The Promise of the CCD: Challenges and Opportunity for Quality Improvement and Population Health

John D. D'Amore, MS[1], Dean F. Sittig, PhD[1], Adam Wright, PhD[2],
M. Sriram Iyengar, PhD[1] Roberta B. Ness, MD, MPH[3]
[1]**School of Biomedical Informatics and** [3]**School of Public Health at the University of Texas Health Science Center, Houston, TX;** [2]**Division of General Medicine, Brigham and Women's Hospital, Boston, MA**

Abstract

Interoperability is a requirement of recent electronic health record (EHR) adoption incentive programs in the United States. One approved structure for clinical data exchange is the continuity of care document (CCD). While primarily designed to promote communication between providers during care transitions, coded data in the CCD can be re-used to aggregate data from different EHRs. This provides an opportunity for provider networks to measure quality and improve population health from a consolidated database. To evaluate such potential, this research collected CCDs from 14 organizations and developed a computer program to parse and aggregate them. In total, 139 CCDs were parsed yielding 680 data in the core content modules of problems, medications, allergies and results. Challenges to interoperability were catalogued and potential quality metrics evaluated based on available content. This research highlights the promise of CCDs for population health and recommends changes for future interoperability standards.

Introduction

Recent incentives and policy decisions are promoting the rapid adoption of electronic health records (EHRs) in the United States. The American Reinvestment and Recovery Act of 2009 commits up to $27 billion in payments, beginning in 2011, to eligible professionals and hospitals that meaningfully use EHRs[1]. Those reimbursements will come in three stages and are expected to propel ambulatory and hospital EHR adoption to over 70% by 2020[2]. The rapid timeline for uptake, however, will lead to a heterogeneous environment of technology. With over 400 EHRs certified for the first stage of 'Meaningful Use,' interoperability will remain a concern.[3] As providers seek to improve quality and population health, technology standards advanced by this federal legislation will enable new methods for data aggregation.

Background

In July 2010, the Department of Health and Human Services adopted the Continuity of Care Document (CCD) as an option to meet the goals of clinical data exchange for 'Meaningful Use'[4]. Using an extensible markup language (XML) based structure, the CCD was collaboratively developed in 2006 by harmonizing standards from the American Society for Testing and Materials and Health Level 7 (HL7)[5]. The CCD provides a flexible format for the communication of free-text and codified data. Given the recent emergence of the standard, most health information exchanges are not routinely using CCDs today, although select institutions have launched pilots to explore their potential[6, 7]. The lack of widespread use means that EHR developers must rely on guidance from standards organizations, such as the Health Information Technology Standards Panel (HITSP), Integrating the Healthcare Enterprise (IHE) and HL7, on how to create and exchange CCDs.

HITSP released for implementation its first CCD patient summary construct, named C32, in 2007[8]. That construct is directly referenced in the final federal rule for Stage 1 of 'Meaningful Use'[4]. The most recent C32 specification references two other constructs developed by HITSP as well as technical frameworks previously released by IHE and HL7[9, 10]. Naturally, documents and specifications from different organizations and developed at different times may lead to varying interpretations about requirements. For Stage 1 of 'Meaningful Use,' the National Institute for Standards and Technology (NIST) has released the definitive testing procedures to determine whether an EHR-generated CCD meets the standards for 'Meaningful Use' certification[11]. These procedures focus on the ability of EHRs to generate, receive and display four categories of coded patient data with specific vocabularies: problem lists,

diagnostic test results, medication lists and medication allergy lists. Although the CCD can encode additional clinical content, these four sections plus patient demographic information form the foundation for what certified EHRs will be capable of exchanging for the first two years of the federal incentive program.

New models of care integration being advanced by private insurers and federal payers require data from multiple clinical entities to determine if patients are receiving appropriate care[12]. While the primary intent of clinical data exchange is provider-to-provider communication, the structured data of the CCD has potential re-use. Specifically, provider based care networks could create CCD extracts for all patients and consolidate this information into a longitudinal clinical data warehouse.

This strategy overcomes four significant barriers facing providers in the United States: 1) providers avoid the costs and competitive threats associated with joining a health information exchange[13], 2) this strategy works in regions that do not have an existing health information exchange, 3) networks do not need to consolidate independent care providers onto a single EHR technology[14], and 4) quality measurement exempts data aggregation from privacy restrictions since protected health information may be shared among clinicians for quality improvement[15].

