Enhancing Electronic Health Records to Support Clinical Research

David K. Vawdrey, PhD1, Chunhua Weng, PhD1,
David Herion, MD2, James J. Cimino, MD3,1

1Columbia University Department of Biomedical Informatics, New York, NY;
2Department of Clinical Research Informatics, NIH Clinical Center, Bethesda, MD;
3Laboratory for Informatics Development, NIH Clinical Center, Bethesda, MD

Abstract

The “Learning Health System” has been described as an environment that drives research and innovation as a natural outgrowth of patient care. Electronic health records (EHRs) are necessary to enable the Learning Health System; however, a source of frustration is that current systems fail to adequately support research needs. We propose a model for enhancing EHRs to collect structured and standards-based clinical research data during clinical encounters that promotes efficiency and computational reuse of quality data for both care and research. The model integrates Common Data Elements (CDEs) for clinical research into existing clinical documentation workflows, leveraging executable documentation guidance within the EHR to support coordinated, standardized data collection for both patient care and clinical research.

Introduction

The separation of research from patient care processes has long been a barrier to achieving the goals of the “Learning Health System” and makes clinical research unnecessarily time-consuming and expensive. The Institute of Medicine called for increased attention to this problem in Crossing the Quality Chasm, and the National Research Council lamented that IT-related activities of health professionals are “rarely well integrated into clinical practice,” and that health IT is “rarely used to link clinical care and research.”

Electronic health records (EHRs) have long been viewed as a catalyst to expedite clinical research. Multiple institutions, such as Kaiser, the VA, Partners HealthCare, and Intermountain Healthcare, have been using EHRs to support clinical research for decades. Mayo Clinic conducts more than 4,000 clinical trials each year, and nearly every trial relies on EHR information. The Cleveland Clinic has been using their EHR for trial recruitment, as has Stanford University. The University of Texas MD Anderson Cancer Center developed ClinicStation that presents integrated views of data from both patient care and research. Despite these examples, clinical research is often poorly integrated with clinical care. Poor integration results in unnecessary duplication of work and limits learning from clinical practice.

Besides using an EHR system, many hospitals with large clinical research programs have implemented clinical trial management systems (CTMS), which maintain administrative and clinical information of research participants and are usually disconnected from EHRs. The design requirements for CTMS and EHR systems differ significantly, especially with respect to their data models. While EHRs are oriented to single-patient, unplanned care-related tasks, CTMS tools are designed to support protocol-based research tasks. Another distinction is that EHRs typically contain information obtained through what may be considered “routine data collection” (i.e., data collected in the process of providing clinical care), while CTMSs may require “specialized data collection” (i.e., data collected with finer granularity or more precision). For example, a routine blood pressure measurement recorded in the EHR in a hospital may not be suitable for inclusion in a clinical research trial dataset because the patient’s body position (e.g., sitting, standing, supine) was not documented. In this case, a research nurse would be obliged to take a separate blood pressure measurement, recording the value and the body position in the CTMS or directly in a research case report form (CRF). Similar examples of redundancy in clinical and research tasks can be seen with ordering of laboratory and imaging tests.

Several groups have published on the barriers to integrating clinical and research information systems. To address some of these issues, interoperability standards have been developed for exchanging data between clinical and research systems. The Clinical Data Interchange Standards Consortium (CDISC) has established global, vendor-neutral, and freely available standards to support the acquisition, exchange, submission and archive of clinical research data and
Partnering with CDISC, members of the Integrating the Healthcare Enterprise (IHE) initiative have labored to link EHR and clinical research systems through efforts such as the Retrieve Form for Data Capture Profile (RFD), which allows clinical trial forms to be embedded in EHRs and pre-populated with certain data. With RFD, no data are retained in the EHR, which is a significant drawback if the data are useful for clinical as well as research purposes. Overall, adoption of clinical research data exchange standards remains low. According to CDISC, which counts over 200 organizations in its worldwide membership, barriers to adoption include a lack of understanding of the relevant standards, the cost of implementation, and the lack of data for exchange.18

Many clinical research institutions have adopted Vanderbilt University’s Research Electronic Data Capture (REDCap) software, or similar electronic data capture (EDC) systems to store research data. In some cases, a CTMS system is used for research subject tracking, billing, and visit/procedure scheduling, and a separate EDC system is used to store research data. While these systems have been thoughtfully designed to accommodate a variety of data import and export options, extracting EHR data from proprietary vendor data models and synchronizing information across multiple systems remains challenging. In the end, far too much effort is spent working around the limitations of EHRs as opposed to addressing the underlying challenges.

