

Using Clinical Data Standards to Measure Quality: A New Approach

John D. D'Amore^{1,2} Chun Li¹ Laura McCrary³ Jonathan M. Niloff¹ Dean F. Sittig⁴
Allison B. McCoy⁵ Adam Wright⁶

¹Diameter Health, Inc., Farmington, Connecticut, United States

²Boston University Metropolitan College, Boston University, Boston, Massachusetts, United States

³Kansas Health Information Network, Topeka, Kansas, United States

⁴School of Biomedical Informatics, University of Texas-Memorial Hermann Center for Healthcare Quality and Safety, University of Texas Health Science Center, Houston, Texas, United States

⁵Department of Global Biostatistics and Data Science, Tulane University School of Public Health and Tropical Medicine, New Orleans, Louisiana, United States

⁶Division of General Internal Medicine, Brigham and Women's Hospital, Boston, Massachusetts, United States

Address for correspondence John D. D'Amore, MS, Boston University Metropolitan College, Boston University, Boston, MA 02481, United States (e-mail: jdamore@diameterhealth.com).

Appl Clin Inform 2018;9:422–431.

Abstract

Keywords

- ▶ Medicare Access and Chip Reauthorization Act
- ▶ quality
- ▶ standards adoption
- ▶ Continuity of Care Document
- ▶ clinical data management
- ▶ Merit-Based Incentive Payment System and Alternative Payment Models
- ▶ health care reform

Background Value-based payment for care requires the consistent, objective calculation of care quality. Previous initiatives to calculate ambulatory quality measures have relied on billing data or individual electronic health records (EHRs) to calculate and report performance. New methods for quality measure calculation promoted by federal regulations allow qualified clinical data registries to report quality outcomes based on data aggregated across facilities and EHRs using interoperability standards.

Objective This research evaluates the use of clinical document interchange standards as the basis for quality measurement.

Methods Using data on 1,100 patients from 11 ambulatory care facilities and 5 different EHRs, challenges to quality measurement are identified and addressed for 17 certified quality measures.

Results Iterative solutions were identified for 14 measures that improved patient inclusion and measure calculation accuracy. Findings validate this approach to improving measure accuracy while maintaining measure certification.

Conclusion Organizations that report care quality should be aware of how identified issues affect quality measure selection and calculation. Quality measure authors should consider increasing real-world validation and the consistency of measure logic in respect to issues identified in this research.

Background and Significance

The U.S. federal government's goal is to have 90% of its health care payments based on care quality by 2018.¹ In addition, private payers have increasingly incorporated quality outcomes in their contracts.² The transition from fee-for-service

to value-based-payment relies on accurate and reliable methods to measure the quality of care delivered. Many programs have advanced this capability; all of which require objective data and measure definitions.

The longest established program for quality measurement in the United States is the Health Effectiveness Data and

received
January 12, 2018
accepted after revision
April 17, 2018

DOI <https://doi.org/10.1055/s-0038-1656548>.
ISSN 1869-0327.

Copyright © 2018 Schattauer

License terms



Information Set (HEDIS) program managed by the National Committee for Quality Assurance (NCQA). This program began in 1991 and is currently used by over 90% of health plans.³ HEDIS has historically used longitudinal information, primarily electronic billing data from multiple providers, to calculate care quality. This program has shown progress in improving quality outcomes.^{4,5} However, it is challenging to use measures calculated from payer administrative data for ambulatory care improvement due to reporting latency, insufficient clinical specificity, payer patient market share, and inadequate risk adjustment.⁶

A more recent national initiative directly focused on ambulatory care improvement is the Physician Quality Reporting System. Started in 2006, this program provided a voluntary reporting bonus. It reached over 600,000 physicians participating in Medicare but relied on methods developed before widespread electronic health record (EHR) adoption.⁷ To accelerate EHR adoption with a goal of improving care quality, the Meaningful Use incentive program was launched in 2010 by the Centers for Medicare and Medicaid Services (CMS). Only 11% of physicians had a basic EHR at that time.⁸ The Meaningful Use program brought widespread EHR adoption with over 78% of ambulatory clinicians using certified EHRs by the end of 2015.⁹ Part of the Meaningful Use program required the calculation and reporting of at least six quality measures. Incentives were paid for reporting but were not tied to performance. Quality calculations for reporting in this program used information available in EHRs; challenges have been noted in this approach.^{10–12} Unlike HEDIS, EHRs often calculate measure compliance using only data documented within that EHR, in part due to lack of health information exchange and interoperability challenges.^{13,14}

The Merit-Based Incentive Payment System, enacted as part of the Medicare Access and CHIP Reauthorization Act (MACRA), succeeded Meaningful Use for ambulatory clinical quality reporting. Beginning in 2017, based on quality performance, high performing clinicians are paid more than lower performing ones.¹⁵ This program also introduces an alternative method of quality reporting, qualified clinical data registries (QCDRs). QCDRs are third-party organizations that accumulate clinical data from various providers for quality measurement. Since QCDRs can collect data on the same patient from different organizations, including those using different EHRs, they can provide a longitudinal approach to performance measurement like HEDIS. This requires the use of interoperability standards to aggregate the data from different EHRs.

