience with eCQMs including HQMF/QDM and CQL, was the population of interest who could have answered our research question. We therefore, focused on experts with this particular expertise.

3. Interview Guide/Survey Creation

Interview guides and surveys were created to outline general topic areas to be explored with interview participants based on the PRECEDE-PROCEED model. The questions may have been modified in light of what is learned during the interview and to fit the expertise of the interviewee.

<u>Outline of Interview Guide (See Appendix 1 for full Interview Guide).</u>

i. Learn about the expert's background, role, career etc.
    a. Each expert will self-assess his or her level of familiarity with eCQMs, HQMFs, CQL, and scripting languages.

ii. Ask experts about enabling factors and challenges they face in eCQM implementations, development, and authoring: particularly, around reviewing the CQL-representations of measures compared to QDM/HQMF representations of measures.
iii. Assess the benefits and challenges of each representation. What is understandable and what is confusing?
iv. Ask about environmental, policy, regulatory, and training materials that would help facilitate interpretability of CQL and implementations, and that would enhance more standardized representations of eCQMs across the ecosystem.

There was no pre-existing survey inclusive of these topics. The questions in the semi-structured interview guide were based on review of the literature about eCQMs, CQL, and pseudocode formats, as well as iteratively reviewed with the thesis committee and eCQM experts to determine the final survey questions. See Appendix 1 for full Interview Guide

4. Analysis

We used a multi-step approach to analyze the transcripts from the interviews. First, we developed a codebook and standardized definitions of the codes based on prior work, (10, 11, 14) pseudocode conventions, and on our understanding of specifications associated with delineating logic-expression for eCQMs. The data was coded independently, and reviewed as a group, and themes were identified using an immersion-crystallization approach. Robert Wood Johnson Foundation defines immersion- crystallization: "immersion is a process whereby researchers immerse themselves in the data they've collected by reading or examining some portion of the data in detail. Crystallization is the process of temporarily suspending the process of examining or reading the data (immersion) in order to reflect on the analysis experience and attempt to identify and articulate patterns or themes noticed during the immersion process. These dual processes continue until all the data have been examined and patterns and claims emerge from the data that are meaningful and can be well articulated and substantiated."(30) All qualitative data – interview transcripts – were entered into Altas.ti software(31) for data management and analysis. Conducting the interviews and immersion-crystallization analysis were conducted in parallel: themes around enabling, pre-disposing, reinforcing factors, challenges in the eCQM ecosystem, challenges in delineating specific eCQM logic criteria, as well as environmental and policy considerations were elicited throughout data collection. Saturation was tracked by determining whether new themes were being added to these categoires as the interviews were

being conducted. When saturation was achieved, new experts were no long recruited to participate.

Results:

Participant characteristics

We interviewed thirteen subjects with backgrounds in clinical informatics and/or Health IT, with varying experience working with eCQMs. Nine subjects work at academic medical centers or large health systems, three at software vendors or as software contractors, and one at a coordinated care organization / local health plan. All thirteen participants had experience with implementing eCQMs (ranging from 2-11 years, mean of 6.5 years); three had experience with developing and authoring eCQMs; and, one had experience in evaluating HQMF/QDM specifications. Six participants had experience working with CQL, six had briefly reviewed CQL specifications, and one had not yet delved into the details of CQL, but had heard of it.

Table 1. Interview subject characteristics

Participant characteristics		Counts (n=13) (%)
Employer	Academic Medical Center / Large Health System	9 (69%)
	Software Vendor / Software Contractor	3 (23%)
	Coordinated Care Organization / Local Health Plan	1 (8%)
Experience	Implementing eCQMs	13 (100%)
	Devloping and Authoring eCQMs	3 (23%)
	Evaluating eCQM specifications	1 (8%)
Familiarity with CQL	Experience working with CQL	6 (46%)
	Briefly reviewed CQL	6 (46%)
	Had no experience with CQL	1 (8%)

Enabling factors mentioned by participants included various web-based resources

Despite having a wide range of experience with eCQMs, participants tended to agree on both the resources that act as enabling factors for enhancing eCQM implementations, as well as the challenges that act as barriers to clearly delineating eCQM logic. CMS' eCQI resource center and NLM's Value Set Authority Center (VSAC) were widely referenced as the most helpful centralized repositories of documents, files, and resources related to eCQM specifications and logic. The Bonnie testing tool was also highlighted as particularly helpful for writing measures, and for understanding more complex representations of logic – especially in troubleshooting. Two distinct participants, for example, highlighted the same enabling factors:

> *The VSAC… hosted by the NLM has also been a tremendous resource. Although the value set representations a lot of times are included with the measure bundles, I've found that VSAC is a very helpful tool, graphically appealing, to be able to explore and review the value set definitions*

themselves. .. And more recently, the eCQI [Resource Center] website is a really, really nice centralized location for information about all the different technologies and things that go into ECQMs. So, I was very, very enthusiastic when that was created and have been using that quite a bit, as well.