In this paper, we propose a model for enhancing EHRs to collect structured and standards-based clinical research data during clinical encounters that promotes efficiency and computational reuse of quality data for both care and research. The model integrates Common Data Elements (CDEs) for clinical research into existing clinical documentation workflows, leveraging executable documentation guidance within the EHR to support coordinated, standardized data collection for both patient care and clinical research.

Methods

Process and Limitations of Electronic Documentation

Spurred by government financial incentives, the United States is experiencing unprecedented adoption of EHRs, including increasing use of electronic clinical documentation. Electronic documentation improves legibility and availability of notes, and it facilitates the collection of structured data for purposes such as quality improvement and research. However, implementing electronic documentation has been reported to adversely impact clinicians’ perceptions of documentation quality, workflow, professional communication, and patient care.19-22

The question of “What should be documented in the EHR?” is relevant and timely. The 2011 AMIA Invitational Health Policy Meeting addressed the current and future state of technology-enabled clinical data capture and documentation, and in February 2013, the Office of the National Coordinator for Health Information Technology (ONC) HIT Policy Committee’s Meaningful Use and Certification and Adoption workgroups held hearings focused on clinical documentation functionality in EHRs and its effect on the delivery of high quality clinical care and provider efficiency and collaboration. Even after decades of experience with EHRs, electronic notes continue to be cluttered and redundant, making it difficult for clinicians to understand the actions and thought processes of their colleagues.19

Our attempt to enhance EHR documentation capabilities to support clinical research acknowledges that the primary purpose of clinician documentation must be to support patient care.

Common Data Elements (CDE)

Frequently, data collection forms used in clinical research contain fields with inadequate definitions and idiosyncratic permissible values.23 Common Data Elements (CDEs) have been developed with the goal of reducing the time and effort spent by researchers deciding what data to collect for a clinical trial, as well as increasing the interoperability of data collected by various groups. An example of a CDE is shown in Figure 1. CDEs are defined in detail using a
metadata dictionary and can be shared in a standardized format across multiple institutions. Our model for enhancing the EHR to support clinical research integrates CDEs with current electronic documentation workflows.

The use of CDEs is a growing trend, although to date, adoption has occurred on a relatively small scale—most commonly in cancer research. The National Cancer Institute (NCI)'s Cancer Data Standards Registry and Repository (caDSR) supports development and deployment of CDEs in cancer research and provides a web-based CDE Browser and application programming interface for public use. CDEs have also been used in epilepsy research, posttraumatic stress disorder research, traumatic brain injury, and substance use disorder. Recently, the National Institute of Neurological Disorders and Stroke (NINDS) has “strongly encouraged” investigators of Phase III and exploratory clinical trials to use CDEs. Additionally, the National Library of Medicine has become an important participant in curating CDEs for clinical research and recently developed the CDE Resource Portal (http://www.nlm.nih.gov/cde/). In the 2013 AMIA Joint Summit meeting, Lin et al. described a method for mapping the clinically-oriented Common Element Model to research variables in dbGaP. The work presented a useful taxonomy of contextual information to be recorded when collecting research data. Our model goes beyond definition to represent clinical workflows, and to create an environment to collect research data during clinical encounters.

**Integrating CDEs with Documentation Workflows**

Figure 2 presents a model that integrates CDEs with existing documentation workflows to improve the process of collecting data for clinical research. The model, which emphasizes clinician and researcher data needs and documentation processes, is informed by the conceptual framework for clinical research informatics proposed by Kahn and Weng. Our model consists of an informatics-enabled clinical research workflow, where providers or clinicians can access a library of disease-specific CDEs and perform CDE-based structured data collection using smart templates. Documentation decision support can guide clinicians in capturing research-quality data. In this implementation, the EHR plays a dual role for both patient care and clinical research and facilitates the interoperability of the processes of both missions.

**Implementation**

The following scenario highlights the EHR's potential to enhance clinical research, as well as the challenges that may be overcome by our proposed model for enhancing EHR documentation to support clinical research.

