

EXPERIENCE REPORT

Use of EHRs data for clinical research: Historical progress and current applications

Amy Harris Nordo¹ | Hugh P. Levaux² | Lauren B. Becnel^{3,4} | Jose Galvez⁵ |
Prasanna Rao⁶ | Komathi Stem⁷ | Era Prakash⁸ | Rebecca Daniels Kush⁹

¹Duke Center for Health Informatics, Durham, North Carolina

²Protocol First, Inc., Headquarters, Salt Lake City, Utah

³Clinical Data Interchange Standards Consortium (CDISC), Austin, Texas

⁴Pfizer, New York, USA

⁵National Institutes of Health, Bethesda, Maryland

⁶IBM Watson Health, Durham, North Carolina

⁷monARC Bionetworks, San Francisco, California

⁸Allscripts Inc, Raleigh, North Carolina

⁹Scientific Innovation Officer, Elligo Health Research and President, Catalysis, Austin, TX, USA

Correspondence

Rebecca Kush, Scientific Innovation Officer, Elligo Health Research and President, Catalysis.

Email: rkush@catalysisresearch.com

Abstract

The benefits of reusing EHR data for clinical research studies are numerous. They portend the opportunity to bring new therapies to patients sooner, potentially at a lower cost, and to accelerate learning health cycles—through faster data acquisition in clinical research studies. Metrics have proven that time can be saved, workflow and processes streamlined, and data quality increased significantly. Pilot projects and now actual investigational trials used for regulatory submissions have shown that these benefits support the transformation of clinical research by leveraging EHRs for research. Panelists at a recent collaborative focused on bridging clinical research and clinical care offered varying perspectives on how the latest standards and technologies could be leveraged to facilitate data transfer from EHR systems into clinical research databases, as well as the associated improvements in data quality. Panelists also discussed other avenues to leverage EHR in clinical research. Improvements and exciting possibilities notwithstanding, much work remains. Data ownership and access, attention to metadata and structured data for data sharing, and broader adoption of global standards are key areas for collaboration. With the steady increase in adoption of EHRs around the world, this is an excellent time for all stakeholders to work together and create an environment such that EHRs can be used more readily for research. The capacity for research can thus be increased to provide more high-quality information that will contribute to rapid continuous learning health systems from which all patients can benefit.

KEYWORDS

clinical research, electronic health records, interoperability, SMART on FHIR

1 | INTRODUCTION

The adoption of electronic health records (EHRs) in patient care settings is increasing rapidly throughout the world, yet the use of EHR data in clinical research has lagged. Reuse of health care data directly for research purposes has the potential to bring significant value and

accelerate learning in several key areas: streamlining clinical research processes at health institutions; improving data quality by reducing the number of transcription errors; evaluating the feasibility of research protocols and the availability of patients to participate in research; providing real world evidence; and, last but not least, enhancing drug safety and early identification of safety events.^{1,2} Unfortunately, a number of

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2019 The Authors. Learning Health Systems published by Wiley Periodicals, Inc. on behalf of the University of Michigan

varied barriers, including but not limited to distrust in the integrity of EHR data and lack of interoperability among EHRs and between research and healthcare systems, have thwarted this value generation while standards, technologies, and processes to enable the use of EHRs for research have been available for almost a decade.³ Specifically, reuse of data to autopopulate case report forms and research databases is a rare occurrence.⁴ Various countries have had differing experiences in this regard, primarily based on their national healthcare systems and electronic health record implementations.⁵

The initial primary purpose of EHRs in the United States was to support clinical care, financial billing, and insurance claims. Some EHRs have now begun integrating core workflows for clinical research. Still, sharing of the EHR data for research purposes is challenged by different and inconsistent data standards among organizations using EHRs and EHR vendors, and between clinical care and research entities. Although an increasing number of countries are adopting certain common healthcare standards and codelists, such as the International Classification of Diseases (ICD) codes, remaining differences between nations, organizations, and languages around healthcare standards have slowed EHR interoperability.

Interest in using EHRs as electronic data source (eSource) for traditional regulated randomized clinical trials has been ongoing for more than two decades. eSource is a research-based term that refers to source data in electronic format and includes the reuse of EHR data and a myriad of other electronic sources of data such as patient reported outcomes, diaries, and wearable devices. Beyond traditional randomized trials, regulators encourage the use of eSource and other data sources for the generation of “real-world evidence” (RWE) based upon real-world data (RWD). Use of RWD may reduce the time needed for data collection and reduce the number of patients required for a study (eg, through the use of “synthetic” standard of care control arms), thereby increasing the speed at which new therapies are delivered to patients. Improved transfer of EHR data to various research databases can provide near real time data to sponsors, allowing more rapid identification of potential safety events and improving access to robust longitudinal patient data for surveillance to assess the safety and efficacy of therapies. Guidance documents from the US Food and Drug Administration (FDA) and communications around the use of eSource and specifically EHRs for research underscore Agency's support.⁶⁻⁸

The Learning Health Community (LHC)⁹ was formed to mobilize, inspire, and empower multiple and diverse stakeholders nationally and globally to work collaboratively on some of these issues, with a goal to realize person-centered learning health systems anchored in a set of shared consensus, *Core Values*. The LHC has expressed interest in accelerating the lengthy learning health cycle from healthcare data to research findings that inform clinical decisions—a cycle that has been estimated to take an excessive 17 years and rely on the use of healthcare data for research. Thus far, the LHC has organized two primary initiatives: (1) the Essential Standards to Enable Learning (ESTEL) initiative and (2) the Governance Initiative. A recent LHC consensus-based action plan has been published to address a variety of opportunities to shorten this learning cycle time.¹⁰

A panel of representatives with varying perspectives on this topic of using EHRs for research was convened at a Bridging Clinical

Research and Health Care Collaborative in April 2018. This panel discussed opportunities and impediments to successfully use electronic health records to streamline clinical research and to shorten the learning health cycle. These panel experts also offered their opinions on what it will take to overcome the current barriers and the key areas to be addressed. The perspectives discussed were primarily focused on the United States, although certain insights from Europe and Japan were shared. To maintain coherence, this manuscript focuses on data transfer from EHRs to clinical research databases, but we will also highlight some additional uses of EHRs in research.

