REVIEW ARTICLE

eSource for Standardized Health Information Exchange in Clinical Research: A Systematic Review of Progress in the Last Year

Maryam Y. Garza[*], Sahiti Myneni[†], Susan Fenton[†] and Meredith Zozus[‡]

Objective: To identify studies conducted using the direct, electronic extraction of electronic health record (EHR) data to electronic data capture (EDC) systems, also known as eSource, and to identify any gaps or limitations present for promoting standardized health information exchange in clinical research.

Materials and Methods: Articles were included only if the solution described (1) utilized eSource to directly exchange data electronically from EHR-to-EDC and (2) was relevant to a prospective clinical study use case.

Results: In total, 20 relevant articles were identified, describing a total of 15 unique eSource interventions. Of the 15 interventions, 12 were single-site, single-EHR (SS-SE) implementations and 3 were multi-site, multi-EHR (MS-ME) implementations. All 15 implementations mentioned the use of standards, but nearly all referenced older data exchange standards.

Discussion: Following the trajectory of work towards direct EHR-to-EDC, eSource data collection, we appear to have arrived at the point where information systems leveraging data standards can offer efficiency and increased quality in clinical research. However, these methods need to be tested for effectiveness and acceptance in the context of real multicenter clinical trials. Several early studies using a single source of data for research and patient care appeared over a decade ago. Since that time, implementations and evaluations have been scarce and almost always confined to single-EHR, single-EDC, single-institution implementations.

Conclusion: These results only further emphasize the observation that the clinical trial use case continues to be the most difficult and least demonstrated eSource-related initiative. Thus, additional work is critically needed in this area to address the gaps identified from the literature.

Keywords: eSource; electronic health records; secondary data use; clinical research; health information exchange

1. Introduction

Since the advent of health information systems and electronic health records (EHRs), clinical researchers have sought to harness the resulting data for use in clinical research.[1,2] The direct use of EHR data, often referred to as 'eSource', has long been an optimistic and highly desired goal in prospective clinical trials[3] because of anticipated increases in data quality and reductions in site burden.[4–7] Over the last decade, sporadic attempts toward this have been reported,[8–10] most of which have been limited to single-EHR, single-EDC (electronic data capture), and single-institution implementations.[11] Successful implementation of a generalizable eSource solution requires (1) use of data standards, (2) process re-design, and (3) the rigorous evaluation of data quality, site effort, cost, and feasibility.

Accordingly, the objective of this manuscript was to build upon our prior work to (1) identify clinical research studies conducted using the direct, electronic extraction of electronic health record (EHR) data to electronic data capture (EDC) systems and (2) identify any gaps or limitations present for promoting standardized health information exchange in clinical research.

2. Materials and Methods

In 2019, Garza and colleagues published the original systematic review of the literature to identify EHR-to-EDC eSource studies.[11] Articles were included only if the solution described (1) utilized eSource to directly exchange data electronically from EHR-to-EDC and (2) was relevant

[*] Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR, US

[†] School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, US

[‡] School of Biomedical Informatics, University of Texas Health Science Center at San Antonio, San Antonio, TX, US

Corresponding author: Maryam Y. Garza , PhD, MPH, MMCi (mygarza@uams.edu)

to a prospective clinical study use case.[11] This resulted in the identification of 14 relevant articles.[8,9,12–23]

Since the publication of the original review, a secondary scan of the literature was conducted to identify eligible articles that had since been published. For this search, the dates of publication were limited to January 2018 through December 2020. The same methods were used to identify, evaluate, and categorize each article as was done in the original review. This resulted in the identification of six additional articles.[24–29] Thus, the final set was comprised of 20 total articles (**Figure 1**).

Each study was categorized according to four critical dimensions, including: (1) whether the study was conducted at a single site or was part of a multi-site study, (2) whether the study utilized a single EHR or multiple EHRs, (3) whether or not the study was conducted as part of an ongoing, prospective clinical study, and (4) whether or not relevant standards were used.[11] It is recommended that readers become familiar with the content of the original review prior to this secondary publication, as it contains important content that will not be repeated here.