Ms. Johnson is a 52-year-old woman who arrives in the emergency department complaining of abdominal pain, fatigue, and jaundice. Reviewing the results of her blood work, physicians discover that Ms. Johnson has tested positive for hepatitis C. She is admitted to the hospital. During the intake process, Ms. Johnson's nurse asks about...
her health history, including her smoking status. Ms. Johnson quit smoking 3 years ago when her granddaughter was born, but prior to that time she had smoked approximately half-a-pack of cigarettes per day since she was a teenager.

Two years before Ms. Johnson’s hospital visit, the institution had implemented a certified EHR and attested to its “meaningful use” of the system, qualifying to receive federal incentive payments under the HITECH act. As part of its EHR/meaningful use implementation, the hospital configured a structured “Nursing Admission History” documentation template. The template contained check-boxes to record smoking status, with options such as “Never,” “Current,” and “Former.”

Consulting with her physicians and family members, Ms. Johnson elects to enroll in a phase 3 clinical trial of a new Hepatitis-C medication called sofoviran. The clinical trial’s purpose is to confirm the effectiveness and safety of the drug, and the clinical trial sponsor agency is interested in collecting detailed information of several types, including the health history of trial participants as well as any adverse events they experience—such as headaches or chest pain—that could be caused by the medication.

One of the data elements that must be recorded in the pharmaceutical company’s case report form (CRF) is the study participant’s smoking history. The CRF specifically requires documentation about the level of cigarette consumption (i.e., packs-per-day) for current and former smokers. Because this level of granularity is not captured in the EHR, a research nurse must re-ask the patient about her smoking habits.

Several hours after Ms. Johnson receives her first dose of sofoviran, she develops a severe headache. The headache is a possible adverse event of the medication, and should be recorded on the CRF. She describes the headache to her physician during evening rounds, and the doctor informally notes the pain in his free-text assessment/plan without identifying the possible connection to the medication.

Aspects of this scenario are probably familiar to clinicians and research investigators in a variety of environments. Applying the model in Figure 2 can expose the overlap between data elements collected during routine patient care and data elements that are captured as part of a specific research protocol. Current documentation workflows can then be augmented by mapping EHR data fields to CDEs. For example, referring to the above scenario, smoking status recorded for every hospital patient can be encoded in a computationally reusable format such as the CDE for “Cigarette Consumption Daily Count” shown in Figure 1.

Moreover, the proposed model will fuse documentation workflows with awareness of clinical research protocol documentation requirements such as recording of adverse events. Applying the model to the scenario above, a decision support algorithm could prompt the physician during the note-writing process to report possible adverse events using standard definitions. If an adverse event is identified, it can be coded using a CDE and appropriately communicated to the trial sponsor and other stakeholders. For serious adverse events, CDE concepts can be leveraged to generate the necessary codes, forms, and messages (e.g., MEDWATCH, ICH E2B, ICSR) for transmission to systems such as the FDA’s Adverse Event Reporting System (FAERS). Similarly, the decision support system can provide the EHR with temporal context that is crucial for most types of clinical research (e.g., alerting the clinician that a certain panel of laboratory tests must be performed during the third week of a protocol, or allowing the clinician to tag the test results as being the “week 3” results).

Significant effort will be required to fully implement the proposed model for improving EHR documentation processes; however, institutions with certified EHR systems are much closer to achieving the vision of a learning health system than they were just a few years ago. Figure 3 illustrates how the data collection model can be encoded in a terminology management system as a set of concepts, including CDEs (with their various attributes), EHR observations that correspond to particular CDEs (such as “smoking history CDE”), EHR clinical documents (such as “Sofoviran Admission Note”), and clinical trial

Figure 3. Representation of relationships between clinical research and EHR documentation concepts for a hypothetical “Sofoviran” clinical trial protocol.
protocols (such as “Sofoviran Protocol”). We intend to implement this model both in the Research Entities Dictionary\textsuperscript{36} at the NIH Clinical Center and the Medical Entities Dictionary at Columbia University Medical Center.\textsuperscript{37}