2 | BACKGROUND

2.1 | Slow evolution: 1990s to early 2000s

The opportunity to directly use electronic health record data to populate electronic case report forms for research was proposed well before the turn of the century.¹¹ At that time, however, the adoption of electronic health records was inadequate for this vision to become a reality. In 2004, the Office of the National Coordinator (ONC) was established within the US Department of Health and Human Services (HHS). The purpose of the HHS/ONC was to be a “resource to the entire health system to support the adoption of health information technology and the promotion of nationwide health information exchange to improve health care.”¹² The Health Information Technology Standards Panel (HITSP) was a public-private partnership formed in 2005 with a mission “to achieve a widely accepted set of standards to enable widespread interoperability among healthcare software applications.”¹³ Use cases were identified by an oversight committee, the American Health Information Community (AHIC) and “interoperability specifications” were developed for these use cases. AHIC approved “EHRs for clinical research” as the last HITSP Use Case, but they had no further funding to offer to develop this use case. Funding for the research use case was raised from various US federal agencies with the assistance of HHS and through the Clinical Data Interchange Standards Consortium (CDISC) as private organizational contributions that were sent directly to the American National Standards Institute (ANSI) to fund this special HITSP project. The Interoperability Specification (IS #158) was developed for the Use of EHRs for Clinical Research and posted in January 2010.¹⁴ Standard, IS #158, identified three key components: HL7 Consolidated Clinical Document Architecture (CCDA), IHE Retrieve Form for Data Capture (RFD), and CDISC Clinical Data Acquisition Standards Harmonization (CDASH) to support the use of EHRs for research.

Implementations of IS #158 include reporting of incidents of the H1N1 (swine flu) outbreak to the Centers for Disease Control and Prevention (CDC); EHRs populating electronic case report forms (eCRFs) for research, for which several cases have been demonstrated at the HIMSS Interoperability Showcase; and the Adverse Spontaneous Triggered Event Reporting (ASTER) project¹⁵ using EHR data to autopopulate safety reports (MedWatch forms) for pharmacovigilance and regulatory purposes. Initial metrics on the use of RFD and CDASH have been collected and published through the ASTER project and Duke University's Comparative Effectiveness Study.² In the ASTER

use case, the workflow benefits from implementing RFD in safety reporting included: (1) reducing the time for a busy clinician to make a safety report from 35 minutes to less than 1 minute; and, therefore, (2) safety reporting that did not happen at all before increasing significantly. Other RFD implementations focused on leveraging EHRs along with the clinical research standard, CDISC CDASH to autopopulate eCRFs for research studies. The use of CDASH for collecting research data at the start of a research study has the potential to decrease the study start-up time/resources by 70% to 90% and downstream cycle times/resources by 40% to 60%,¹⁶ assuming that the data capture method/system can reuse form builds, edit checks, and other setup features. A more recent study leveraging RFD by Duke with an EHR-electronic data capture (EDC) system combination demonstrated improved data quality²; this work is further discussed below.

In Europe, during the HITSP and HITECH activities in the United States, the Innovative Medicines Initiative (IMI)¹⁷ initiated a 5-year project called EHR4CR. A proposal for “cross border semantic interoperability for learning health systems” was developed in the context of the EHR4CR semantic resources and services in 2016.⁵ This IMI project team held the belief that each patient healthcare interaction provided an opportunity to generate evidence to improve clinical practice. They also expressed an interest in integrating genome, phenome, and exposome data. The project originally identified four use cases, generally addressing protocol feasibility, identification of patients, clinical trial conduct, and safety surveillance. During the allocated 5-year timeframe, the focus was on the first two of these. The EHR4CR project is now continuing within the environment of the European Institute for Innovation through Health Data, I~HD.¹⁸ This Institute was formed as one of the sustainable entities emanating from EHR4CR. This collaboration is one of a number of projects supported by the European Commission. I~HD was established with a goal of scaling up “innovations that rely on high quality and interoperable health data.” Another relevant EU-funded project was TRANSFoRm, which leveraged EHRs to populate case report forms for a study on gastroesophageal reflux disease.¹⁹ In addition to using RFD for this purpose, they also developed a BRIDG-based ontology and a patient-reported outcomes tool that extended the CDISC Operational Data Model (ODM) to function on mobile hand-held devices. The ODM extension is now publicly available through CDISC.²⁰

In Japan, a common storage standard for EHR data (SS-MIX) was leveraged to populate eCRFs for research.²¹ The CDISC ODM standard for data archiving and exchange was implemented by the University of Osaka and also in Europe to transport data from EHRs for research data collection purposes.^{22,23}

In summary, data standards for regulated research have been increasingly harmonized and adopted globally, through efforts by organizations such as the International Council on Harmonization (ICH),²⁴ CDISC²⁵, and Health Level Seven (HL7).²⁶ CDISC, HL7, and others work collaboratively through an HL7 working group now called the Biomedical Research and Regulations (BR&R). These standards development organizations, partnering with the National Cancer Institute (NCI), and the FDA have developed and continue to maintain an information model to bridge research and healthcare—the above-mentioned Biomedical Research Integrated Domain Group (BRIDG) model.²⁷ It is now a research-healthcare standard through the

International Standards Organization (ISO),²⁸ CDISC, and HL7 with a broad scope of protocol-driven research, including genomics. Nevertheless, additional collaborative efforts are needed to harmonize standards around the world for both healthcare and research.