Really the starting point is always the measure specification. Either they are coming from CMS or some other entity, so that's the starting point always, and that's obviously the first resource we look for. And then, well, it depends on again, on the measure's steward or where it is coming from and then obviously, we'd be going into more details and CMS has a bunch of resources out there, eCQI is another one and off the top of my head, I'm not really sure what else that I have been using. And the Value Set Authority Center, if we need to go that much detail.

Challenges mentioned by participants included temporal relations, encounter-types, baseline population, and granularity of value sets

One global challenge that came up consistently is that all implementations of eCQMs require some level of human intervention to be able to translate eCQM specifications into a query that retrieves the required data elements and outputs the results. Pragmatically, eCQMs require human configurations and manual implementations of the specifications. Therefore, the decisions that implementers make are often made in uncertainty. Some of specific logic criteria that were mentioned as particularly challenging to interpret included: temporal relationships, navigating specific types of encounters, transforming units for calculation, accurately retrieving the initial patient population, and issues related to mapping and granularity of codes included in value sets. For example, one participant said:

"One thing that we really struggled with …is trying to understand what the baseline population is actually supposed to be or how you define the denominator. Because a lot of these measures… they'll just say…within your population, people who've had at least one visit and are over 18. But they don't say, what kind of visit that's limited to. And they don't say how exactly you're defining our population.
We [also] had at least one measure that required that people be on a medication for at least three months consecutively. And there's not a good way for us to tell that because if you look at the earliest date that it was started, it's just historical data and you have no way of knowing like how accurate that is, if the person remained on the medication the entire time after that or not. So, that was an issue."

Similarly, another participant said:

I think it's important to find the initial patient population (IPP_ – well, mostly measures do have all of these categories, like the IPP and then denominator and numerator exclusions and all those items there. But it depends on the specifications and how granular they are. Sometimes they go into the real details of defining what, for instance, an office visit would be but in other cases they will just mention that you know, you just need to look for an office visit, a completed office visit or something. So, we have to interpret if that office visit includes the prenatal visits, or does it also include

telemedicine visits and [inaudible] all those items there. So, an ideal measure specification to me should include the details, the line level, or the item level details and the very granular basic level. It's easier for us to interpret and implement.

Speaking generally about eCQM logic expression one participant described an ideal representation of logic:

So most of the stuff that we have found to be very valuable and I think would be incredibly valuable for making sure that the Electronic Clinical Quality Measures are successful is making the logic less opaque, simple, and traceable for an actual clinician so that a technical person can sit down and go look, here's what we're doing and you can have faith in this because this is right, right? And then showing them, you know, this is what's going on. It's hard to explain because we always try to do things in very discrete finite terms.

CQL vs HQMF/QDM comparison: mixed results about beneficial aspects

The aforementioned global challenges with delineating eCQM logic criteria were highlighted when we showed subjects a HQMF/QDM representation and CQL representation for logic expression of CMS154v4, Appropriate Treatment for Children with Upper Respiratory Infection (see Figure 2). Our qualitative analysis showed mixed results about whether subjects thought that HQMF/QDM or CQL versions were better at addressing the challenges outlined above. Some participants indicated that CQL was a better representation in terms of readability, syntax of expression, and naming conventions, while others thought that HQMF/QDM was better. Similarly, with temporal relationships, interpretability varied with some participants favoring the HQMF/QDM's notation (e.g., <= 3 days) while others preferred CQL's use of natural language (e.g., occurs 3 days or less). Some subjects preferred HQMF/QDM's expression of Boolean logic to represent encounter types, while others favored CQL's syntax. Although not included in Figure 2, value sets were displayed after the logic expression in the HQMF/QDM version, and before the logic expression in the CQL-representation in the actual interviews; there was also discord about which format was better for interpretability.