The primary standards that support clinical data exchange today from EHRs are Health Level 7 (HL7) messaging and the Consolidated Clinical Document Architecture (C-CDA). Previous research has demonstrated that clinical documents, such as the C-CDA, provide many of the necessary data elements for quality measure calculation.^{16,17} Research is lacking, however, on the implementation of quality measurement by QCDRs, particularly those integrated with health information exchanges. In addition, studies have called into question the validity and reliability of quality

measures calculated by EHR reporting systems. This is due to challenges in data completeness, accuracy, appropriate codification, gaps between structured fields and available free-text, as well as inconsistency of measure logic implementation.^{18–20} Examination of clinical data from multiple EHRs provides an opportunity to explore how data transformation may improve quality measure calculation while recognizing these concerns. Furthermore, quality measure definitions for HEDIS and other reporting programs are specified using the Health Quality Measure Format and Quality Data Model (QDM). These specifications expect Quality Reporting Document Architecture (QRDA) documents as the clinical data format while this research explores the applicability of C-CDA documents to quality measurement.

Objective

The purpose of quality measurement is to evaluate the care quality delivered to the patient. This research seeks to detail and address challenges that affect the use of interoperability standards to achieve this intent of quality measurement by a QCDR. The Doctors Quality Reporting Network, offered as part of the Kansas Health Information Network (KHIN), was approved as a QCDR in 2017 by CMS and is the locus for this research. Through its use of data in KHIN, its potential reach extends to nearly 10,000 providers and over 5 million patients. The quality measures selected for evaluation included 17 electronic clinical quality measures adjudicated using technology certified by the NCQA.

Methods

We sampled the KHIN data from 11 ambulatory care sites during the 1-year period from July 1, 2016 to June 30, 2017. Sites were selected based on size (> 300 visits per month), continuous submission of clinical documents to KHIN, and independence from an acute care institution since all the quality measures in this study relate to ambulatory care. Selected facilities were not contacted in advance, so the data sample represents a sample of data regularly used in health information exchange. Patient data use in this research was approved by the UTHealth Committee for the Protection of Human Subjects.

One hundred unique patients were randomly selected from each facility; the same patient was never selected from more than one facility. Data from a single clinical document during the time frame was used for quality measurement. Documents included a wide range of clinical data, including patient diagnoses, immunizations, medications, laboratory results, problems, procedures, and vital signs. These clinical domains are required by Meaningful Use as part of Continuity of Care Documents. Multiple EHRs were represented, including Allscripts (Chicago, Illinois, United States), Computer Programs and Systems, Inc. (Mobile, Alabama, United States), eClinicalWorks (Westborough, Massachusetts, United States), General Electric (Chicago, Illinois, United States), and Greenway Medical (Carrollton, Georgia, United States). The data were processed by Diameter Health's (Farmington,

Connecticut, United States) Fusion and Quality modules (version 3.5.0), technology certified by NCQA for electronic clinical quality measurement.²¹ This software includes both transformation logic associated with clinical data and measure logic necessary to calculate and report quality performance. An example of how quality measure compliance may be calculated in the software application for a fictional patient not derived from any real patient information is shown in **Fig. 1**.

Twenty-four measures were available in the certified software, although 7 were excluded from this study. Five measures were excluded since they require data on multiple care encounters, which may not be accurately represented in a randomly selected clinical document (e.g., multivisit initiation and maintenance for drug dependence therapy). One was excluded due to the lack of behavioral assessment data in the sample and one was excluded since it had been discontinued for use by CMS. The 17 examined measures constituted a broad range of process and outcomes measures across diseases and preventative care as shown in **Table 1**. Each measure's logic was specified according to the QDM and was eligible for use in CMS quality reporting programs.²²

The quality measures were first calculated using clinical data without any transformation logic. Since clinical documents generally available to KHIN were used in this study, the software-aligned clinical data from these extracts to quality measure criteria as specified in the QDM. The measures were then recalculated using an iterative approach where techniques were added to improve adherence to national standards, such as terminology and free-text map-

pings. This included techniques to deal with data heterogeneity in clinical document as detailed in prior research.¹⁴ Quantitative metrics on clinical encounters, problems, medications, laboratory results, and vital signs were analyzed for the 1,100 patients and illustrative issues affecting quality measurement were recorded. Changes in measure calculation were then extensively tested against test cases made available by NCQA to determine if certification was affected by the iterative improvement. Population counts of both denominators and numerators were captured both before and after the iterative improvement process.

Results

Depth of Clinical Data

All 1,100 selected clinical documents were loaded into the quality measurement software without error. Of the facilities selected, 4 (36%) submitted Healthcare Information Technology Standards Panel C-32 Continuity of Care Documents and 7 (64%) submitted HL7 C-CDA 1.1 Continuity of Care Documents. Patient age ranged from 0 to 99 at the beginning of the measure period. A total of 589 (53.5%) of patients were female and 510 (46.4%) were male with one patient not having gender recorded as male or female.