2.2 | New impetus: Late 2000s to present

On the basis of this global progress, it appears that technical capabilities and global research standards to conduct a regulated research study using EHRs have been available at least since 2010, but the actual adoption of such a process by industry has been extremely slow. Better approaches were clearly needed. Progress took place over the last decade from health care authorities, improved standards and renewed focus on data transfer from EHRs into research databases.

HHS/ONC was legislatively mandated in 2009 through the Health Information Technology for Economic and Clinical Health (HITECH act), which provided federal funding incentives for physicians and hospitals for “meaningful use” of EHR systems. HITSP was dissolved when the HIT Standards Committee and the HIT Policy Committee were formed to focus on meaningful use certification criteria and related projects. Meaningful Use is a set of incentives to encourage providers to adopt EHR systems to facilitate (1) patient engagement, (2) health information exchange, and (3) public health and reporting.²⁹

With respect to the standards, whether directly or not, it appears that the landscape within the healthcare arena has been shifting with Meaningful Use. The HL7 Board initiated a Board Task Force (Fresh Start) a couple of years ago to recommend a solution to HL7’s concern with having competing standards (V2, V3, CDA). The recommendation was that HL7 adopt Fast Healthcare Interoperability Resources (FHIR).³⁰ HL7 V2 still remains in wide use and there is limited support for V3, with use primarily centered on the structured product label for FDA. However, the CCDAs, originally named as the standard to be used to comply with Meaningful Use requirements in the United States, is now being replaced by HL7 FHIR,³¹ which has been consistently gaining traction.

Concurrent with the emergence of FHIR was another promising open standard, Substitutable Medical Apps, Reusable Technology (SMART).³² SMART Health IT is an open, standards-based technology platform that enables innovators to create apps that seamlessly and securely run across the healthcare system. SMART was funded initially by the US Government; the Boston Children’s Hospital Computational Health Informatics Program and the Harvard Medical School Department for Biomedical Informatics are nonprofit institutions that lead the SMART program. The SMART Advisory Committee now consists of a number of organizations committed to transforming healthcare and supporting SMART.

Another project called Argonaut was formed by the five largest EHR vendors in the United States with a goal of standardizing the SMART API in HL7 specifications. Thus, “SMART on FHIR” is collaborating on the FHIR “resources,” which are the “building blocks” of FHIR and will use HL7 FHIR as the means to exchange data required by SMART apps. Resources offer¹ a common way to define and represent data,² a common set of metadata and³ a human readable part.³³

FHIR resources are, therefore, the key in retaining the meaning when information is exchanged. Realizing the value and return on investment of a prospective regulated research study, successfully leveraging EHRs for data exchange between healthcare and research via FHIR will rely on FHIR Resources that align with global research standards. These are currently in development, using the BRIDG model (an ISO, HL7, and CDISC standard)²⁷ as the common information model. The exchange of information while retaining meaning is at the heart of semantic interoperability.

While healthcare standards development takes place within standards development organizations (SDOs), other organizations are creating “common data models (CDMs)” and “common data elements (CDEs).” For example, NCI and NLM offer repositories of CDEs.^{34,35} CDMs have been developed by ODHSI (OMOP),³⁶ the Sentinel initiative³⁷ and PCORI (PCORNet).³⁸ While these CDEs and CDMs each serve their given purpose, they are not harmonized among one another³⁹; thus, they add burden to researchers, while making it increasingly difficult to integrate, aggregate, or compare data among these models or studies. A project, funded by the PCOR Trust Fund and led by FDA, is designed to harmonize these CDMs with the BRIDG model and test through a pharmacovigilance use case for RWE. This project was also reported at the Bridging Collaborative.

While the lack of global harmonization and adoption of data standards is an impediment that forms the basis for others such as the lack of interoperability, data integration, efficiency, and processes redesign, it is not the only impediment faced in this desire to bridge research and healthcare. Others include aspects of data privacy, legal agreements, increasing complexity of research requirements, costs associated with data access and integration, and related issues.

2.3 | Panelists

The Panelists consisted of seven participants from various types of organizations: NIH Clinical Center, Duke University, monARC,

Allscripts, IBM Watson Health, CDISC, and Protocol First. The panelists discussed the wide array of stakeholders affected by EHRs for research before addressing specific aspects of leveraging EHRs.

As depicted in Figure 1, the panel was composed of a reasonable representation of the stakeholders interested in the reuse of EHR data for clinical research; however, it was not inclusive of all stakeholders' perspectives. Historically, one of the main challenges to finding a solution for eSource and the use of EHRs for research has been that only a subset of the relevant stakeholders' perspectives was included. The panelists had varied backgrounds and represented perspectives outside of their current organization.

2.4 | Data transfer from EHRs to research databases

A recent study conducted by the Tufts Center for the Study of Drug Development assessed the evolving eClinical Landscape.⁴¹ Out of 257 unique global companies, it was found that all are using electronic data capture; however, there are still over 30% of research studies where data are collected on paper case report forms. Although 50% of the companies anticipate having a strategy for using EHRs by 2020, only 20% are employing them currently. Reentry of data from source medical records is still the norm, and the time from patient visit to reentry of data for research has increased from an average of 6.9 days to 8.1 days over the past decade. Reentry of data not only lengthens the access time but also negatively impacts data quality, as evidenced by the Nordo study (see below).