For example, here's a quote where a subject thought that HQMF/QDM was the better representation:

I guess it's clearer so you kind of flipped it upside down. The numerator and denominator are more towards the top of the screen. At the bottom of the screen, you've defined all the data elements. They have "and" at the top that's really clear. The initial patient population equals, your criteria, although the logic's still not super clear. Denominator, initial patient population, exclusions, you know, a bunch of "or" statements. I mean the two main differences that I see is that the elements of the thing

> *are more spelled out. They're just more wordy so they're more descriptive so I don't have to reference their definitions to know what they are*

And, here's a quote where the respondent thought that CQL is clearer:

> *I think that the flow of the way that the CQL has defined the measure is more in line with how a person who is implementing the measures would want the measure spec to look like. But again, the value sets are kind of like a black box even on here, so you still have to go in there to see what's actually in there. And sometimes you might see errors that you want to have fixed. I thing the pre-processing time of the specification might be less. [pause] And I guess, just basic understanding and development of the workflow of the implementation could also be reduced by this.*

Specifically, in regard to representing encounters/diagnoses one subject indicated that CQL was an improvement compare to HQMF/QDM.

> *If the encounter diagnosis was present for more than two years, I couldn't do that in QDM logic. I would have to say: The diagnosis was present for at least two years before the measurement period and there was an encounter diagnosis that's the same thing. Now, I can just say: The encounter diagnosis started more than two years ago. So, it's a lot cleaner. So, from the authoring side, the changes to QDM helped and the CQL helped tremendously to be able to create those expressions and make it clearer.*

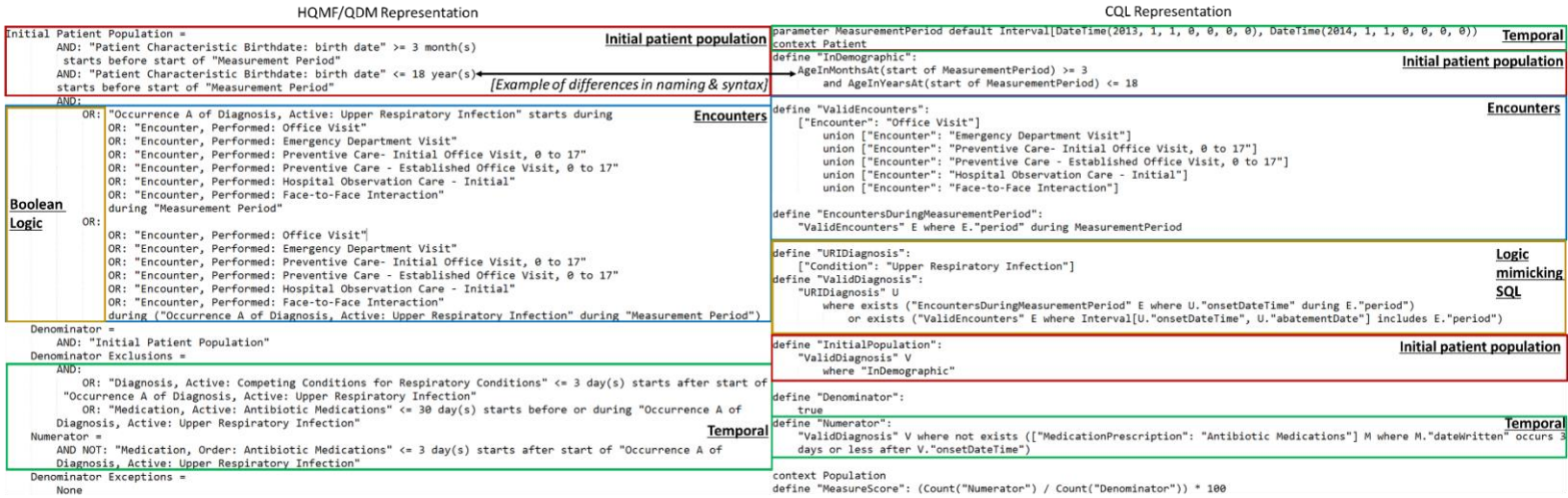
While another participant also thought that CQL was the clearer specification overall, they did not agree that CQL was better for representing encounters/diagnoses:

> *But the thing that was a little bit confusing was the encounters during the measurement period or the valid encounters. Because you know, thinking through that, I actually think that the exists valid encounters where the diagnosis includes the encounter period? That disconnect from the measurement period and if any encounter where that was true, it would also be true for the encounters during measurement period because the valid encounters is a superset of encounters during measurement period. So that – that was the part that was confusing to me but there appears to be redundant criteria here.*

Figure 2, below, shows a side-by-side comparison of the HQMF/QDM logic representation versus the CQL logic representation for CMS 154. The colored boxes intend to show snippets of pseudocode that are equivalent between the two specifications. Despite the mixed results between preferences for HQMF/QDM versus CQL, there were some potentially beneficial aspects of CQL that were elicited in our interviews. As indicated in the quote above, one potential benefit was around temporal relations. In HQMF/QDM the syntax might read as follows: "Greater than three months starts before start of x". In CQL, it would simply read "Starts three months before x". This was interpreted as being a clearer statement, as the phrasing

"starts before start of" is open to more interpretation and can be interpreted as being ambiguous. Furthermore, CQL contains an explicit definition of the measurement period near the top of the specification, whereas HQMF/QDM refers to a measurement period but never defines it within the pseudocode itself.