This research examines the promise of using C32 compliant CCDs to create a normalized database for quality improvement and population health management. The context for such data aggregation would be a care model where providers with different certified EHRs have entered into a data use agreement to share identified medical records. This research focuses on the clinical content modules of the CCD being tested for Stage 1 EHR certification as part of 'Meaningful Use' (See Table 1 for all CCD sections and associated vocabularies). To examine the potential for quality measurement, each of the ambulatory measures endorsed as part of Stage 1 'Meaningful Use' were evaluated based on the parsed CCD clinical content modules.

Table 1. Sections and Vocabularies of the CCD and Modules Examined in this Research.

Clinical Content Modules in C32 Standard	Primary Vocabulary for Clinical Content	Required by C32 Standard	Required by Stage 1 EHR Testing Procedures	Extracted with CCD Parser
Advanced Directive	Systematized Nomenclature of Medicine (SNOMED CT)			
Allergy & Drug Sensitivity	Uniform Ingredient Identifier (UNII) & RxNorm		Medication Allergies Only	✓
Comment	Free Text			
Condition (Problems)	SNOMED CT & International Classification of Diseases (ICD)*		✓	✓
Encounters	Uniform Billing (UB) Standard			
Healthcare Provider	National Provider Identifier			
Immunizations	Vaccine Value Sets			
Insurance Provider	X12 Billing Standard			
Language Spoken	Language Value Set	✓	Implied, but not specified	
Medication	RxNorm		✓	✓
Person Information	Free Text & HL7	✓	✓	✓
Plan of Care	Free Text			
Procedure	SNOMED CT, ICD and Current Procedural Terminology (CPT)*		Required for hospital but not ambulatory CCD exchange	
Results	Logical Observation Identifiers Names and Codes (LOINC)		✓	✓
Support	Free Text & HL7	✓	Implied, but not specified	
Vital Signs	LOINC			

* While the HITSP preferred vocabulary for problems and procedures is SNOMED, the Final Rule for EHR certification provides flexibility for other vocabularies as specified in the Table 1. Note: Information Source and Pregnancy status are modules not included in the above table since they overlap with content provided in the CCD header and condition modules.

Methods

To assess the feasibility of a CCD-based aggregation strategy, samples needed to be collected from multiple EHR vendor products. No large collection of sample CCDs from multiple sources has been made available to the public, so the research team contacted vendors and healthcare organizations for EHR-generated samples of fictitious patient data conformant to the HITSP C32 standard. A request for sample CCDs was included in the February *eNews* distribution email sent from the Certification Commission for Health Information Technology to approximately 12,000 recipients. Additional contacts were made in person during HIMSS 2011 at the *Interoperability Showcase* and with individuals in the exhibit hall. As an incentive to submit, organizations providing sample CCDs were offered feedback on the parsing results of their samples. Organizations that submitted samples were also assured that this research not would identify specific EHRs and that their CCDs would not be publicly released. These two methods yielded the majority of participating organizations, but additional CCDs were also examined from standards organizations and a research library of synthetic EHR data from ExactData (Rochester, NY).

To aggregate and analyze the collected CCDs, a program was written in Python 3.1 (Python Software Foundation, Wolfeboro Falls, NH) to parse clinical content modules and patient demographic details. The Python program utilized the document object module (DOM) library for XML parsing and used HITSP documentation for the identification of relevant sections in the CCD. Any data provided outside the modules identified for this research were ignored. String-based lists encoded clinical content for each of the modules. Data elements were imported based on the tags identified in the C32 construct. Separate lists were created to store the primary content of a module (e.g. medication code), its associated vocabulary (e.g. RxNorm) and other relevant content (e.g. dose, dosing interval and brand name). At least one clinical content module was imported from each CCD.

The Python program included a timestamp on each of the CCDs as they were imported to report efficiency of XML parsing. All processing was performed on a quad-core Intel Core2 computer running Windows 7 (Redmond, WA) with 3Gb of RAM. Results included the processing time of each CCD as well as total counts of data elements for each of the primary content modules. Generic medication codes were counted for the medication module, problem codes for the condition module, laboratory result codes for the results module and allergy codes for the allergy module. Null values were only counted if the section tags were present or corresponding detail was provided as translation codes.