3. Results

Across the 14 manuscripts identified as part of the primary review, there were 10 distinct eSource interventions evaluated. Eight of the ten interventions were single-site, single-EHR implementations.[8,9,12–14,16–18] Of the two multi-site, multi-EHR interventions remaining – (1) the European EHR for Clinical Research (EHR4CR) centralized commercial, fee-for-service platform (www.ehr4cr.eu)[19–22,30] and (2) the European TRANSFoRm project (www.transformproject.eu)[23] – only one (TRANSFoRM) was part of an ongoing, prospective study. These approaches all leveraged older data exchange standards and evaluative measures remained inconsistent across studies.[11,31]

Upon review of the six additional manuscripts, there were five distinct eSource interventions evaluated, bringing the total to 15 eSource interventions across all 20 publications. One of the manuscripts[29] described the protocol development process for a future eSource evaluation study, the TransFAIR study, developed as part of the EHR2EDC Project (https://eithealth.eu/project/ehr2edc/). The TransFAIR study is meant to be a proof-of-concept for an EHR-to-EDC eSource technology for use
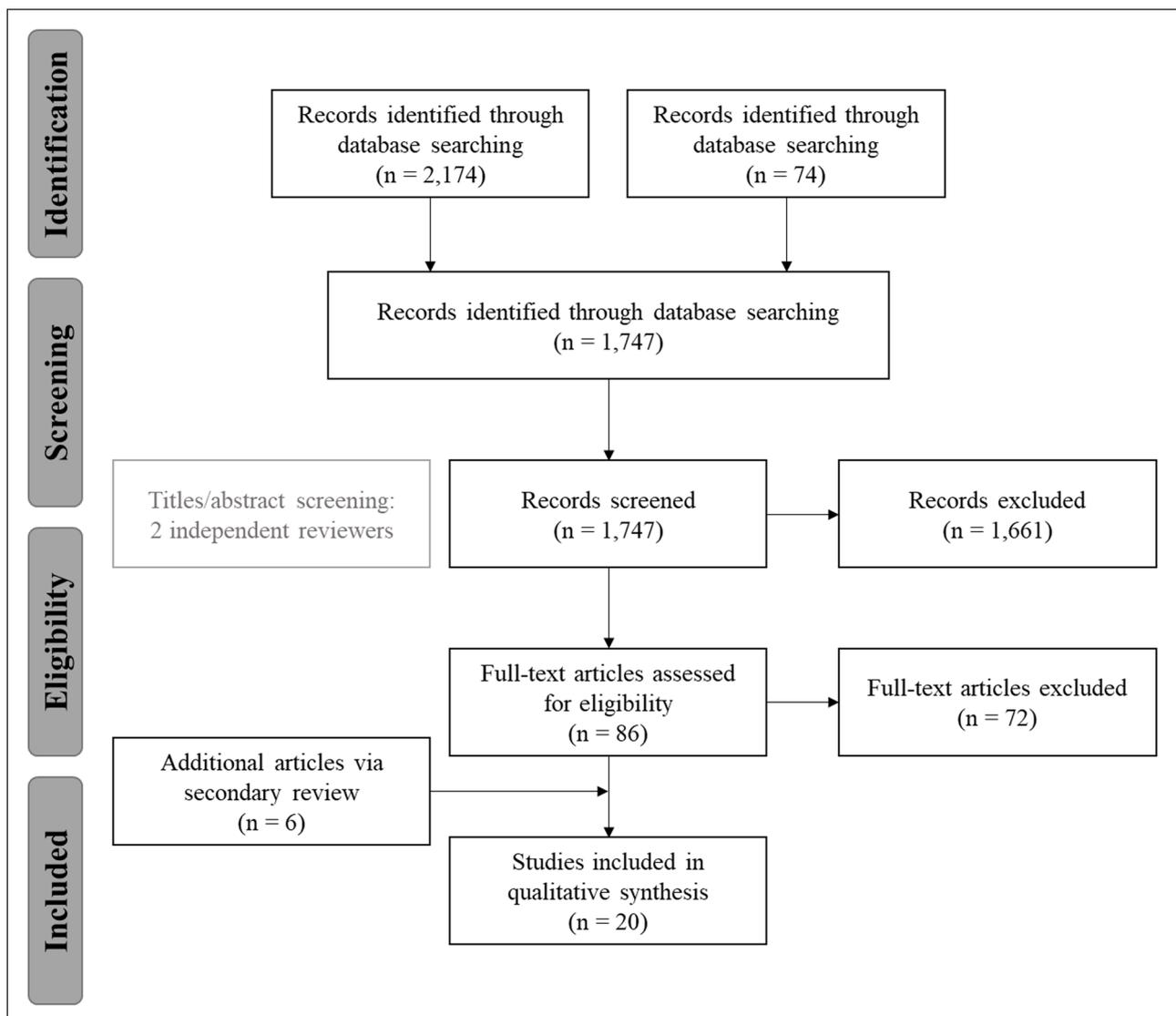