Figure 2. *CQL vs HQMF/QDM representations for eCQM CMS 154v4: Appropriate Treatment for Children w/ URI*



Another potentially beneficial aspect of CQL was around readability. More subjects tended to note that CQL was simplified and easier to read, with clearer naming expressions, and improved visual aesthetic. Regarding HQMF/QDM one participant said:

> *"So, looking at the indentation, I forgot that one of the things I disliked and personally found confusing until you just got used to it, was the way that they put the ands together and the or indent. To me, it is kind of hard to unwind and reapply that at a first glance. And this is a minor point, but the fact that it starts with And. You know, initial population equals AND that – and that and that – I think that if it started –with: Initial patient population equals patient characteristic birthdates, you know, blah, blah, blah, blah and – It just makes it a little bit more natural language as you're reading through it. And hearkening back to the way that they did the age and months? I think it's a lot cleaner than the way that age is represented as part of the demographics within here (CQL).*

Regarding readability, a participant with eCQM authoring experience said:

> *One of the things that helps the most in terms of readability, was really just naming of expressions used within the quality measures. You know, a significant barrier to understanding any logic is that you know, you're often constrained by the syntax of the formalism, to express things in certain ways, and being able to communicate the intent through those barriers is often quite difficult. And so naming and CQL was designed around this notion of being able to communicate logic, allows you to name your constructs using natural language. And so, we use, you know, the primary identifiers can be quoted identifiers and you can use fairly long descriptive names with spaces and title*

casing to try to clearly indicate the intent. It's almost like a short description of the criteria rather than an identifier for it.

Another promising aspect of CQL is the concept of abstraction, or components that can be defined globally and re-used across measures. One subject noted:

I know that one of the things with CQL is that you can reuse components, or reuse definitions. Not that pediatric age takes a lot of time to rewrite but you can also kind of just set that up once and reuse it. So, even if it is something where the syntax gets a little bit tricky, it's easy to reuse and pull it in again, whereas here, you know, I know that we would have to write this every time for every different measure.

Similarly, another subject said:

From a data management standpoint, having explicit buckets of information you're going to pull from, so you're defining these as reasonable elements means that they're reusable elements. So the logic will be easier and frankly it will be easier to test for conformance each of the individual components, you know. You can say hey, did we get this right or this right or this right for purposes of the testing for one. The part that I really like is it's just easier for somebody to understand and communicate and even from the standpoint of breaking work down and looking at it, like hey, we need you to go find this, what is this.

<u>*Participants suggested additional resources from the eCQM community around CQL*</u>

Regarding eCQM Policy, Regulatory, and Environment (PROCEED), our subjects tended to agree that more documentation, training, and community forums would help in the uptake and interpretability of CQL. Many asked for annotated examples in addition to community linkages to help understand the new components within CQL. For instance, one subject said:

I mean to be fair, you're coming from existing understandings of things so it's also a matter of explaining from your knowledge of HQMF, here's how we are different and better. What the goals are and how it meets those goals and not to diss HQMF because it works great, but this is just a different way of looking at it, how do I think differently about this…then separately honestly, examples, examples, examples, and stepping me through examples so I can understand this because that to me is actually how I learn things. You know, why didn't this work, tear it apart and try to figure it out or why did it work, how do I come at this. I think we understand how to shape my thinking process.

Discussion:

CQL is intended to be an evolutionary step in expression logic for harmonizing standards between clinical decision support and eCQM reporting. There are significant, new features of CQL compared to the HQMF/QDM format, such as the use of 'libraries' or collections of CQL definitions and function statements that can be shared across measures. This is intended to result in greater consistency across measures, as different eCQMs and CDS rules can use the same

logic statements. Ideally, this potential benefit will be realized when the quality measurement calculation technology evolves to a point where CQL can be fully machine-readable and automatically applied to disparate sources of clinical data through the ELM. This ideal is aspirational; pragmatically, the eCQM ecosystem will still involve human intervention and interpretation of human-readable eCQM expression logic in the near term future. Despite not being fully automated, the human-readable updates in CQL compared to HQMF/QDM are still intended to help with the interpretably of delineating eCQM expression logic. In this study, however, we found that there may be variation in the preference of syntax, and in ability to interpret pseudocode representing the measure logic and clinical data elements between CQL vs HQMF/QDM.