High-level classes can be defined to represent concepts that need to be mapped between the clinical and research realms—not only data definitions, but also the data workflows and context necessary for computational reuse. For EHR concepts, classes include the data source and the EHR data context. Intermediate-level concepts within those classes include 1) for data source: clinical narrative note, clinical structured template, clinical flow sheet, laboratory test, registration data; and 2) for EHR data context: time of data collection, visit the data are attached to, status (final, preliminary), linked clinical order. For clinical research concepts, classes include the data type hierarchy and the research data context. Intermediate-level concepts may include symptoms, signs, laboratory tests, diagnoses, and procedures. The research data context includes concepts that define the constraints on research data collection that are usually executed by the research staff that carry out a trial. Examples include the time that data are collected (either absolute time with respect to entering a clinical trial or relative to other trial events), allowable sources (e.g., only values measured in a special laboratory, or only diagnoses confirmed by a physician), and other constraints (e.g., measurements taken after a meal).

In our model, documentation decision support can be encoded as an open source set of computer-interpretable process rules for coordinating clinical care and research workflows to facilitate knowledge sharing. The result of a rule firing can be the automatic addition of data elements to a template that is about to be used, a message to a user, or some other type of decision support. Given the similarity between clinical guidelines and process rules guidelines, it makes sense to leverage existing standards for clinical guidelines to formalize the process rules. Many languages have been developed to represent and share formal knowledge of research protocols or clinical guidelines, such as the Arden Syntax,\textsuperscript{38} GELLO,\textsuperscript{39} PROforma,\textsuperscript{40} EON,\textsuperscript{41} GLIF,\textsuperscript{42} and SAGE.\textsuperscript{43}

Discussion

There are compelling arguments for integration of patient care and clinical research, both in terms of workflow processes and electronic systems.\textsuperscript{44,45} A recent decision support panel identified four areas where advances in decision support lie: the state of the knowledge base (the set of rules, content, and workflow opportunities for intervention); necessary database elements to support decision support functions; operational features to promote usability and to measure performance; and organizational structures to help manage and govern current and new decision support interventions.\textsuperscript{46} The panel’s findings stress the central importance for decision support functions and workflow changes to be mutually supportive to each other so that decision support facilitates workflow changes and relies on workflow support and integration. Mandl and Kohane emphasized the value of flexibility in healthcare system design, arguing that “system[s] will have to function under evolving policies and in the service of new health care delivery mechanisms...and emerging information technologies.”\textsuperscript{47} Their SMART platform enables lightweight, modular “apps” to be integrated with EHRs, overcoming the proprietary “silos” that exist in current systems.\textsuperscript{48} As this system architecture paradigm gains momentum, EHR implementers will be increasingly in a position where the ‘right choice’ in terms of designing data collection forms is also the ‘easy choice’—flexible and efficient user interfaces will enable clinicians to capture discrete, coded data that are computationally reusable.

Our proposed model for enhancing EHRs to support clinical research builds on the foundation of CDE standards, bridging the adoption gap by incorporating them directly into electronic documentation tools in the EHR. The model facilitates reuse of routinely collected data and seamless inclusion of data capture specific to a patient’s research studies while minimizing the impact on clinician effort. The model is consistent with next-generation EHR architectures such as the SMART platform, enabling documentation decision support within the EHR to support coordinated, standardized data collection for both patient care and clinical research.

Conclusion

The clinical research informatics community has emphasized the need for innovative information technology to support clinical and clinical research processes; however, the complexity of the patient care and clinical research environments makes coordination among the multiple stakeholders difficult to achieve. We propose a model for enhancing EHRs to collect structured and standards-based clinical research data during clinical encounters that promotes efficiency and computational reuse of quality data for both care and research. While we believe that the model will be useful in a variety of healthcare delivery settings, further research is warranted to demonstrate its effectiveness.
References