Content extracted from the clinical documents included 12,308 clinical encounters, 3,678 immunizations, 20,723 medications, 25,921 problems, 17,959 procedures, 45,704 diagnostic results, and 32,944 vital sign observations. All 11 sites produced clinical documents with information in the domains of patient medications, problems, procedures, results, and vital signs. The majority of clinical encounters

Fictional Patient
400 Nowhere Way, Noplace, NA
(999) 999-9999
Male 52 Years-Old (Jan 9th, 1965)
White
MRN:

OUTPATIENT VISIT
Attributed Clinician: Harry Winston MD
Start Date: 01/19/2017
Depart Date: 01/19/2017

Compliant Measures (3) Non-compliant Measures (3) All Measures (17)

Hemoglobin A1c Poor Control

Eye Exam

Medical Attention for Nephropathy

GENERATE QRDA CATEGORY 1 REPORT FOR THE YEAR: 2018

Smoking Gun for Hemoglobin A1c Poor Control

Encounter	Encounter Date	Code
Office/outpatient visit, new, level 1	01/19/2017	99201 CPT

Problem	Category	Onset Date
Type 2 diabetes mellitus with other diabetic kidney complication	Diabetes mellitus with complications	01/26/2009

Test Performed	Test Value	Normality
4548-4 HbA1c	8.5 %	High

Fig. 1 Quality measure presentation in software application. Quality calculation shown for a fictional patient for calculated measures, with clinical detail shown for a specific measure. Note 1: Tabs along the top show three eligible measures with compliance and three eligible measures with noncompliance. Note 2: The button labeled “Smoking Gun” provides specific clinical detail that substantiates measure eligibility and compliance calculation. Note 3: The clinical detail of the eligible encounter, diagnosis and laboratory result that supports compliance for the selected measure (cms122v5 Diabetic HbA1c < 9%). Copyright and reprinted with permission of Diameter Health, Inc.

Table 1 Quality measures selected in this research

CMS identifier	Measure description	Measure type (reason)	Measure steward
74v6	Primary caries prevention	Process (preventative)	CMS
82v4	Maternal depression screening	Process (preventative)	NCQA
122v5	Diabetes: Poor HbA1c control	Outcome (disease control)	NCQA
123v5	Diabetes: Annual foot exam	Process (preventative)	NCQA
124v5	Cervical cancer screening	Process (preventative)	NCQA
125v5	Breast cancer screening	Process (preventative)	NCQA
127v5	Pneumonia vaccination of older adults	Process (Preventative)	NCQA
130v5	Colorectal cancer screening	Process (preventative)	NCQA
131v5	Diabetes: Annual eye exam	Process (preventative)	NCQA
134v5	Diabetes: Attention for nephropathy	Outcome (disease control)	NCQA
146v5	Appropriate testing for children with pharyngitis	Process (utilization)	NCQA
153v5	Chlamydia screening for women	Process (preventative)	NCQA
154v5	Appropriate treatment for children with upper respiratory infection	Outcome (utilization)	NCQA
155v5	Pediatric weight assessment	Process (preventative)	NCQA
156v5	High risk medication use in elderly	Outcome (patient safety)	NCQA
165v5	Controlling high blood pressure	Outcome (disease control)	NCQA
166v6	Use of imaging studies for back pain	Outcome (utilization)	NCQA

Abbreviations: CMS, Centers for Medicare and Medicaid Services; NCQA, National Committee for Quality Assurance.

represented annual wellness visits and typical evaluation and management distributions for ambulatory encounters. For nine of the sites, data included information for prior clinical visits dating back months and years prior. Historical data are important for several of the quality measures that examine prior clinical information (e.g., past colonoscopies for colon cancer screening). For two of the sites, the clinical documents were more limited, sending data primarily related to the most recent clinical encounter.

Nonnormalized Measure Calculation and Focus Areas for Improvement

Using the clinical data without any transformation, quality measures were calculated using certified technology for a 12-month period from July 2016 to June 2017. Results for individual patients was collected using standard reporting formats of the software and presented in ▶Table 2 (the "Calculation before Iterative Improvement" column).

Of the 17 measures, most measures showed unexpectedly low proportions of eligible patients (i.e., denominators) both relative to disease prevalence and patient demographics. For example, a recent report identified 9.7% of adults in Kansas as having diabetes, but only 1.8% of the 1,100 patients qualified for the diabetes measures examined.²³ Consequently, one area for examination and iterative improvement was to increase the number of eligible patients ("Iterative Improvements for Patient Inclusion").

Of the 15 measures with at least 1 eligible patient, 9 showed no clinical events associated with the measure numerator, resulting in either 0 or 100% compliance. These

rates called into question the validity of the calculation. Consequently, a second area for iterative improvement was to examine if data transformations would improve the accuracy of compliance rates ("Iterative Improvements for Quality Measure Compliance").

Iterative Improvements for Patient Inclusion

Eligible Population Improvement for Encounters. Each of the 17 quality measures as defined by the measure steward requires a face-to-face encounter or office visit in the measurement period for the patient to be eligible for quality measure calculation. Since our information drew directly from interoperable documents from EHRs, the codes used in encounter documentation often lacked this specificity. An example is shown in ▶Fig. 2, where no specific code is shown in the yellow highlighted XML, although the human-readable text provides the context of the visit.

Using automated mapping available in the software, the reference between the human-readable narrative and machine-readable content were used to assign a code for this encounter based on the text of "Office Visit." The software uses a simple text-matching algorithm using exact keywords in the text (e.g., "Office Visit," "Hospitalization," "ER Visit") to assign an appropriate code when not appropriately codified in the machine-readable portion. The code selected was "308335008 (Patient encounter)" from the Systemized Nomenclature of Medicine (SNOMED) which qualified this patient encounter for quality calculation. Analogous encounter normalization techniques were performed on all 1,100 patients.