A typical eCQM consists of populations (denominator, numerator, etc.) where each population is a combination of the measure's logic and clinical data elements from expressions. In the HQMF/QDM expressions, these expressions are formatted using logical operators as typically used with Boolean logical values. In HQMF/QDM, within a single AND/OR statement, there is a precedence of operators defining events filtered by codes, data type (e.g., active, ordered, resolved), negation rationale, value set criteria, temporal constraints, value restriction, etc. (32) Therefore, the order of these operators is crucial for executing eCQMs accurately, and this has been problematic in eCQM implementations across sites. Despite these challenges, the interpretation of HQMF/QDM logic does not require specific programming knowledge, and therefore clinically trained and non-technical people can decipher the logic and plan for implementations. With CQL, on the other hand, the standards for expression logic intend to mimic a high-level query language. Not only does CQL include Boolean logic operators like HQMF/QDM, but it also includes additional logic operators (e.g., '+', '-', 'and, 'or', 'intersect', 'union'), more structured 'values' (e.g., 5 mmHg), functions (e.g., CalcuateAge()), and identifiers (e.g., 'Inpatient Encounter'). These elements are combined and then labeled with identifiers so they can be used to define a top-level population. Therefore, while not required, having programming experience is helpful in the ability to interpret CQL.

In this study, most subjects had some technical background or experience in implementing measures. It is unclear whether clinical users, or those without some programming/scripting experience will be able to readily use and decipher CQL. Even for users with considerable programming experience, interpretability varied, as there is no annotation in

current CQL examples which limits human-readability. Therefore, one suggestion is to add annotation so that measure developers, authors, and implementers can easily decipher the criteria and elements of CQL. Importantly, hyperlinked CQL examples with clickable links that define key identifiers, functions, and operators would be particularly useful.

Having interpretable eCQM specifications may help decrease the vast variation currently seen in eCQM implementations. Ideally, eCQMs intend to promote evidence-based guidelines and the specifications are intended to reflect 'perfect' implementations. A 'perfect' implementation of an eCQM can be thought of as an implementation that accurately includes all pertinent patients for a therapy or intervention, and accurately excludes those that might be harmed. The numerator, similarly, is implemented 'perfectly' if it accurately reflects evidence-based clinical action or therapy. From an operational perspective, in a 'perfectly' implemented eCQM, the patients *failing* a measure (or those who are included in the denominator, but not included in the numerator) are the patients for whom quality improvement interventions can be prioritized. This list of *numerator-negative* patients is the list of patients where there is a care gap and adverse events, or even death, can be prevented for these patients if interventions are implemented effectively. However, a major barrier to achieving 'perfect' eCQM implementations centers around the lack of universal interpretability of eCQM logic. Therefore, as our prior work has shown, we currently see variation in eCQM implementations, with different institutions implementing vastly different versions of measures, including major discrepancies in the underlying logic, despite the institutions referring to the same HQMF/QDM specifications. This is not only a threat to patient-level quality improvement efforts, but also to broader payment reform, as comparability of quality performance across sites is the crux upon which value-based care can be achieved.

CQL is expected to help this problem of eCQM logic interpretability and therefore the portability of eCQM implementations across sites. Notably, CQL is an effort to solve the 'curly brace problem' from CDS artifacts (outlined above). Our study shows that there may be benefits of CQL that may help alleviate some aspects of the 'curly brace problem'. The beneficial aspects elicited in this study involved improved interpretability around logic readability, abstraction of re-usable, general eCQM logic components, and eCQM temporal relations between clinical data elements. Despite these potential benefits to interpretability of expression-logic, there are prevailing challenges of the 'curly brace problem' that persist even with CQL. At the most basic

level, when considering issues related to mapping a concept to a EHR and its corresponding database table/field, the curly brace problem still exists with CQL.