1. Weng C, Appelbaum P, Hripcsak G, et al. Using EHRs to integrate research with patient care: promises and challenges. *J Am Med Inform Assoc.* Sep-Oct 2012;19(5):684-687.
2. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. *Sci Transl Med.* Nov 10 2010;2(57):57cm29.
3. Institute of Medicine (U.S.). Committee on Quality of Health Care in America. *Crossing the quality chasm: a new health system for the 21st century.* Washington, D.C.: National Academy Press; 2001.
4. Stead WW, Lin H, National Research Council (U.S.). Committee on Engaging the Computer Science Research Community in Health Care Informatics. *Computational technology for effective health care: immediate steps and strategic directions.* Washington, D.C.: National Academies Press; 2009.
5. FastCures. Think Research: Using EMR to Bridge Patient Care and Research. 2009; [http://www.fastercures.org/objects/pdfs/white_papers/emr_whitepaper_summary.pdf](http://www.fastercures.org/objects/pdfs/white_papers/emr_whitepaper_summary.pdf).
6. Cimino JJ, McNamara TJ, Meredith T, et al. Evaluation of a proposed method for representing drug terminology. *Proc AMIA Symp.* 1999:47-51.
7. Cimino JJ, Patel VL, Kushniruk AW. Studying the human-computer-terminology interface. *J Am Med Inform Assoc.* Mar-Apr 2001;8(2):163-173.
8. STRIDE; [http://clinicalinformatics.stanford.edu/STRIDE/](http://clinicalinformatics.stanford.edu/STRIDE/). Accessed April 14, 2013.
9. ClinicStation. [http://www.informatics-review.com/wiki/index.php/ClinicStation](http://www.informatics-review.com/wiki/index.php/ClinicStation). Accessed March 16, 2013.
10. *Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions* National Research Council;2009.
11. El Fadly A, Daniel C, Bousquet C, Dart T, Lastic PY, Degoulet P. Electronic Healthcare Record and clinical research in cardiovascular radiology. HL7 CDA and CDISC ODM interoperability. *AMIA Annu Symp Proc.* 2007;216-220.
12. Kuchinke W, Wiegelmann S, Verplancke P, Ohmann C. Extended cooperation in clinical studies through exchange of CDISC metadata between different study software solutions. *Methods Inf Med.* 2006;45(4):441-446.
13. Kuchinke W, Aerts J, Semler SC, Ohmann C. CDISC standard-based electronic archiving of clinical trials. *Methods Inf Med.* 2009;48(5):408-413.
14. Bruland P, Breil B, Fritz F, Dugas M. Interoperability in clinical research: from metadata registries to semantically annotated CDISC ODM. *Stud Health Technol Inform.* 2012;180:564-568.
15. Dworkin RH, Allen R, Kopko S, et al. A standard database format for clinical trials of pain treatments: an ACTTION-CDISC initiative. *Pain.* Jan 2013;154(1):11-14.
16. El Fadly A, Lucas N, Rance B, Verplancke P, Lastic PY, Daniel C. The REUSE project: EHR as single datasource for biomedical research. *Stud Health Technol Inform.* 2010;160(Pt 2):1324-1328.
17. El Fadly A, Rance B, Lucas N, et al. Integrating clinical research with the Healthcare Enterprise: from the RE-USE project to the EHR4CR platform. *J Biomed Inform.* Dec 2011;44 Suppl 1:S94-102.
18. (http://www.cdisc.org/stuff/contentmgr/files/0/fdf8540f5324c81f48d3630923b95fd6/misc/cdigrd_onder_etal_p2.pdf) Accessed March 13, 2013.
19. Embi PJ, Yackel TR, Logan JR, Bowen JL, Cooney TG, Gorman PN. Impacts of computerized physician documentation in a teaching hospital: perceptions of faculty and resident physicians. *J Am Med Inform Assoc.* Jul-Aug 2004;11(4):300-309.
20. Hartzband P, Groopman J. Off the record--avoiding the pitfalls of going electronic. *N Engl J Med.* Apr 17 2008;358(16):1656-1658.
21. Hirschtick RE. A piece of my mind. Copy-and-paste. *JAMA.* May 24 2006;295(20):2335-2336.
22. Weir CR, Hurdle JF, Felgar MA, Hoffman JM, Roth B, Nebeker JR. Direct text entry in electronic progress notes: An evaluation of input errors. *Methods Inf Med.* 2003;42(1):61-67.
23. Grignon ST, Miller K, Marler JR, et al. National Institute of Neurological Disorders and Stroke Common Data Element Project - approach and methods. *Clin Trials.* Jun 2012;9(3):322-329.
24. Jiang G, Solbrig HR, Chute CG. Quality evaluation of value sets from cancer study common data elements using the UMLS semantic groups. *J Am Med Inform Assoc.* Jun 2012;19(6):e129-136.
25. Jiang G, Solbrig HR, Chute CG. Quality evaluation of cancer study Common Data Elements using the UMLS Semantic Network. *J Biomed Inform.* Dec 2011;44 Suppl 1:S78-85.
26. Nadkarni PM, Brandt CA. The Common Data Elements for cancer research: remarks on functions and structure. *Methods Inf Med.* 2006;45(6):594-601.
27. Patel AA, Kajdacsy-Balla A, Berman JJ, et al. The development of common data elements for a multi-institute prostate cancer tissue bank: the Cooperative Prostate Cancer Tissue Resource (CPCTR) experience. *BMC Cancer.* 2005;5:108.