Table 2 Quality measure calculation before and after iterative improvement

CMS identifier	Measure description	Calculation before iterative improvement		Calculation after iterative improvement	
		Denominator	Compliance	Denominator (% change)	Compliance (absolute change)
74v6	Primary caries prevention	107	4.7%	164 (+53%)	3.0% (−1.7%)
122v5	Diabetes: Poor HbA1c control	20	45.0%	78 (+290%)	37.2% (−7.8%)
123v5	Diabetes: Annual foot exam	20	0.0%	78 (+290%)	0.0% (NA)
124v5	Cervical cancer screening	88	0.0%	182 (+107%)	7.1% (+7.1%)
125v5	Breast cancer screening	64	0.0%	120 (+88%)	9.2% (+9.2%)
127v5	Pneumonia vaccination of older adults	113	55.8%	204 (+81%)	55.9% (+0.1%)
130v5	Colorectal cancer screening	117	1.7%	237 (103%)	14.3% (+12.6%)
131v5	Diabetes: Annual eye exam	20	0.0%	78 (+290%)	0.0% (NA)
134v5	Diabetes: Attention for nephropathy	20	35.0%	78 (+290%)	69.2% (+34.2%)
146v5	Appropriate testing for children with pharyngitis	0	NA	50 (NA)	9.1% (NA)
153v5	Chlamydia screening for women	0	NA	5 (NA)	20.0% (NA)
155v5 Rate 1	Pediatric weight assessment: BMI percentile	81	0.0%	123 (+52%)	22.0% (+22%)
155v5 Rate 2	Pediatric weight assessment: Nutrition counseling		0.0%		0.0% (NA)
155v5 Rate 3	Pediatric weight assessment: Activity counseling		0.0%		0.0% (NA)
156v5 Rate 1	High risk medication use in elderly: 1 medication	109	100%	196 (+80%)	98.5% (−1.5%)
156v5 Rate 2	High risk medication use in elderly: 2 or more medications		100%		100% (NA)
165v5	Controlling high blood pressure	44	34.1%	190 (+332%)	36.4% (+2.3%)
Measures not included in iterative improvement					
82v4	Maternal depression screening	1	0.0%	Not available	
154v5	Appropriate treatment for children with upper respiratory infection	44	100% (73% excluded)	Not available	
166v6	Use of imaging studies for back pain	2	Not available (100% excluded)	Not available	

Abbreviations: BMI, body mass index; CMS, Centers for Medicare and Medicaid Services; NA, not available.

Eligible Population Improvement for Problem Inclusion. Several of the quality measures require patients to have a specific diagnosis before a specific date for inclusion in the quality measure. For example, for inclusion in the diabetes measures, a patient must have an eligible SNOMED, International Classification of Diseases (ICD)-9, or ICD-10 code on or before the measure period. Real-world documentation of onset dates, however, is often lacking in EHRs. This may be due either to the information not being known or from clinicians skipping over fields when documenting in the EHR.

Nine measures selected for this sample require a specific problem to be documented. These include diabetes (measures 122v5, 123v5, 131v5, 134v5), pharyngitis (146v5), pregnancy or sexually transmitted disease (153v5), respiratory infections (154v5), hypertension (165v5), and back pain (166v6). We examined all 25,291 problems that were docu-

mented on the 1,100 patients to determine the documentation of the time of problem onset. Note that 51.7% of problems had no onset date documented. In addition to the omission of problem onset date, we also examined other sections in the clinical documents which may contain problems that were not on the problem list. These included the history of past illness and the encounters sections. We found 5,483 incremental problems or diagnoses in these sections, which represented a meaningful percentage (21.1%) of overall problems.

To address these issues, we used all sections of clinical documents that may include problems and changed our measure logic to address problem onset omission. Specifically, if a problem was documented as active, we assessed that the onset date must have been prior to visit date (i.e., it is not reasonable that any clinician would document a problem to occur in the future).



```

Human Readable Content
<text>
  <table>
    <tr>
      <content ID="enc200417">Office Visit - Hypertension</content>
    </tr>
  </table>
</text>

Machine Readable Content
<entry>
  <encounter classCode="ENC" moodCode="EVN">
    <templateId root="1.3.6.1.4.1.19376.1.5.3.1.4.14" />
    <templateId root="2.16.840.1.113883.10.20.1.21" />
    <templateId root="2.16.840.1.113883.3.88.11.83.16" />
    <templateId root="2.16.840.1.113883.3.88.11.32.17" />
    <id extension="22222222" root="1.2.3.4.5.6.7" />
    <code codeSystem="2.16.840.1.113883.5.4" codeSystemName="ActEncounterCode">
      <originalText>
        <reference value="#enc200417" />
      </originalText>
    </code>
    <text>
      <reference value="#enc200417" />
    </text>
    <effectiveTime>
      <low value="20170210090802-0500" />
      <high value="2017021091834-0500" />
    </effectiveTime>
  </encounter>
</entry>

```

No encounter code included but reference (#enc200417) points to human readable

Fig. 2 Illustrative example of encounter normalization. This example from a clinical document, edited to protect patient identity, demonstrates how code omission in the XML (highlighted in yellow) would normally exclude this patient from being included in quality measures. Using the text of “office visit” in the reference tag, however, allows a valid code to be selected from appropriate terminology.