The Tufts study also indicates that global drug development programs are becoming increasingly complex, the number and breadth of data points collected for each protocol is increasing and the number and variety technology applications employed is growing. Interestingly, this does not appear to be improving efficiency. In fact, the opposite was observed; cycle times such as time to lock a database at the end of a study have lengthened during the past decade. Continuation of

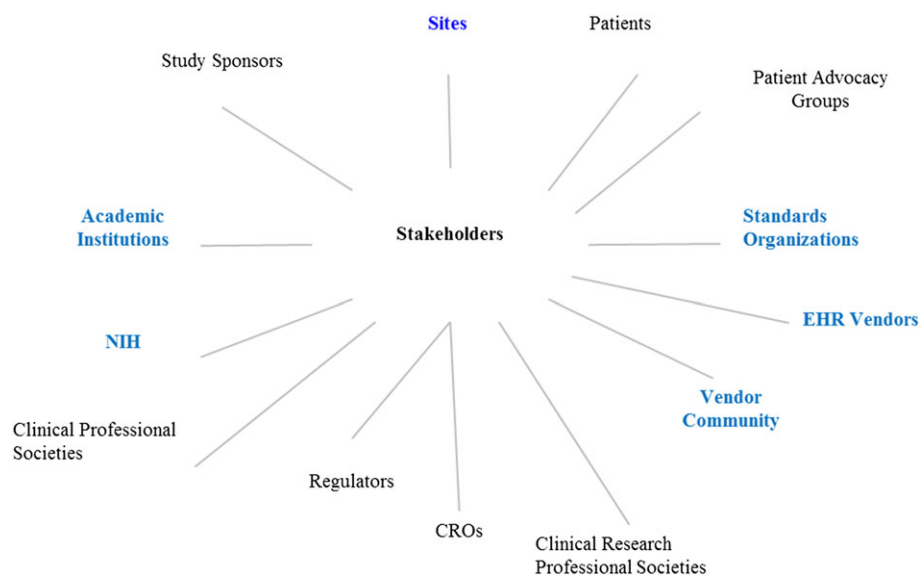


FIGURE 1 List of stakeholders (blue represents panelists) concerned with or affected by issues around the use of EHR data for clinical research.⁴⁰ (blue represents organizations most related to those who participated in the panel discussions)

these trends not only increases the time and cost of developing therapies and burden on sites, but also impedes data sharing efforts.

Fortunately, data transfer from EHRs directly into research databases offers multiple benefits. Two panelists offered direct experience of implementing such solutions.

2.5 | Streamlining clinical research process

Amy Nordo, who worked for Duke University and recently joined Pfizer, elaborated on the recent eSource study she conducted at Duke University.² This pilot was designed to compare eSource-enabled data collection and the traditional manual transcription of data into eCRF based on the RFD standard Table 1 summarizes the results of the eSource pilot. This project has progressed from a single center project into a multistakeholder collaboration between standards organizations, regulatory agencies, vendors, sponsors, and several academic medical centers. The project uses a standards-based approach leveraging HL7 FHIR, CDISC CDASH, and ODM, following the clinical workflow in an effort to move toward “structured reporting.” The two diagrams below show the current data collection process (Figure 2) and the eSource solution data collection process at an investigative site (Figure 3).

2.6 | View on leveraging SMART on FHIR standards by Dr. Hugh Levaux

Dr. Hugh Levaux and his Protocol First team⁴² have developed a product named Clinical Pipe that can facilitate the transfer and mapping of clinical data from EHR to EDC. They have built an EHR application using the standards SMART and HL7 FHIR. Most major EHR systems comply with these standards, but there are still some that do not. The Clinical Pipe application transforms the data to an internal consistent format, which in turn will transform the data into formats such as CDASH and ODM, which EDC systems in clinical research can understand (Figure 4). Not all EDCs have the technical maturity to include a RESTful API, which allows for interoperability with the HL7 FHIR standard. The Clinical Pipe application supports parsing of HL7 V2 messages and can work with EDC that do not conform to the RESTful API, HL7 FHIR, CDASH, or ODM standards. Clinical Pipe can sit on top of existing systems without requiring upgrades of the clinical technology stack for either the site or the sponsor. The application only transfers structured, discreet fields. As such, and depending on the study design, a significant portion of data (approximately 30%-60%) can be transferred versus manually transcribing and verifying data. The application is used in production on a major leukemia program, the Beat AML Master Trial.⁴³ Clinical Pipe's use in a production on a trial is an encouragement that the many factors impacting the reuse

TABLE 1 Data entry time comparison²

Phase	N=	Non-eSource ¹	eSource ¹	Difference (95% CI)	P value ²
Initiation	21	66.3 (50.5)	21.3 (19.6)	45.0 (19.7-70.4)	0.001
Demographic	21	212.5 (49.4)	133.5 (38.1)	79.1 (56.7-101.4)	0.000
Non-eSource	21	1476.1 (406.7)	1447.9 (463.2)	28.2 (-126.6 to 183.1)	0.708
Total time	21	1755.0 (396.5)	1602.6 (470.0)	152.3 (-1.1 to 305.7)	0.051

¹ Mean (Standard deviation)