Our study also showed, however, that there is a major paradigm shift as we move from HQMF/QDM representations to CQL. Some of our subjects have nearly a decade of experience with working with HQMF/QDM. They are coming with an existing understanding of what the various HQMF/QDM components mean and how to decipher them. Therefore, while CQL is intended to help reduce the burden of interpretation, the rapid roll-out of CQL will require explicit documentation and training to help with uptake – particularly, with the re-usable components. While there is promise, there are still aspects of the 'curly brace problem' that seem to still exist, or are perhaps exacerbated, with CQL. While the re-usable 'libraries' are intended to serve as general CQL definitions that can be used across measures, this may propagate errors as expression logic evolves. If one global CQL definition contains an error, it could easily be re-used in several other measures, and therefore a more widespread threat to achieving 'perfect' eCQM implementations.

Moving forwarded, further documentation and examples are needed so that users can understand the new components of CQL. There are some existing resources which the eCQI Resource center and CQL community have already released in this regard. These include the Measure Author's Guide (33)with description of CQL formatting and how to use it. There are also monthly 'Cooking with CQL' webinars often led by the creators/authors of CQL, as well as a 'Developer's Guide within the CQL specification itself which defines many of the specific components of CQL. While these resources are a good start, there may be additional resources and tools required to support the paradigm shift. For instance, the Measure Authoring Tool has many components that are linked with HQMF/QDM and can readily interpret HQMF XML specifications, and many vendors have tools that touch on HQMF/QDM files. With the rapid release of CQL specifications, these tools will need to be updated so that they can interpret CQL.

An opportunity for future research is to determine the impact that programming language (i.e., pseudocode contained in CQL, HQMF/QDM) has on the quality of programming (i.e., SQL, SPARQL queries). As people gradually adopt new CQL specifications for eCQMs, initiatives like the Quality Payment Program can study the "natural experiment" of HQMF/QDM vs. CQL by comparing the impact of CQL on various trends including eCQM submission rates, performance scores, and prevalence of data quality and validation errors (e.g., number of

submissions with extreme performance). Another opportunity for future work involves an experiment that could be deployed in which one set of implementers with varying levels of experience could implement eCQMs using HQMF/QDM specifications, and another set of implementers could write queries for calculating the same eCQMs based on CQL specifications. A good target group for this study might be newly hired people in the Health IT workforce who would not be influenced by working knowledge of HQMF/QDM. This study could help highlight areas for improvement in CQL, and thereby help achieve higher quality expression logic, and hopefully more standard eCQM implementations.

Limitations

This study faced a number of limitations. One factor, is that we only included people with experience with eCQMs, most of which had extensive experience with HQMF/QDM. A more comprehensive analysis could be done in the future to include more novice users, as well as more experts who have specific experience with CQL. Another limiting factor, was the reliance on subjective measures of preference, rather than any kind of objective measure of interpretability. However, because there has not been a formal evaluation of interpretability of CQL, it would have been difficult to choose appropriate measures. Therefore, we chose to use a qualitative approach in the study design so that the results could be more hypothesis-generating and open-ended. To that end, our purposive sample of experts was justified in that we needed a targeted panel of experts to be able to answer our research question. Our results may be able to inform quantitative metrics for future interpretability studies with a broader sample of users. Another limitation is that it can be hard to separate out the influence of the interpretability of CQL vs. HQMF/QDM from the quality of the authoring. There are reasons to think authoring for both CQL and HQMF/QDM used in this study for CMS 154 were of high quality: 1) The HQMF/QDM specification for this measure has been around for a number of years, and newer and older versions have stayed largely consistent indicating that it did not need major revision, and therefore of good quality; 2) HL7's Clinical Quality Information working group did iteratively review CQL examples and conducted informal comparisons to HQMF/QDM to ensure that measure logic was accurately translated to CQL, therefore we were confident of the CQL representation of CMS 154. But, ultimately the potential discrepancy in authoring quality between CQL and HQMF/QDM syntax is a confounding factor that must be considered.

Conclusions:

While our subjects agreed about specific elements of eCQM logic criteria that are challenging to interpret, our qualitative analysis showed that individuals may have differing preferences for the specific models of logic-expression to resolve these challenges. While CQL is intended to be an improvement, there still may be prevailing challenges in delineating eCQM logic criteria. As a result, implementers may continue to make local decisions based on their intuition and clinical knowledge. Further training materials, community forums, and examples may be required to help in the new paradigm shift as the eCQM community moves from HQMF/QDM representations of logic-expression to CQL-representations.