Iterative Improvements for Quality Measure Compliance

Compliance Improvement through Value Set Mapping. Electronic clinical quality measures use a set of codes, often referred to as “value sets,” to determine whether a specific activity was performed. For example, with breast cancer screening (125v5), the measure specifies a value set of mammography studies that would qualify a mammography as being performed. Through the examination of specific records, we found the specific codes used in these value sets have a material impact on quality measure calculation. With mammography, all the specified codes were from Logical Observation Identifier Names and Codes (LOINC). As shown in ▶Table 2 for mammography, none of the eligible patients for this measure had one of those LOINC codes in the appropriate time period since the compliance rate was 0%. This electronic clinical quality value set for mammography, however, varies from the value set for the equivalent HEDIS measure for mammography, which allows for Current Procedural Terminology, ICD-9, and ICD-10 codes.

We contacted NCQA, who is a measure steward for 16 of the 17 measures included in this research, to discuss this specific concern. They agreed that for the measures where codes were included in HEDIS, equivalent concepts are acceptable through mapping (Smith A, Archer L, at National Committee for Quality Assurance, phone call, November 2017). This significantly increased compliance for the cancer preventative screening measures (124v5, 125v5,

130v5). This process would be expected to have had an impact on the two diabetes measures (123v5, 131v5) although no change was observed based on the small eligible populations for these measures.

Compliance Improvement through Medication Normalization. Electronic clinical quality measures use a national standard vocabulary, RxNorm, established by the National Library of Medicine for medication-related logic. RxNorm is a normalized naming system that contains concepts spanning ingredient, coordinated dose forms, generic name, and brand names. When value sets are created for medication usage, however, they often include only generic concepts, omitting branded and ingredient concepts. There are significant challenges in using such a limited value set. First, we found that 3,095 (14.9%) of medications collected in this sample are not coded in RxNorm. These likely included medications affecting measure calculation, but without terminology mapping would provide inaccurate results. Second, we found that the term types of RxNorm codes in real-world data often did not match the measure value set. Specifically, only 12,146 (69.3%) of RxNorm-coded medications were mapped to a generic drug concepts that align with quality measure value sets. The combined effect of medications not coded in RxNorm and not mapped to generic medication concepts are that only 58.6% of real-world medications from our samples appropriately functioned with quality measures that include medication logic.

The resolution to this inability to identify medications for measure calculations was to use terminology mapping of

medications that were available in the research software. This mapping included relationships between the RxNorm term types publicly available as well as proprietary technology for the free-text mapping of medications names. This successfully mapped 18,767 (90.6%) of the original medications to a usable RxNorm concept which could then be applied to the quality measure logic. For the remaining 1,956 medications that were not mappable, manual review showed that 460 were vitamins (e.g., multivitamins that did not specify content), 360 were medical supplies (e.g., lancets, test strips, nebulizers), and 191 were “unknown” or null entries. These types of entries were not applicable to the quality measures selected. This left 945 (4.5%) of medication entries not available to quality measure logic. Several of these were actual medications, but others were concepts recorded in a manner which did not detail a specific ingredient (e.g., “allergy immunotherapy” or “hormones”). The effective yield of usable medication data was approximately 95% (18,767 mapped medication entries vs. 945 unmapped medication entries).

Once translations were performed, it was also necessary to adjust the logic associated with medication administration before medication quality logic would function appropriately. Specifically, 17,505 (84.5%) of all medications were recorded in clinical document as medication orders (i.e., HL7 moodCode of “INT”). Of those, however, 14,318 (81.8%) had an associated start date at or before the clinical encounter. For medications that had a start date in the past, we treated them as administered medication events rather than intentional. This allowed the medication duration logic of High Risk Medications in the Elderly (156v5) to function (i.e., have at least 1 numerator event). This issue may stem from poor implementation of the clinical document standards as detailed in prior research.¹⁴

Compliance Improvement through Laboratory and Vital Sign Normalization. Often laboratory information recorded in EHRs does not meet the value set of laboratory results in quality measures. This impacted the diabetes control measure (122v5) which required HbA1c results. Using all the result data in the collected information, 4.1% of all HbA1c results were found to not have the appropriate LOINC code. In addition, 14.8% of these HbA1c results did not use the appropriate unit of measure (i.e., %) for the laboratory results. An even larger impact was shown among laboratory results related to the diabetes nephropathy measure (136v5), where 18.3% of results were not shown to have appropriate code. For pediatric body mass screening measure (155v5), while vital signs used the appropriate LOINC code for body mass index (BMI), 35.1% did not use the appropriate unit (i.e., kg/m²). The solution to this was to normalize laboratory and vital signs using both code mapping and unit translation to transform data for the above examples which affected measures 122v5, 134v5, and 155v5.

Compliance Improvement through Logic Changes. Finally, additional logic changes were attempted for three pediatric-related measures. For the pediatric testing of pharyngitis, the relationship between the timing of the encounter, medication start, and problem onset were simplified. For the treat-

ment of childhood upper respiratory infections (154v5), we found that the relationship between encounter timing, problem onset, and medication timing could not be simplified to make this measure include a reasonable portion of patients. Attempted resolutions for this measure were unsuccessful. For the measure relating to pediatric weight (155v5), we found that the requested vital sign of BMI percentile was never recorded in interoperable clinical documents we examined. Using the data that was recorded on BMI, gender, and patient age, however, permitted the calculation of the appropriate percentile for part of this measure (i.e., BMI percentile was unambiguously known from information provided).