² Paired samples t-test

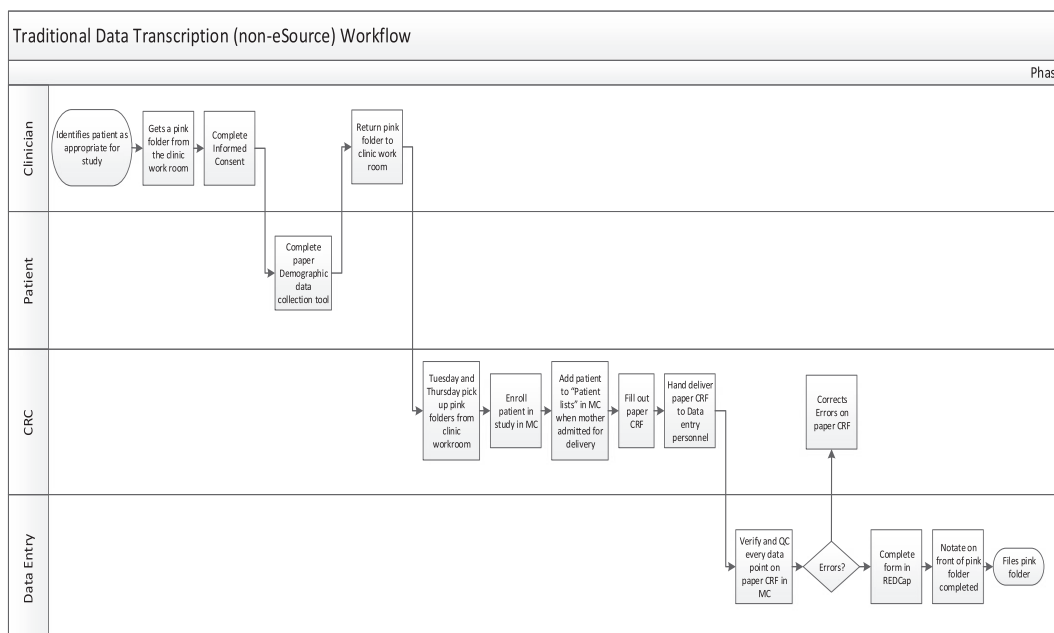


FIGURE 2 An example of traditional data transcription (non-eSource) workflow²

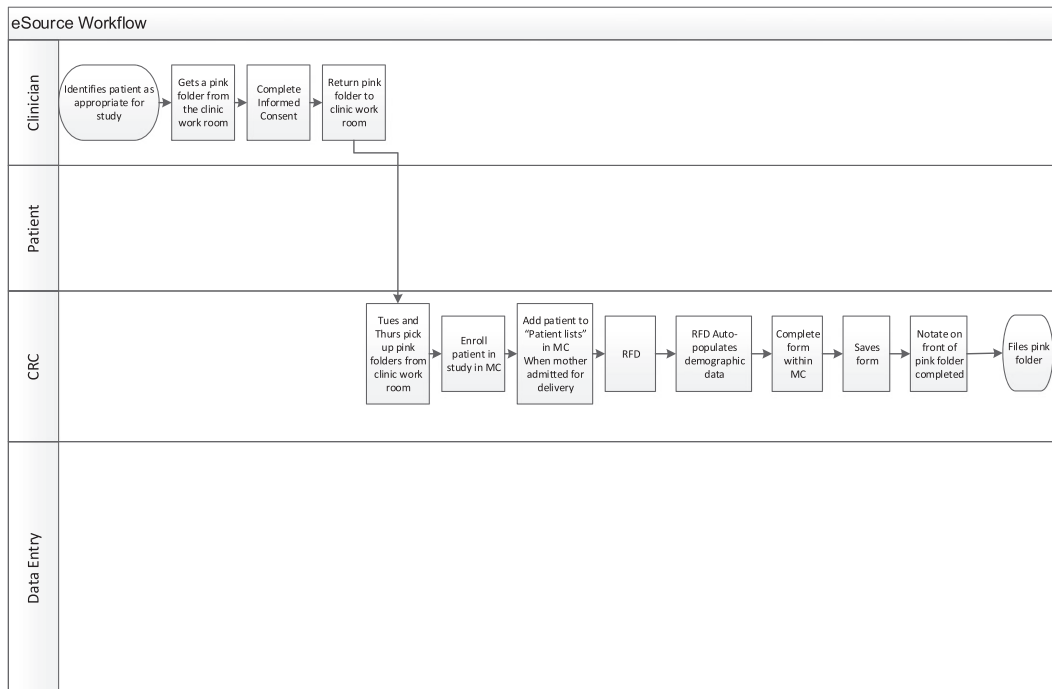


FIGURE 3 An example of eSource workflow²

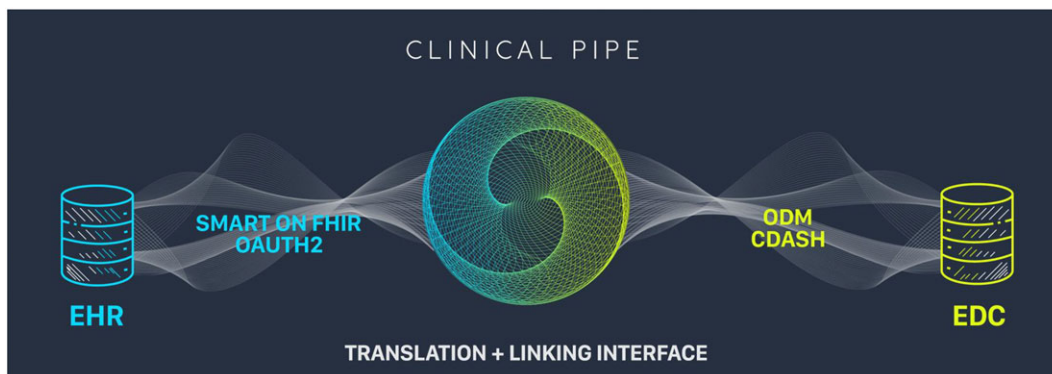


FIGURE 4 A stylized representation of the clinical pipe application: translation + linking interface

of EHR data are being bridged. Successful solutions that support structured data interoperability are not meant to imply that they completely optimize EHR data for clinical research and additional opportunities and challenges exist, including the harvesting of nonstructured data.