The programming included error traps and empty string notifications to catalog the challenges of CCDs imported from various EHRs. These warnings prompted formative evaluation of approximately 30 XML files to manually identify the causes for unsuccessful parsing. At least 1 CCD was selected for manual inspection from each submitting organization. All inspections were conducted by a single reviewer to maintain evaluative consistency, although no formal instrument was used given the wide range of issues encountered. In addition, the data were examined for what potential conflicts may exist if imported into a single relational database. Since the imported content was under 1,000 clinical content modules, those data were reviewed in Microsoft Excel 2007 (Redmond, WA).

To determine if the collected content was adequate for objective quality assessment, each of the 44 ambulatory quality measures included in Stage 1 'Meaningful Use' was evaluated[16]. If coded data in patient detail and clinical modules were sufficient to calculate the measure, then that quality metric was recorded as possible using the CCD parser. If not, the fact that content was missing was recorded. The electronic standards for ambulatory quality measures have been approved through the National Quality Forum and retrieved from the federal website for electronic specifications[16].

Results

In total, 196 CCDs were collected from 14 different organizations representing at least 10 different EHRs (Table 2). Several CCDs (n=57) were excluded from analysis since no clinical data were present for parsing or they were redundant with other submitted CCDs. This left 139 CCDs which were successfully parsed with at least one clinical data element. Parsing time averaged 80ms (SD 53ms) per CCD and the mean number of clinical data elements was 4.9 (SD 5.4) per CCD (Figure 1 shows timing and content distribution of CCDs). The counts of parsed data elements by clinical module were: 109 for allergies, 220 for problems, 168 for medications and 183 for diagnostic results.

Table 2. CCD Collection by Source

Data source	Number of organizations	Number of CCDs Received	Number of CCDs Parsed
EHR Vendors & Synthetic EHR Data	6	93	87
Healthcare Providers	5	93	49
Standards Organizations	3	10	3
Total	**14**	**196**	**139**

Figure 1. Parsing Time and Content for 139 CCDs.*
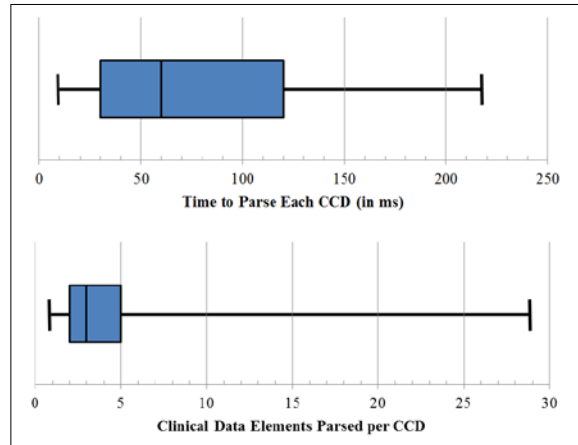
*Chart whiskers show absolute minima and maxima; the blue box represents lower and upper quartile around the median center line.

Challenges were categorized into three major themes: 1) Problematic CCD hierarchy and organization, 2) Inconsistency in data representation and 3) Data conflict or redundancy within the CCD (Illustrative examples included in Table 3).

1. Problematic CCD Hierarchy and Organization

One of the common issues encountered when working through the CCDs was the lack of consistent template root identifiers for the clinical content modules. This is critical to extract codified data since identifiers reference the technical specification applied in XML formatting. Specifically, a majority of EHR samples did not include the root template identifiers for HITSP C83 content modules even though these identifiers are referenced in the 2009 HITSP C32 specification and CCD samples[8,10]. To accommodate the missing identifiers, the Python program was modified to include checks so that any template root identifier from a standards organization would be sufficient to parse the content modules analyzed in this research.

Next, the use of tabs and line breaks was inconsistent between different EHRs. Some EHRs do not use line breaks or spacing, some use tabs without consistent line breaks and others use a combination to provide formatting similar to the examples from standard organizations. While line breaks and spacing do not affect the ability of programs utilizing the DOM library for XML parsing, human review of the CCD to identify problematic sections and data elements becomes significantly more difficult.