Resultant Quality Measure Calculations

Of the original 17 measures selected, we found two measures (166v6 and 82v4) where the eligible population remained under 5 patients from the sampled population of 1,100. In addition, all attempted changes to the treatment of upper respiratory infections measure (154v5) were not able to reasonably reduce the exclusion rate. These three measures were considered to be nonfunctional despite attempts to increase the eligible populations in (→Table 2—“Measures not included in Iterative Improvement”). For the remaining 14 measures, we report both the original and the normalized quality measure rates in →Table 2 (“Calculation after Iterative Improvement”).

The overall impact of the iterative improvement on the eligible population increased the denominator populations across these 14 measures from 803 to 1,783 (+122%). This counts the same patient multiple times when the patient qualifies for multiple measures. The number of unique patients included in at least one measure increased from 315 to 601 (+91%).

The overall impact of the iterative improvement in compliance was varied. Five measures saw an increase from no applicable compliance to a nonzero number. One measure decreased from 100% compliance to a lower rate. Three measures had at least one rate component remain at zero compliance despite attempts to improve compliance. Other measures had small or moderate changes in reported compliance.

Once these changes were made, the 14 revised measures were extensively tested to determine if certification compliance was maintained. Appropriate Testing for Children with Pharyngitis (146v5) was found to not maintain certification. While data are presented for this measure, the revised logic could not be used in reporting. Certification for the other 13 measures was unaffected since techniques for free-text normalization, terminology mapping, or missing data as addressed through the iterative improvement do not affect certification test data, which include only properly structured data.

Discussion

Implications of these results can be categorized into two domains: considerations for measure authors and stewards and considerations for organizations performing quality calculation.

Considerations for Measure Authors and Stewards

Quality measure development is a difficult task often done in the abstract; authors lack heterogeneous clinical data sets to validate logic and examine how real-world documentation practices affect calculations. Our findings support the need for measure developers to better understand how the routine collection of clinical data impacts quality measurement, as policymakers have acknowledged.²⁴ That requires access and testing with real-world data before a measure is released for use. This will help measure authors evaluate the inherent limitations of terminologies, value sets, discrete data entry, and cohort definitions in the process of measure development. It also helps identify gaps between clinical data collection and the available data for reporting. This study validated that interoperability standards for clinical documents, as promoted by the Meaningful Use program, is a viable strategy. In addition, the use of interoperability standards provides a clear audit trail back to the source EHR. Auditing using interoperability standards can include both the original source information and any data transformations performed. This becomes increasingly important as both private and public payers use quality measure performance for provider payment.

Another finding is the importance of measure consistency across programs. We observed that value sets for terminologies varied substantially from HEDIS to electronic clinical quality measures. Specifically, some terminologies included in HEDIS were excluded in clinical quality measures. This caused several preventative measures to report zero compliance, when any observer would find evidence of the clinical prevention in the data. We strongly believe that there should be alignment and compatibility of value sets across measure programs, particularly since providers have been encouraged to document in a way which supports older programs such as HEDIS. This need for consistency also applies to how patients are qualified for measures as documented in other research.²⁵ Electronic clinical quality measures incorporate the concept of a specific type of visit before a patient is eligible for quality measure calculation. The lack of proper encounter coding in EHRs creates a burden in this domain. HEDIS measures apply to broader member populations based on billing profiles, while electronic clinical quality measures are artificially restricted. Such attribution logic also overlooks patients who go 12 to 24 months between physician visits and emerging modalities where virtual encounters are used for patients in good health. We believe that measure eligibility logic should recognize these concerns to ensure greater consistency across programs.

Finally, poor documentation practices, such as free-text order entry or missing qualifiers, should never result in better compliance. In the example of high risk medications in the elderly, we found higher compliance when medication data were not normalized. This rewards clinicians and technologies that do not record medications in the standard terminology. Since we found 41% of medications were not in the expected term type of RxNorm, this issue of normalization for complex clinical data, such as medications, will remain important for the near term.

Considerations for Organization Performing Quality Calculation

This study validates that the strategy promulgated by MACRA to establish QCDRs for quality measurement is technically feasible for at least several measures. It also demonstrates viability of collecting clinical data from various sources using interoperability standards that could be adopted by integrated delivery systems with multiple in-house EHRs. While the compliance rates reported for selected measures vary from known benchmarks, we believe that to be reasonable given the limited data examined and the fact that selected facilities were not known to have any focus on the selected measures. Measure selection by QCDRs will be important based on the findings of this research. Also important will be the selection of a technology vendor to collect and normalize clinical data. Our findings substantiate the value in transforming clinical data collected using interoperability standards, as had been previously demonstrated for individual EHRs.²⁶

In addition, clinical documentation practices should always remain a priority when working with providers who intend to use a QCDR to support electronic clinical quality measurement. For several of the measures with low or zero compliance rates, the information required is often not structured in the appropriate place to be available for quality measure calculation, as documented in prior research.²⁷ For example, we never found nutritional or physical activity counseling to be documented as a particular code for the pediatric weight assessment measure, but we fully expect this was performed on at least some of the 123 eligible pediatric patients. Previous research has validated that practice type, size, and experience with EHR technology have significant impacts in data availability for quality reporting.²⁸ Further work with local practices and EHRs will be required to implement tactics that will increase data completeness.