2.7 | Reduced resource requirements and improved data quality

The Duke University Comparative Effectiveness Study autopopulated 1.75% of the eCRF data collected, resulting in statistically significant improved data quality and saved time. Times to complete the forms and data quality were assessed. The results indicated that time was reduced by 37% (from 213 s to 133 s) when eSource process was implemented to collect demographic data (Table 1), and the quality of data improved from a 9% error rate to a 0% error rate when eSource was implemented. Scaling these findings to 60% of the eCRF data being

autopopulated would yield an impressive improvement in data collection time and data quality, both of which would positively impact the time necessary to conduct clinical research studies and potentially allow the therapies to come to market faster. Another potential downstream effect of decreased time lapse in data availability to the sponsor⁴⁴ is the earlier detection of adverse events, thereby allowing faster response from the sponsor leading to improved patient safety. Additional research is necessary to prove potential downstream effects.

3 | OTHER APPLICATIONS OF EHRS FOR RESEARCH

3.1 | Real world data

While the above example implementations offered examples pertaining to observational research and investigational trials,

respectively, there is increased interest in leveraging EHRs' RWD to accelerate RWE. Regulators have indicated such interest through a federal requirement for the FDA, within the 21st Century Cures Act,⁴⁵ to increase the use of RWE based upon RWD. The challenges with RWD, however, remain the reliability of the data sources, the quality of the data, the cost, linkage and integration across sources, data consistency through application of appropriate standards, and unclear regulatory guidelines. Several panelists offered examples of their work and interest in this developing field of research.

3.2 | Views on empowering patients by Komathi Stem

Komathi Stem discussed the need for a paradigm shift in the healthcare ecosystem from prioritizing data ownership to data access. monARC⁴⁶ is creating a new world where patients own their data and are fully vested in sharing their data with researchers to accelerate the search for new treatments or cures. She shared how monARC's Smart Health Record engages patients early and continuously to share their EHR data directly with researchers to inform trial designs and prequalify for trials. This direct collaboration with patients has reduced screening time and effort, reduced screen failures, and demonstrated exponential increase in enrollment rates. monARC's Patient Research Networks and Direct-to-Patient Trial platform is also leveraging smartphones, sensors, and EHRs to broaden access for more patients to participate in clinical trials, while also making it simpler for researchers to access real world data for enhanced registries, natural history trials, and pragmatic research studies based on RWD.

4 | FUTURE NEEDS IN LEVERAGING EHRs FOR RESEARCH

4.1 | Views on interoperability and resources by Dr. Lauren Becnel

Dr. Lauren Becnel, who worked for Baylor Medical College prior to joining CDISC and has recently joined Pfizer's Global Real World Evidence team as the oncology lead, recognized the importance of FHIR, which may become the standard way to access healthcare data from EHRs for use in clinical research. CDISC has been actively participating in FHIR Connectathons and updating its ODM to a new version that will provide a translation layer from core FHIR resources, which should be utilized by a multi-stakeholder project currently led by CDISC and Duke University. CDISC participates in the HL7 BR&R working group that helps guide the development of HL7 standards connecting research and healthcare, including the nascent research study and research subject FHIR resources. She stated that exchange formats are still a challenge today. Further, study sites have very limited resources, so the creation of one-off solutions for multiple stakeholders' access to data, where appropriate, should be avoided in favor of standard, interoperable services and tools.

4.2 | Views on standards by Era Prakash

Era Prakash of Allscripts⁴⁷ presented the concept of a *collaborative ecosphere*. The clear message is that EHRs can enable a number of different data sharing opportunities, including with patients themselves, for clinical research, and with providers, regulators, and vendors. She stated that the challenge today is that in each case, the request is for data in a different format. The question remaining to be answered is whether there is an economic incentive for EHRs to provide data for research, especially with such varying requirements.

4.3 | Views on AI, machine learning, and Blockchain by Prasanna Rao

Prasanna Rao, who worked for IBM Watson Health and recently joined Pfizer,⁴⁸ spoke about the work within Watson Health to train machines to look at information and learn—specifically, he mentioned “how to get machines thinking like humans.” The overarching goals of these activities are to transform the way clinical trials are done today, moving away from the existing linear process, and to ensure that the patient is at the center of the process. He also mentioned that Blockchain⁴⁹ is a solid technology that is slowly being adopted as an accepted security method for unlocking patient data for clinical research.

4.4 | Views on data quality and Medidata by Dr. Jose Galvez

Finally, most panelists emphasized the prospect of increased data quality through leveraging EHRs, as well as the inevitable semantic challenges to be tackled. Dr. Jose Galvez spoke about the activities of the National Institutes of Health (NIH) Clinical Center,⁵⁰ which is the largest dedicated research hospital; patients at this center are essentially all research participants. He oversees the NIH Biomedical Translational Research Information System (BTRIS), which brings together clinical research data from the clinical center and other intramural NIH institutes and centers. A key issue with such broad data sharing for any site is how to get contextual information about the data and its subtleties in order to maximize its reuse and give it longevity beyond any one study. For example, the US NCI is engaged in a pilot study assessing to what extent clinical, pathological, and radiographic data and metadata from publicly accessible data repositories can be aggregated, harmonized using global standards such as CDISC and DICOM, to be shared with the community for reuse. The use of EHRs for research will require increased use of standards. Contrary to common thinking, big data actually requires more standards and metadata, not less. Protocols should also be simplified and case report forms should be consistent; creating “works of art” and one-off requirements for data to be collected in each study are a burden and a barrier to streamlining research.