The optionality of data elements within each clinical content module presented additional difficulty in creating a normalized database. Relevant data elements with a problem, medication, allergy or result include information on the time of onset, current status, units of measurement, dosing interval, severity and result interpretation. Unfortunately, the C32 construct leaves most of the associated data as optional, which can therefore be omitted in a compliant CCD. Our parsing results display a large number of records that omit data like a date of a problem's onset or drug dosing interval.

2. Inconsistency in Data Representation

Since EHR data typically undergo a translation between the source system and the normalized vocabularies for Stage 1 'Meaningful Use', there is the opportunity for mappings to no known code. Incomplete mappings were observed in this research and have been previously identified in medication interoperability research and harmonization discussions for the CCD[6,17]. An example of this occurs when a null value is set for a clinical content code such as generic medication in RxNorm, while a translation code is populated, such as brand name in the National Drug Classification (NDC).

In addition to missing values, non-uniformity was observed in the associated data content for each code. One example repeatedly noted in our research was inconsistency of effective time for problems and laboratory results. While the examples provided by HITSP are generally eight characters in YYYYMMDD format, some EHRs adapt this format to include six more characters for time in HHMMSS format. Others also append a hyphen and four digits for HHMM. One EHR did not comply with the character date format at all, instead inserting values such as '% 2m %', presumably meaning two months. Inconsistency was also noted in the dose quantity tag for solid oral medications, which included alternative unit labels such as 'tablet,' 'Tablet,' 'tab,' 'tab(s),' 'mg,' and 'g.' Only the mass terms qualify as standard units, although some flexibility is permitted in specifications to identify non-units, like tablet or red blood cell count, when deemed important[18].

Another data representation challenge identified in this research includes content relation between laboratory results. One example is that corresponding procedures coded within the same module could not always be related to appropriate results. While the published example CCD from NIST approaches this by collapsing a single procedure, such as a blood draw, with all corresponding results into a single XML entry, other EHRs included all procedures and results within the same entry. This thereby eliminates any codified relation between the procedure and result. Another example was the use of a comprehensive code for a testing panel with multiple lab values. Since multiple lab results are returned for comprehensive panels, such as blood urea nitrogen and transaminases within a metabolic panel, the generic label of a lab panel was insufficient to interpret the content of each test result.

3. Data Conflict or Redundancy

The final rule for Stage 1 'Meaningful Use' specifies either SNOMED-CT or ICD-9 for the codification of problems and both vocabularies were observed in the sample CCDs collected. In addition to conflict generated from two acceptable vocabularies for the same information, many sample CCDs utilized code systems that did not conform to standards. Observed misapplied terminologies included NDC for generic drug code, NDC for medication allergy and Multum Drug Allergy codes for medication allergy. In addition, a few CCDs had codes that did not correspond to any recognized standard vocabulary.

Lastly, there was a level of redundancy of content transmitted in portions of the CCD. One example is the RxNorm vocabulary, using codes like '309362' for 'clopidogrel 75 mg oral tablet.' This code contains information on drug dose and route of administration, but many CCDs included additional XML elements that also code such information. In other cases, some CCDs omitted redundant content by skipping optional data elements altogether.

Table 3. Illustrative Challenges from 139 CCDs from Multiple EHRs.

Category	Example Issues	Example XML*

Problematic CCD Organization & Hierarchy	This section on medications has the HL7 root identifier, shown in yellow, but is missing the corresponding root identifiers for HITSP and IHE specifications. Since dosage and route of administration are optional tags in the medication section, those data are skipped in this XML section although known based on the RxNorm code, shown in blue.	`<section classCode="DOCSECT" moodCode="EVN">`
		`<templateId root="2.16.840.1.113883.10.20.1.8" assigningAuthorityName="HL7 CCD"/>`
		`<code code="10160-0" codeSystemName="LOINC" displayName="History of medication use"/>`
		`<entry contextConductionInd="true" typeCode="COMP">`
		`<substanceAdministration classCode="SBADM" moodCode="EVN">`
		`<templateId root="2.16.840.1.113883.10.20.1.24"/>`
		`<id extension="12039106" root="1.3.6.1.4.1.16517"/>`
		`<statusCode code="active"/>`
		`<effectiveTime xsi:type="IVL_TS">`
		`<low value="20100630"/>`
		`</effectiveTime>`
		`<consumable typeCode="CSM">`
		`<manufacturedProduct classCode="MANU">`
		`<templateId root="2.16.840.1.113883.10.20.1.53"/>`
		`<manufacturedMaterial classCode="MMAT" determinerCode="KIND">`
		`<code code="309362" codeSystemName ="RxNorm" displayName="clopidogrel 75 MG Oral Tablet">`
		`<originalText>PLAVIX (CLOPIDOGREL)</originalText>`
		`</code>`
		`</manufacturedMaterial>`
		`</manufacturedProduct>`
		`</consumable>`
		`</entry>`
		`</section>`