Since QCDRs have access to real-world data and the ability to author measures, they are in a unique position to advance the state of quality measure development. We believe that cross-industry collaboration between QCDRs and payers needing quality measurement for value-based contracting measure will be critical. These collaborations could include deidentified data repositories for new measures, measure validation using real-world clinical data, and best practices in data transformation to support quality measurement.

Finally, some QCDRs are tightly integrated with a health information exchange, and we believe this research highlights an important implication. Improving clinical data will not only improve clinical quality measurement but will also improve care transitions and improvement objectives supported by HIEs. We believe that using interoperability standards to empower quality measurement provides an incentive and feedback loop to improve interoperability generally.

Limitations

This study was limited in several dimensions. First, it used a single clinical document to calculate the quality measures. Had multiple documents been used, the rates for both patient

inclusion and compliance would likely have been different. Other data sources, such as QRDA or Fast Health Interoperability Resource extracts, may have provided additional data than what was recorded in available clinical documents but were not examined in this research.^{29,30} Moreover, using electronic data capture for quality measurement has been shown to differ from manual abstraction and is not examined in this research.²⁰ Next, only a single measurement technology was used in this research. Nine vendors have been certified by the NCQA to calculate quality measures and dozens more are certified by other authorized testing laboratories.²¹ We fully expect that other technologies will generate different results, even based upon the same data set. Data transformations performed by any software may introduce variability and potential data anomalies to quality measurement, although the process of software certification helps minimize inadvertent errors. Finally, no facility was contacted in advance about this study so no effort was specifically expended to improve measure documentation or compliance. Further research should establish how longitudinal, multisource clinical data may impact quality measure calculation as it may be anticipated that such data would provide better rates than those observed from the point-in-time information examined in this research.

Conclusion

Quality measure calculation is possible using interoperability standards to collect data from a variety of EHRs. Quality measure stewards should increasingly use real-world data in their measure development to validate measure integrity, reliability, and consistency. The selection of specific quality measures by QCDRs will be an important consideration since quality measures may have issues affecting inclusion and compliance calculation, even when using certified technology. The use of interoperability standards to support quality measurement provides a long-term incentive to jointly improve interoperability, clinical documentation, and care quality. This will be paramount as payers transition to value-based contracting.

Clinical Relevance Statement

The use of clinical data exchanged routinely from EHRs can empower quality measurement. The results described in this article specify how to improve patient inclusion and measure accuracy using an iterative approach. Organizations that report quality measurement should be aware of how such techniques affect compliance rates for reported quality measures.

Multiple Choice Question

Why can the transformation of medication data from certified EHRs improve quality measure calculation?

- Medication administration instructions are different among EHRs
- Medication data need to align with the subset of codes, known as a "value set," used by the quality measure

- Medication doses can change for the same patient over time
- All medications recorded by clinicians were unstructured and need codification before quality measurement can occur

Correct Answer: The correct answer is option b. This research found that over 40% of medication data coding from certified EHRs varied from the "value sets" used by quality measure logic. Consequently, transformation of the medication data is required for the appropriate calculation of measures. Terminology mapping is one technique that markedly improves the usability of medication data within interoperable clinical document. This research made similar observations in other clinical domains, such as problems, encounters, laboratory results, and vital signs.

Protection of Human and Animal Subjects

This study was approved by the Institutional Review Board for the University of Texas, Health Science Center, Committee for the Protection of Human Subjects. Technical and administrative safeguards were utilized to protect the privacy of such information throughout this research.

Funding

Support for this research was provided by the Kansas Health Information Network and Diameter Health, jointly donating time and resources to the research team.

Conflict of Interest

John D'Amore, Chun Li, and Jonathan Niloff receive salaries from and have an equity interest in Diameter Health, Inc., whose software provided the quality measure calculation used in this research. Dean Sittig serves as a scientific advisor with an equity interest in Diameter Health.

Acknowledgments

We would like to thank many who assisted in this study. From KHIN, we would like to acknowledge Andy Grittman and Vince Miller who provided the technical infrastructure for the research. We would also like to thank Mary Matzke and Jody Denson from KHIN in their research assistance of issues related to quality reporting. From Diameter Health, we acknowledge Judith Clark and Dan Andersen for their assistance regression testing measures for certification and Tom Gaither in assembling the research team. From NCQA, we would like to thank Ben Hamlin, Anne Smith, and Latasha Archer who were responsive to questions and discussion.

References

- Better Care. Smarter Spending. Healthier People: Paying Providers for Value, Not Volume. Centers for Medicare and Medicaid Services; 2015. Available at: <https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2015-Fact-sheets-items/2015-01-26-3.html>. Accessed January 7, 2018
- Japsen B. UnitedHealth, Aetna, Anthem Near 50% Value-Based Care Spending. Forbes, Inc. Available at: <https://www.forbes.com/sites/brucejapsen/2017/02/02/unitedhealth-aetna-anthem-near-50-value-based-care-spending/>. Accessed January 7, 2018