5 | DISCUSSION

Integration challenges include both technical and semantic reasons, and semantic issues have been ascribed to the tendency to overestimate the reliability of the data being exchanged.⁵¹ Challenges in terms of the standards must address harmonization of semantics and will require consensus-building around the use of FHIR resources that support healthcare and research entities if these are to be useful and adopted for research, especially global regulated research. In a blog entitled “Be careful how you fan the FHIRs of interoperability,” it is cautioned that lessons can be learned from clinical imaging standards development, wherein defined clinical protocols and data exchange specifications alone were not sufficient for true interoperability.⁵² Common nomenclatures, definitions, and other metadata were also required for data and images to be reliably and meaningfully shared. Beyond lessons from the domain of clinical imaging, it would also be prudent to learn from the research arena, which has a wealth of experience in the area of provenance, audit trails, traceability, and retaining data integrity used for regulatory purposes.

On the health care side of the ledger, EHRs must substantially increase the ability to manage metadata and allow for creating data sets based less on individual patients and more on their metadata characteristics. As indicated by the panel discussions, improved structured reporting and interoperability necessitate the collaborative adoption and development of harmonized and complementary standards. As in all quality programs, building the standards and quality, the process at the beginning is very important. Implementing standards at the end increases work, resource needs, and cost.

On the research side of the ledger, many research studies are still utilizing traditional data collection methods that are cumbersome and time-consuming, particularly since data are often reentered or transcribed multiple times during the conduct of a study. More pilot and actual implementation studies are needed to learn and improve data transfers from EHRs to research databases.

6 | CONCLUSION

The benefits of reusing EHR data for clinical research studies are numerous. The FDA, HHS/ONC, IHE, CDISC, EU, HL7, IMI, I-HD, Japan's authorities, and others have encouraged the use of EHRs for research based on the reasons provided in this document. The technology to connect EHRs to research databases is available as demonstrated at Duke or with Clinical Pipe. Patients want to share their data as highlighted by monARC. Metrics have proven that time can be saved, workflow and processes streamlined, and data quality increased significantly. Pilot projects and now actual investigational trials used for regulatory submissions have shown that these benefits support the transformation of clinical research by leveraging EHRs. The value propositions extend to the opportunity to bring new therapies to patients sooner, potentially at a lower cost, and learning more rapidly from healthcare information—thereby accelerating learning health cycles.

Unfortunately, while the benefits are numerous, remaining impediments are multi-faceted and will require collaboration among

numerous stakeholders. Broad adoption and harmonization of global standards, along with processes and tools from the recent example of data transfer from EHRs to research databases, will be a starting point to offer new opportunities to overcome the remaining barriers. With the steady increase in adoption of EHRs around the world, this is an excellent time for all clinicians, researchers, and other stakeholders to collaborate. Together, these groups can change the environment such that EHRs can be used more readily for research and the capacity for research can be increased to provide high quality information that will contribute to rapid, continuous learning health systems from which all patients can benefit.

CONFLICT OF INTEREST

The authors confirm that they have no conflicts of interest.

REFERENCES

- Eisenstein EL, Collins R, Cracknell BS, et al. Sensible approaches for reducing clinical trial costs. *Clin Trials*. 2008;5(1):75-84.
- Nordo AH, Eisenstein EL, Hawley J, et al. A comparative effectiveness study of eSource used for data capture for a clinical research registry. *Int J Med Inform*. 2017;103:89-94.
- Kush RD. What the patient should order. *Sci Transl Med*. 21 October 2009;1(3):1-5.
- Lencioni A, Hutchins L, Annis S, et al. An adverse event capture and management system for cancer studies. *BMC Bioinformatics*. 2015; 16(Suppl 13):S6.
- 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*. 2011;44(Suppl 1):S94-S102.
- Guidance for Industry: Source Data in Clinical Investigations (eSource). September 2013. <https://www.fda.gov/downloads/drugs/guidance-compliance/regulatoryinformation/guidances/ucm328691.pdf>.
- Guidance for industry: use of electronic health record data in clinical investigations; July 2018. <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM501068.pdf>.
- Use of real-world evidence to support regulatory decision-making for medical devices; Guidance for Industry and Food and Drug Administration Staff. August 2017. <https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf>.
- Learning Health Community, www.learninghealth.org
- Rubin JC, Silverstein JC, Friedman CP, et al. Transforming the future of health together: the leaning health systems consensus action plan. *Learning Health Systems Journal*. 20 March 2018. <https://doi.org/10.1002/lrh2.10055>
- Kush, RD, Bleicher, P, Kubick, WR, Kush, ST, Marks, R, Raymond, SA, Tardiff, B., *eClinical Trials: Planning and Implementation*, published by Thompson CenterWatch. 2003
- Office of the National Coordinator for Health Information Technology (ONC) Office of the Secretary, United States Department of Health and Human Services. Connecting health and care for the nation: a shared nationwide interoperability roadmap. Department of Health and Human Services. Washington D. C. 2015 <https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf> accessed 1.13.18.
- Health Information Technology Standards Panel (HITSP)- http://hitsp.org/about_hitsp.aspx.
- Health Information Technology Standards Panel (HITSP) http://www.hitsp.org/about_hitsp.aspx.