Inconsistency in Data Representation	A panel laboratory code is used to refer to multiple laboratory values. The difference between BUN and AST, shown in blue, is written into the text field but each LOINC value only codes for the comprehensive panel, shown in yellow. Consequently individual test values, shown in green, are not connected with corresponding lab results from the panel.	`<text>`
		`<tbody>`
		`<tr>`
		`<td> <content ID="result-4">Comprehensive Metabolic Panel</content> </td>`
		`<td>  BUN=16 mg/dL,  AST=20 mg/dL </td>`
		`</tr>`
		`</tbody>`
		`</text>`
		`<component>`
		`<observation classCode="OBS" moodCode="EVN">`
		`<templateId root="2.16.840.1.113883.3.88.11.83.15" assigningAuthorityName="HITSP C83" />`
		`<code code="24323-8" displayName="Comprehensive Metabolic Panel" codeSystemName="LOINC" />`
		`<text> <reference value="#result-4" />  </text>`
		`<value xsi:type="PQ" value="16" unit="mg/dL" />`
		`</observation>`
		`</component>`
		`<component>`
		`<observation classCode="OBS" moodCode="EVN">`
		`<templateId root="2.16.840.1.113883.3.88.11.83.15" assigningAuthorityName="HITSP C83" />`
		`<code code="24323-8" displayName="Comprehensive Metabolic Panel" codeSystemName="LOINC" />`
		`<text> <reference value="#result-4" /> </text>`
		`<value xsi:type="PQ" value="20" unit="mg/dL" />`
		`</observation>`
		`</component>`

Data Redundancy & Conflict	An incomplete mapping of internal code sets from NDC to RxNorm can exist in medications. Consequently, generic information is set to a null value, shown in yellow, while branded data are codified, shown in blue. Redundancy is seen in the multiple template root identifiers, shown in green.	`<consumable>`
		`<manufacturedProduct>`
		`<templateId root="2.16.840.1.113883.3.88.11.83.8.2" assigningAuthorityName="HITSP C83" />`
		`<templateId root="2.16.840.1.113883.10.20.1.53" assigningAuthorityName="CCD" />`
		`<templateId root="1.3.6.1.4.1.19376.1.5.3.1.4.7.2" assigningAuthorityName="IHE PCC" />`
		`<manufacturedMaterial>`
		`<code xsi:type="CE" nullFlavor="UNK">`
		`<originalText>Acetaminophen 650 mg; po q 6h prn<reference value="#med_" /></originalText>`
		`<translation code="00364724812" displayName="Acetaminophen 650 mg" codeSystemName="NDC" />`
		`</code>`
		`</manufacturedMaterial>`
		`</manufacturedProduct>`
		`</consumable>`

* Portions of XML have been excerpted and codeSystem tags translated into names to simplify display.

Potential Quality Metrics Based on Parsed Clinical Modules

Our analysis showed that 12 (27%) of the ambulatory quality measures included in Stage 1 of 'Meaningful Use' were calculable based on patient information and the four clinical content modules parsed in this research (Table 4). While this may appear low, 35 (80%) of all measures could be calculated by incrementally adding the procedures module to the sections in this research. The remaining data which would be required to calculate all quality metrics are vital signs, smoking status, patient communication and communication between providers.

While only a fraction of the quality measures were calculable based on the information in the CCD as currently constructed for Stage 1 of 'Meaningful Use,' important clinical questions could be answered with available data. Several queries proposed in the research include the relationship between problems and the utilization of therapeutic medications, the relation between medications and corresponding laboratory results over time and prescription of medications that may evoke an allergic response based on the known allergens of a patient. All of these theoretical queries could be posed to a relational database based on information extracted from the CCD.

Table 4: Ambulatory Quality Measures for Stage 1 of 'Meaningful Use' Possible with Proposed CCD Parsing.