References

1. About Chronic Diseases [Internet]. Putting Patients First. National Health Council; 2014. Available from: http://www.nationalhealthcouncil.org/sites/default/files/NHC_Files/Pdf_Files/AboutChronicDisease.pdf.
2. About Chronic Disease [Internet]. Centers for Disease Control and Prevention; 2018 [cited 2018Aug23]. Available from: https://www.cdc.gov/chronicdisease/about/index.htm.
3. Schneider E, Squires D. From Last to First — Could the U.S. Health Care System Become the Best in the World? [Internet]. Commonwealth Fund; 2017 [cited 2018Aug23]. Available from: https://www.commonwealthfund.org/publications/journal-article/2017/jul/last-first-could-us-health-care-system-become-best-world?redirect_source=/publications/in-brief/2017/jul/last-to-first-could-us-health-system-become-best-in-world.
4. Centers for Medicare and Medicaid Services, MACRA, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-MIPS-and-APMs.html>; last accessed November 23, 2016.
5. Martin J. 5Stepsto Prepare for MACRA Now. Texas medicine. 2016;112(10):39–46. [PubMed].
6. Rhodes B, Bradley J. Getting Started with Clinical Quality Language: Technical Implementation for Vendors [Internet]. The MITRE Corporation; 2018. Available from: https://health.mitre.org/wp-content/uploads/2018/03/HIMSS18_Getting-Started-with-CQL_Technical-implementation-for-vendors.pdf.
7. Dorr DA, Cohen AM, Williams MP, Hurdle J. From simply inaccurate to complex and inaccurate: complexity in standards-based quality measures. AMIA Annual Symposium proceedings AMIA Symposium 2011;2011:331–8.
8. 2015 Edition Health Information Technology (Health IT) Certification Criteria, 2015 Edition Base Electronic Health Record (EHR) Definition, and ONC Health IT Certification Program Modifications. Final rule. Federal register. 2015;80(200):62601–759. [PubMed].
9. PRIME Registry. The American Board of Family Medicine, Inc. [updated February 2017]. Available from: http://primenavigator.org/.
10. Cholan RA, Weiskopf NG, Rhoton DL, Colin NV, Ross RL, Marzullo MN, et al. Specifications of Clinical Quality Measures and Value Set Vocabularies Shift Over Time: A

Study of Change through Implementation Differences. AMIA Annu Symp Proc. 2017;2017:575-84.
11. Cholan RA, Weiskopf NG, Rhoton D, Sachdeva B, Colin NV, Martin SJ, et al. From Concepts and Codes to Healthcare Quality Measurement: Understanding Variations in Value Set Vocabularies for a Statin Therapy Clinical Quality Measure. EGEMS (Wash DC). 2017;5(1):19.
12. Jean-Jacques M, Persell SD, Thompson JA, Hasnain-Wynia R, Baker DW. Changes in disparities following the implementation of a health information technology-supported quality improvement initiative. Journal of general internal medicine. 2012;27(1):71–7. .
13. HL7 Cross-Paradigm Specification. Clinical Quality Language, Release 1. Health Level Seven International. [updated 2017]. Available from:http://www.hl7.org/implement/standards/product_brief.cfm?product_id=400.
14. Colin NV, Cholan RA, Sachdeva B, Nealy BE, Parchman ML, Dorr DA. Understanding the Impact of Variations in Measurement Period Reporting for Electronic Clinical Quality Measures. EGEMS (Wash DC). 2018;6(1):17.
15. Jiang G, Prud'Hommeax E, Solbrig H. Developing A Semantic Web-based Framework for Executing the Clinical Quality Language Using FHIR. Proceedings of SWAT4LS Workshop. Avalailable from: www.swat4ls.org/wp-content/uploads/2017/11/SWAT4LS-2017_paper_40.pdf.
16. Chassin MR, O'Kane ME. History of the Quality Improvement Movement [Internet]. American Academy of Pediatrics. March of Dimes; 2010 [cited 2016Apr20]. Retrieved from: https://www2.aap.org/sections/perinatal/pdf/1quality.pdf
17. Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2015 [Internet]. dashboard.healthit.gov. [cited 2018Aug23]. Available from: https://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php.
18. Wright A, Henkin S, Feblowitz J, McCoy AB, Bates DW, Sittig DF. Early results of the meaningful use program for electronic health records. N Engl J Med. 2013;368(8):779-80.
19. Mennemeyer ST, Menachemi N, Rahurkar S, Ford EW. Impact of the HITECH Act on physicians' adoption of electronic health records. J Am Med Inform Assoc. 2016;23(2):375-9.
20. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Wash DC). 2016;4(1):1244.
21. Certified Health IT Product List (CHPL). [cited 2018Aug28]. Available from: https://chpl.healthit.gov/.
22. Cimino J. Desiderata for Controlled Medical Vocabularies in the Twenty-First Century. Methods of information in medicine. 1998;37(4-5):394-403.
23. McCallie DP. Healthcare information management systems: cases, strategies, and solutions. Cham: Springer; 2016.
24. HL7 Version 3: Reference Information Model (RIM) [Internet]. Introduction to HL7 Standards. [cited 2018Aug28]. Available from: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=77.
25. QUICK Data Model [Internet]. QUICK Data Model. HL7; [cited 2018Aug28]. Available from: https://www.hl7.org/FHIR/2015May/quick/help.html.
26. Hemler JR, Hall JD, Cholan RA, Crabtree BF, Damschroder LJ, Solberg LI, et al. Practice Facilitator Strategies for Addressing Electronic Health Record Data Challenges for Quality Improvement: EvidenceNOW. J Am Board Fam Med. 2018;31(3):398-409.