15. Neuer, A. ASTER study jumpstarts adverse event reporting, Clinical Informatics News, http://www.clinicalinformaticsnews.com/eCliniqua_article.aspx?id=93784, July 2009.
16. Rozwell C, Helton E, Kush R. Saving time and money. *Appl Clin Trials*. 2007;16(6):70-74.
17. Innovative Medicines Initiative (IMI), Europe <https://www.imi.europa.eu/>.
18. Kalra, D., Stroetmann, V., Sundgren, M., Dupont, D., Schlunder, I., Thienpont, G., Coorevits, P., DeMoor, G. The European Institute for Innovation through Health Data (I-HD). *Learn Health System* 2016: 1-8, wileyonlinelibrary.com/journal/lrh2 13 April 2016.
19. Delaney BC, Curcin V, Andreasson A, et al. Translational medicine and patient safety in Europe: TRANSFoRm—architecture for the learning health system in Europe. *Biomed Res Int*. 2015. <https://doi.org/10.1155/2015/961526>
20. ODM Extension <https://www.cdisc.org/standards/foundational/odm/odm-xml-extension-library>.
21. Kimura M, Nakayashu K, Ohshima Y, et al. SS-MIX: a ministry project to promote standardized healthcare information exchange. *Methods Inf Med*. 2011;50(02):131-139. <https://doi.org/10.3414/ME10-01-0015>
22. Matsumura Y, Hattori A, Manabe S, et al. Case report form reporter: A key component for the integration of electronic medical records and the electronic data capture system. MEDINFO 2017: Precision Healthcare through Informatics. Volume 245, 516 – 520. DOI: <https://doi.org/10.3233/978-1-61499-830-3-516>
23. Hume S, Aerts J, Sarnikar S, Huser V. Current applications and future directions for the CDISC operational data model standard: a methodological review. *J Biomed Inform*. 2016;60:352-362.
24. International Council for Harmonization: <http://www.ich.org/home.html>
25. CDISC: www.cdisc.org
26. Health Level Seven: www.hl7.org
27. Becnel LB, Hastak S, Ver Hoef W, et al. BRIDG: a domain information model for translational and clinical protocol-driven research. *J Am Med Inform*. 2017.
28. International Standards Organization (ISO): <https://www.iso.org/home.html>.
29. Meaningful Use: <https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/EHRIncentiveprograms>
30. Fresh Start Recommendation: <http://www.healthintersections.com.au/?p=137>.
31. FHIR Revision Hx: <http://wiki.hl7.org/index.php?title=FHIR&offset=20130130213636&action=history>.
32. SMART: <https://smarthealthit.org/an-app-platform-for-healthcare/about/>.
33. FHIR Resources: <https://www.hl7.org/fhir/resource.html>
34. National Cancer Institute caDSR: <https://wiki.nci.nih.gov/display/caDSR/caDSR+Content>.
35. National Library of Medicine CDE Repository: <https://www.nlm.nih.gov/cde/>.
36. Observational Health Data Sciences and Informatics DHSI -OMOP Model: <https://www.ohdsi.org/>.
37. FDA's Sentinel Initiative. U.S. Food and Drug Administration. [Online] [Cited: February 10, 2018.] <https://www.fda.gov/Safety/FDASentinelInitiative/ucm2007250.htm>.
38. Patient Centered Outcomes Research Institute (PCORI)-PCORNet Model: <http://www.pcornet.org/>.
39. Garza M, Del Fiol G, Tenenbaum J, Walden A, Zozus MN. Evaluating common data models for use with a longitudinal community registry. *J Biomed Inform*. 2016;64:333-341.
40. Nordo A, Eisenstein E, Zozus M. eSource for Regulated Clinical Trials. DataFITS Consortium. (2018,in preparation for publication).
41. Wilkinson M, Young R, Harper B, Machion B, Getz K. Baseline Assessment of the Evolving 2017 eClinical Landscape. *Ther Innov Regul Sci*. 2018;1-10. <https://doi.org/10.1177/2168479018769292>
42. Protocol First: <https://www.protocolfirst.com/>.
43. Beat AML® Master Trial: <https://www.ils.org/beat-aml/beat-aml-master-trial-more-information>.
44. Kiechle M, Paepke S, Shwarz-Boeger U, et al. EHR and EDC Integration in Reality. *Applied Clinical Trials Online*. 2009.
45. Century Cures Act FDA Information: <https://www.fda.gov/regulatory-information/lawsenforcedbyfda/significantamendmentstothehdact/21stcenturycuresact/default.htm>.
46. monARC: <https://www.monarcbio.com/>.
47. Allscripts: <https://www.allscripts.com/>.
48. IBM Watson Health: <https://www.ibm.com/watson/health/>.
49. Baara M, Lipset C, Kudumala A, Fox J, Israel A. Blockchain opportunities for patient data donation and clinical research. Deloitte Development LLC 2018.
50. NIH Clinical Center: <https://clinicalcenter.nih.gov/>.
51. McComb D. *Semantics in Business Systems*. Elsevier: Morgan Kaufmann Publishers; 2004.
52. Marion, Be Careful How You Fan the Flames of FHIR, <https://www.healthcare-informatics.com/blogs/joe-marion/be-careful-how-you-fan-fhir-s-interoperability> (4 April 2018).

How to cite this article: Nordo AH, Levaux HP, Becnel LB, et al. Use of EHRs data for clinical research: Historical progress and current applications. *Learn Health Sys*. 2019;3:e10076. <https://doi.org/10.1002/lrh2.10076>