Quality Group	Problem	Medications	Results	Other Required Modules	Possible Based on Developed Parser
Asthma (NQF 0001, 0036 , 0047)	✓	✓		Procedures	2 of 3
Cancer (NQF 0385, 0387, 0389)	✓	✓	✓	Procedures	0 of 3
Cardiac (NQF 0013, 0018, 0067, 0068, 0070, 0073, 0074, 0075, 0081, 0083, 0084)	✓	✓	✓	Procedures, Vitals	2 of 11
Diabetes (NQF 0055, 0056, 0059, 0061, 0062, 0064, 0088, 0089, 0575)	✓	✓	✓	Procedures, Vitals, Other	3 of 9
Mental Health & Substance Abuse (NQF 0004, 0105)	✓	✓		Procedures	1 of 2
Pediatric (NQF 0002, 0024, 0038)	✓	✓	✓	Allergies, Procedures, Vitals, Other	1 of 3
Prenatal (NQF 0012, 0014)	✓	✓	✓	Procedures	0 of 2
Wellness (NQF 0027, 0028, 0031, 0032, 0034, 0041, 0043, 0421)	✓		✓	Procedures, Vitals, Other	3 of 8
Uncategorized NQF 0033 Chlamydia Screening NQF 0052 Lower Back Pain NQF 0086 Glaucoma Evaluation	✓ ✓ ✓		✓	Procedures Procedures Procedures	0 of 3

Discussion

Many organizations and developers are continuing work on improving EHR interoperability and quality measurement, but Stage 1 of 'Meaningful Use' provides a waypoint to evaluate progress. Thousands of eligible professionals and hospitals are expected to install certified EHRs and achieve Stage 1 goals before the end of 2012. This research provides real-world evidence that four major clinical domains can be extracted and aggregated from different EHRs for quality measurement. This aggregation is particularly promising for two reasons.

First, certified EHRs using CCDs provide a foundation for care teams to share information both during patient handoffs and retrospectively to evaluate performance and improvement opportunities. With tools for CCD-based data aggregation in place, one can envision a clinically integrated organization surveying populations to determine what therapeutic regimens are working best and when patients need additional support through disease management. While such analytics are possible today using older standards, they often require expensive interfaces with sophisticated vocabulary normalization or consolidation onto a single EHR across providers. A CCD-based clinical data warehouse would require only a shared, secure server for integrated care teams to extract XML documents and appropriate parsing and aggregation tools to analyze performance.

Second, the CCD as implemented in the C32 standard has clear potential for growth. This research limited its examined clinical modules to the four with clear vocabularies and testing procedures for Stage 1 of 'Meaningful Use.' As more standard vocabularies are endorsed for clinical modules of the CCD, it is clear that further clinical analytics and population management will be possible.

The potential for CCDs as a normalizing force for clinical data has also been identified and funded from the Office of the National Coordinator for Health Information Technology. The popHealth project is an open source reference implementation software service that enables quality measurement of ambulatory EHRs by extracting CCDs[19]. The technology has been developed by the non-profit MITRE Corporation (Bedford, MA) and has begun pilots this year in ambulatory practices. As demonstrated in this research, CCDs can provide needed data for established quality measurement, although the inclusion of procedures remains a key element missing from requirements for Stage 1 of 'Meaningful Use.'

The technology developed in this research can be deployed on shared networks among clinicians with different EHRs. This offers a low-cost option for organizations adopting emerging models of care that require internal data

sharing and quality measurement[12]. Moreover, centralized aggregation and analysis at the provider level reduces the competitive threats and privacy concerns of data sharing with external parties[13]. The analysis of normalized data via CCD aggregation also has application to payer, public health and research organizations. While sharing of individual medical data typically requires patient consent, aggregate analytics can be advanced through distributed queries of CCD-based databases. This could, for example, provide a uniform method to evaluate practices for prospective clinical trials while avoiding the exchange of protected health information. While outside the scope of this research, future models for health information exchanges could also be potential aggregators of CCD-based clinical content[20]. Further research is required to evaluate how the CCD usage will evolve with patient privacy, public health needs and the evolving role for health information exchange.