27. Purposive sampling [Internet]. Lærd Dissertation; [cited 2018Sep16]. Available from: http://dissertation.laerd.com/purposive-sampling.php - maximum-variation-sampling.
28. Smed J, Hakonen H. Algorithms and Networking for Computer Games, Secound Edition [Internet]. Wiley/Blackwell (10.1111); [cited 2018Aug23]. Available from: http://onlinelibrary.wiley.com/book/10.1002/9781119259770.
29. PRECEDE-PROCEED Model - Rural Health Promotion and Disease Prevention Toolkit [Internet]. Defining Health Promotion and Disease Prevention - RHIhub Toolkit. [cited 2018Aug28]. Available from: https://www.ruralhealthinfo.org/toolkits/health-promotion/2/program-models/precede-proceed.
30. Immersion/Crystallization [Internet]. RWJF - Qualitative Research Guidelines Project | Interviewing | Interviewing. Robert Wood Johnson Foundation; [cited 2018Sep16]. Available from: http://www.qualres.org/HomeImme-3829.html.
31. The Qualitative Data Analysis & Research Software [Internet]. atlas.ti. [cited 2018Sep16]. Available from: https://atlasti.com/.
32. Electronic Clinical Quality Measure Logic and Implementation Guidance and Technical Release Notes. Centers for Medicare & Medicaid Services Office of the National Coordinator for Health Information Technology; 2015.
33. CQL - Clinical Quality Language [Internet]. eCQI Resource Center. CMS; [cited 2018Sep10]. Available from: https://ecqi.healthit.gov/cql-clinical-quality-language.

Appendix 1.

Introductory questions

-Can you tell me a little about yourself and your role?

-Can you tell me about the organization you work for and what it does?

-For how long have you worked with your organization

Background with CQMs

-Could you describe the level of experience you have with implementing CQMs?

Probe: How long have you been implementing and/or developing CQMs? What is your experience in reviewing measure logic and specifications?

-What resources do you think made implementation easier/effective?

-Probe: What would you identify as ideal details of specifications of CQMs? Why do you feel that these details are desirable?

-Probe: What are some potential barriers that you feel present a challenge during interpreting CQM specifications and implementing measure? How do you address these?

-Probe: What solutions have you employed when CQM specifications have any limitations to clearly delineating logic and/or criteria?

-Probe" Could you describe for us the successful strategies you or others have used for successful interpreting CQM logic and implementing them?

CQL

-Have you reviewed CQL? If so, how familiar would you say you are

Demo HQMF pseudocode and CQL pseudocode Experts will be shown an example measure with logic in CQL vs. HQMF format

- What is understandable and what is confusing?
- What is ambiguous?
- Are there specific data elements of CQM logic (eg temporal elements, value sets, lookback period etc) that are easier/harder to interpret in CQL?

Ecosystem

-Overall, how do you think implementation of CQL could work for improving CQM implementations?

-Going forward, what 3 things do you need in order to continue to effectively interpret CQL and CQM logic expression?

-What advice or input would you like to share with the CQL community about what has worked well and what could be done differently in the interpretability of CQM logic?

-What are the advantages of your strategies for implementing CQMs on EHR data?

-What lessons have you learned about CQM pseudocode and specifications that you would want to share with other implementers?

-What types of standards, policies, or industry changes do you think are needed to help CQM specifications achieve standard representations of logic?

And finally, we'd like to ask you:

-Are there any questions we did not ask that you think we should have asked?

-Do you have any questions for us?That's all the questions we have for you today. Thank you for your time and for sharing your insights on these topics.