- 3 HEDIS & Quality Measurement. National Committee for Quality Assurance. Available at: <http://www.ncqa.org/hedis-quality-measurement>. Accessed January 7, 2018
- 4 Jung K. The impact of information disclosure on quality of care in HMO markets. *Int J Qual Health Care* 2010;22(06):461–468
- 5 Eddy DM, Pawlson LG, Schaaf D, et al. The potential effects of HEDIS performance measures on the quality of care. *Health Aff (Millwood)* 2008;27(05):1429–1441
- 6 Mainous AG III, Talbert J. Assessing quality of care via HEDIS 3.0. Is there a better way? *Arch Fam Med* 1998;7(05):410–413
- 7 Participation continues to rise in Medicare Physician Quality Reporting System and Electronic Prescribing Incentive Program. Centers for Medicare and Medicaid Services; 2015. Available at: <https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2015-Fact-sheets-items/2015-04-23-1.html>. Accessed January 7, 2018
- 8 Hsiao C-J, Decker SL, Hing E, Sisk JE. Most physicians were eligible for federal incentives in 2011, but few had EHR systems that met meaningful-use criteria. *Health Aff (Millwood)* 2012;31(05):1100–1107
- 9 2016 Report To Congress on Health IT Progress. The Office of the National Coordinator for Health Information Technology (ONC) Office of the Secretary, United States Department of Health and Human Services. Available at: https://www.healthit.gov/sites/default/files/2016_report_to_congress_on_healthit_progress.pdf. Accessed January 7, 2018
- 10 Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. *Med Care Res Rev* 2010;67(05):503–527
- 11 Torda P, Tinoco A. Achieving the promise of electronic health record-enabled quality measurement: a measure developer's perspective. *EGEMS (Wash DC)* 2013;1(02):1031
- 12 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
- 13 Furukawa MF, King J, Patel V, Hsiao C-J, Adler-Milstein J, Jha AK. Despite substantial progress in EHR adoption, health information exchange and patient engagement remain low in office settings. *Health Aff (Millwood)* 2014;33(09):1672–1679
- 14 D'Amore JD, Mandel JC, Kreda DA, et al. Are Meaningful Use Stage 2 certified EHRs ready for interoperability? Findings from the SMART C-CDA Collaborative. *J Am Med Inform Assoc* 2014;21(06):1060–1068
- 15 Clough JD, McClellan M. Implementing MACRA: implications for physicians and for physician leadership. *JAMA* 2016;315(22):2397–2398
- 16 Cottingham S. PopHealth primer. ONC funds open-source software to streamline clinical quality measures reporting for meaningful use program. *J AHIMA* 2011;82(09):48–50
- 17 D'Amore JD, Sittig DF, Wright A, Iyengar MS, Ness RB. The promise of the CCD: challenges and opportunity for quality improvement and population health. *AMIA Annu Symp Proc* 2011;2011:285–294
- 18 Taggart J, Liaw S-T, Yu H. Structured data quality reports to improve EHR data quality. *Int J Med Inform* 2015;84(12):1094–1098
- 19 Gardner W, Morton S, Byron SC, et al. Using computer-extracted data from electronic health records to measure the quality of adolescent well-care. *Health Serv Res* 2014;49(04):1226–1248
- 20 Kern LM, Malhotra S, Barrón Y, et al. Accuracy of electronically reported “meaningful use” clinical quality measures: a cross-sectional study. *Ann Intern Med* 2013;158(02):77–83
- 21 Current Vendor Information. National Committee for Quality Assurance. Available at: <http://www.ncqa.org/hedis-quality-measurement/data-reporting-services/emeasure-certification/Current-Vendor-Information>. Accessed January 7, 2018
- 22 Quality Data Model from eCQI Resource Center. Centers for Medicare and Medicaid Services: Office of the National Coordinator for Health IT. Available at: <https://ecqi.healthit.gov/qdm>. Accessed January 7, 2018
- 23 Percent of Adults with Diagnosed Diabetes; 2015. Kansas Health Matters. Available at: <http://www.kansashealthmatters.org/index.php?module=indicators&controller=index&action=view&indicatorId=2270&localeId=19>. Accessed January 11, 2018
- 24 Conway PH, Mostashari F, Clancy C. The future of quality measurement for improvement and accountability. *JAMA* 2013;309(21):2215–2216
- 25 McCoy RG, Tulledge-Scheitel SM, Naessens JM, et al. The method for performance measurement matters: diabetes care quality as measured by Administrative Claims and Institutional Registry. *Health Serv Res* 2016;51(06):2206–2220
- 26 Amster A, Jentzsch J, Pasupuleti H, Subramanian KG. Completeness, accuracy, and computability of National Quality Forum-specified eMeasures. *J Am Med Inform Assoc* 2015;22(02):409–416
- 27 Parsons A, McCullough C, Wang J, Shih S. Validity of electronic health record-derived quality measurement for performance monitoring. *J Am Med Inform Assoc* 2012;19(04):604–609
- 28 Heisey-Grove DM, Wall HK, Wright JS. Electronic clinical quality measure reporting challenges: findings from the Medicare EHR Incentive Program's Controlling High Blood Pressure Measure. *J Am Med Inform Assoc* 2018;25(02):127–134
- 29 HL7 CDA® R2 Implementation Guide: Quality Reporting Document Architecture - Category I (QRDA I) STU Release 5–US Realm. Health Level 7; 2017. Available at: http://www.hl7.org/implementation/standards/product_brief.cfm?product_id=35. Accessed March 1, 2018
- 30 Fast Health Interoperability Resource STU. 3. Health Level 7; 2017. Available at: <https://www.hl7.org/fhir/>. Accessed Mar 1, 2018