Despite the potential highlighted through this research, many challenges were identified for using the CCD in a codified fashion for data aggregation. The largest concerns for CCD use for analytics relate to inconsistent data representation, ambiguous relations between data elements and the optionality of key data. Several of the specific issues identified in this research have already been recognized by standards organizations and are included as notes in the sample CCDs for EHR developers[17]. One likely source of these concerns is the multitude of documents, organizations and technical frameworks that are referenced throughout the revision history of HITSP C32 construct. Research and development on CCD interoperability would benefit from a single document including all standards, structures and vocabularies without reference to external sources. Another source for these challenges is likely the limited number of CCD examples that have been made available for the EHR development community to review.

Other issues regarding data redundancy and multiple vocabularies are more amendable to software solutions. Until redundancy has been eliminated in the source CCD standards, parsing technologies should be programmed with logic to search for multiple acceptable formats. Once data are entered into a relational structure, database rules and business query logic can be used to accommodate multiple vocabularies. For example, all the problem codes for a particular diagnosis can be searched in both SNOMED-CT and ICD-9 to accommodate different terminologies. In addition, database transformation logic can normalize alternative spellings, such as 'tab(s)' and 'tablets,' and populate omissions when deduced from other data. The downsides of post-processing logic are added expense to the development of clinical warehouses and a less compelling use case for open source software.

Many limitations of this research are based in the collection methodology for sample CCDs. Participants willing to submit sample CCDs may not be representative of the EHR development community at large. This has the potential to create selection bias where CCD aggregation from other EHRs would raise significant concerns not observed in this research. In addition, received samples were deliberately fictitious. These CCDs may be significantly less complex than actual patient records, so parsing speed may be overly optimistic. In addition, while HITSP C32 conformant samples were requested in all correspondence with research participants, no elements of the CCD identify the version information of the source EHR. Therefore it was not possible for the researchers to verify that the CCDs were generated from certified technology meeting federal standards for 'Meaningful Use.' Finally, only a fraction of the CCDs were manually reviewed to categorize potential issues for use in an aggregated data store. Consequently, the challenges represented here are unlikely comprehensive of all concerns.

Based on this research's findings, several recommendations can be made for future changes to CCD specifications. First, consolidated guidelines should be available that unify information necessary for CCD development and resolve issues on data representation. Although this document would likely be large, it would be helpful to eliminate cross-document and external organization references. This document should also include graphics to help developers understand section relationships, since tables do not easily convey such complex detail. Second, a large library of fictitious CCDs from multiple EHRs should be made available to the public. Specifically, it would be desirable if every certified EHR had the XML documents submitted as part of its testing procedures released into the public domain. Examples of CCDs are not proprietary instruments and developers will benefit from observing the approaches of different EHRs. Making a large repository available for public review would also yield more findings to what was observed in this research.

Next, the required data and vocabularies for CCDs should be increased. Currently, a number of data elements within the HITSP C32 construct can be optionally omitted. This optionality may represent consideration that not all EHRs natively capture these data, but they are valuable in the application of population health. Key information related to medications, allergies, results and problems should be required if known as part of certification for clinical data exchange. Moreover, the addition of normalized data on procedures, vital signs and communication would substantially increase the capacity of CCDs to measure ambulatory quality. Since CCDs are also a common standard

for acute care settings, it is also recommended that new data relevant to hospitalizations be added to the specification. While not evaluated in this research, inpatient quality metrics often require information on the timing of care and such detail is not present in the C32 construct.

Lastly as future terminologies are incorporated into the CCD, those vocabularies should become available to developers in the medical domain free of charge. While the current required terminologies do not present any concerns, the potential to use CPTs, a proprietary standard of the American Medical Association, as the future vocabulary for procedures raises fair use concerns for technologies that aggregate medical data.

Conclusion

This research takes a novel approach to interoperability by collecting CCDs from various EHRs and analyzing the opportunity and challenges encountered when aggregating data for quality improvement. Through parsing and aggregating these CCDs, it provides tangible evidence that normalized clinical databases are possible using the approved standards for Stage 1 of 'Meaningful Use.' This strategy overcomes many barriers facing providers that seek to improve population health in heterogeneous EHR environments. This research also identifies issues in CCD interoperability by cataloging concerns from actual samples and provides recommendations for future specification revisions.

Disclosures

Acknowledgements

References

1. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N. Engl. J. Med. 2010 Aug 5;363(6):501-504.