28. Winget MD, Baron JA, Spitz MR, et al. Development of common data elements: the experience of and recommendations from the early detection research network. *Int J Med Inform.* Apr 2003;70(1):41-48.

29. Loring DW, Lowenstein DH, Barbaro NM, et al. Common data elements in epilepsy research: development and implementation of the NINDS epilepsy CDE project. *Epilepsia.* Jun 2011;52(6):1186-1191.

30. Kaloupek DG, Chard KM, Freed MC, et al. Common data elements for posttraumatic stress disorder research. *Arch Phys Med Rehabil.* Nov 2010;91(11):1684-1691.

31. Maas AI, Harrison-Felix CL, Menon D, et al. Common data elements for traumatic brain injury: recommendations from the interagency working group on demographics and clinical assessment. *Arch Phys Med Rehabil.* Nov 2010;91(11):1641-1649.

32. Ghitza UE, Gore-Langton RE, Lindblad R, Shide D, Subramaniam G, Tai B. Common data elements for substance use disorders in electronic health records: the NIDA Clinical Trials Network experience. *Addiction.* Jan 2013;108(1):3-8.

33. [http://c-path.org/Events/data-standards-in-clinical-trials/Creating-Common-Data-Elements-for-Neurologic-Diseases.pdf](http://c-path.org/Events/data-standards-in-clinical-trials/Creating-Common-Data-Elements-for-Neurologic-Diseases.pdf). Accessed March 19, 2013.

34. Lin K, Hsieh A, Farzaneh S, Doan S, Kim H. Standardizing phenotype variables in the database of genotypes and phenotypes (dbGaP) based on information models. Paper presented at: AMIA 2013 Summit on Translational Bioinformatics2013; San Francisco, CA.

35. Kahn MG, Weng C. Clinical research informatics: a conceptual perspective. *J Am Med Inform Assoc.* Jun 2012;19(1e):e36-42.

36. Cimino JJ, Ayres EJ. The clinical research data repository of the US National Institutes of Health. *Stud Health Technol Inform.* 2010;160(Pt 2):1299-1303.

37. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. *J Am Med Inform Assoc.* Jan-Feb 1994;1(1):35-50.

38. Hripcsak G, Ludemann P, Pryor T, Wigertz O, Clayton P. Rationale for the Arden Syntax. *Comput Biomed Res.* 1994;27(4):291-324.

39. Tu S, Musen M. Modeling data and knowledge in the EON guideline architecture. Paper presented at: MedInfo2001.

40. D'Aunno T. Managing the Care of Health and the Cure of Disease: Arguments for the Importance of Integration. *SO - Health Care Management Review Winter 2001;26(1):85-87.

41. Marsolo K. Informatics and operations—let's get integrated. *Journal of the American Medical Informatics Association.* September 1, 2012 2012.

42. Teich JM, Osheroff JA, Piiper EJ, Sittig DF, Jenders RA, the CDSERP. Clinical Decision Support in Electronic Prescribing: Recommendations and an Action Plan: Report of the Joint Clinical Decision Support Workgroup. *J Am Med Inform Assoc.* July 1, 2005 2005;12(4):365-376.

43. Mandl KD, Kohane IS. No Small Change for the Health Information Economy. *N Engl J Med.* March 26, 2009 2009;360(13):1278-1281.

44. Mandl KD, Mandel JC, Murphy SN, et al. The SMART Platform: early experience enabling substitutable applications for electronic health records. *Journal of the American Medical Informatics Association.* July 1, 2012 2012;19(4):597-603.