2. Elmendorf DW. Letter to the Honorable Charles B. Rangel. Washington (DC): Congressional Budget Office (US). 2009 Jan 21. Available from cbo.gov/ftpdocs/99xx/doc9966/HITECHRangelLtr.pdf.

3. Ackerman K. Heavy hitters hit HIMSS stages to stump for health IT. iHealthBeat 2011 Feb 24. Available from: ihealthbeat.org/features/2011/heavy-hitters-hit-himss-stages-to-stump-for-health-it.aspx.

4. United States. Department of Health and Human Services. Health information technology: initial set of standards, implementation specifications, and certification criteria for electronic health record technology. Washington (DC): Office of the National Coordinator for Health Information Technology (US). 2010 Jul 13. Regulation Identification Number 0991-AB58.

5. Ferranti JM, Musser RC, Kawamoto K, Hammond WE. The clinical document architecture and the continuity of care record: a critical analysis. J Am Med Inform Assoc. 2006 Jun;13(3):245-252.

6. Phansalkar S, Robinson G, Getty G, Shalaby J, Tao D, Broverman C. Challenges in exchanging medication information: identifying gaps in clinical document exchange and terminology standards. AMIA Annu Symp Proc. 2009;2009:526-530.

7. Simonaitis L, Belsito A, Cravens G, Shen C, Overhage JM. Continuity of Care Document (CCD) Enables Delivery of Medication Histories to the Primary Care Clinician. AMIA Annu Symp Proc. 2010;2010:747-751.

8. United States. Health Information Technology Standards Panel (HITSP). HITSP summary documents using HL7 continuity of care document (CCD) component C32. New York (NY): HITSP (US). 2009 Jul 08. Version 2.5.

9. United States. Health Information Technology Standards Panel (HITSP). HITSP clinical document and message terminology component C80. New York (NY): HITSP (US). 2010 Jan 25. Version 2.0.

10. United States. Health Information Technology Standards Panel (HITSP). New York (NY): HITSP CDA content modules component C83. HITSP (US). 2010 Jan 25. Version 2.0.

11. United States. National Institute for Standards and Technology (NIST). Approved test procedures for exchange of clinical information and patient summary record, 170.304(i) and 170.306(f). Gaithersburg (MD): NIST (US). 2010 Dec 3. Version 1.1. Available from: healthcare.nist.gov/use_testing/effective_requirements.html.

12. Rittenhouse DR, Shortell SM, Fisher ES. Primary care and accountable care--two essential elements of delivery-system reform. N. Engl. J. Med. 2009 Dec 10;361(24):2301-2303.

13. Grossman JK, Kushner KL, November EA. Creating sustainable local health information exchanges: can barriers to stakeholder participation be overcome? Research brief no. 2, February 2008. Available from: hschange.com/CONTENT/970/970.pdf.

14. Mostashari F, Tripathi M, Kendall M. A tale of two large community electronic health record extension projects. Health Aff (Millwood). 2009 Apr;28(2):345-356.

15. Nass SJ, Levit LA, Gostin LO, Institute of Medicine (US). Beyond the HIPAA privacy rule: enhancing privacy, improving health through research. Washington D.C.: National Academies Press; 2009.

16. United States. Department of Health and Human Services. EHR incentive program electronic specifications: eligible professional clinical quality measures. Washington (DC): Centers for Medicare and Medicaid Services (US). Available from: cms.gov/QualityMeasures/03_ElectronicSpecifications.asp.

17. United States. National Institute for Standard and Technology. CDA guideline validation documents. Gaithersburg (MD): NIST (US). 2010 Oct 13. Available from: xreg2.nist.gov/cda-validation/mu.html.

18. Schadow G, McDonald CJ. The unified code for units of measure [Internet]. Indianapolis (IN): Regenstrief Institute (US). 2009 [cited 2011 Mar 15]. Version 1.8.2. Available from aurora.regenstrief.org/~ucum/ucum.html.

19. popHealth. An open source population health reporting prototype [Internet]. 2011 [updated 2011 Jan 27; cited 2011 Mar 15]. Available from: projectpophealth.org/index.html.

20. Sittig DF, Joe JC. Toward a statewide health information technology center (abbreviated version). South. Med. J. 2010 Nov;103(11):1111-1114.