Figure 1: PRISMA Diagram.

Table 1: Breakdown of the 15 distinct eSource Interventions.

	NOT Part of an Ongoing Trial		Part of an Ongoing Trial		
	NO FHIR	FHIR	NO FHIR	FHIR	
SS-SE	11	1	–	–	12 (80%)
MS-ME	2	–	1	–	3 (20%)
	13 (86%)	1 (7%)	1 (7%)	0 (0%)	

Note: SS-SE = single-site, single-EHR implementation; MS-ME = multi-site, multi-EHR.

in clinical research. The authors indicate that the tool will be used and evaluated as part of six ongoing clinical trials, across three hospitals (a multi-site, multi-EHR implementation).[29] To date, the implementation, use, and evaluation of the intervention has not been completed (no published results). As such, it was not counted as part of the "distinct eSource interventions" count. Still, if carried out, this work has the potential to be a significant advancement in the field and worth additional review. Thus, while this study is not included in the total intervention count, additional information is included in Appendix A.

Of the five newly identified interventions, four were single-site, single EHR implementations[24–27] and one was a multi-site, multi-EHR implementations.[28] Most of the interventions (four of five) leveraged older standards, i.e., Health Level Seven (HL7®) v2.5, HL7® Clinical Document Architecture (CDA®), Integrating the Healthcare Enterprise (IHE) Retrieve Form for Data Capture (RFD).[24–26,28] However, one of the five did utilize the HL7® Fast Healthcare Interoperability Resources (FHIR®) standard.[27] Even so, the FHIR-based solution was a single-site, single-EHR implementation not yet evaluated within the context of a clinical trial.

When consolidated with the results of the primary literature review, the final breakdown is as follows (**Table 1**). Across the final 20 manuscripts identified, there were 15 distinct eSource interventions evaluated. Of the 15 interventions, 12 (80%) were single-site, single-EHR (SS-SE) implementations and 3 (20%) were multi-site, multi-EHR (MS-ME) implementations. All 15 implementations mentioned the use of standards (some in more detail than others), but nearly all (14 of 15, or 93%) implementations were based on older data exchange standards. Again, only one[27] referenced the use of FHIR®, but it was a single-site, single-EHR implementation and not yet evaluated in the context of a clinical trial. Additional information on the full set of manuscripts (n = 20) is available in the Appendix A, including study categorization, findings, and limitations.

4. Discussion
4.1. Single-Site, Single-EHR Implementations
Use of EHR data for research has been occurring since the earliest uses of computers in medicine.[1] Early demonstrations of what can be referred to as first-generation EHR-to-EDC eSource implementations, occurred sporadically, often as single-study, single-site implementations.[2] Reports of the direct use of EHR data in prospective studies appeared as late as 2007.[8,9] With one exception – the STARBRITE proof-of-concept[12] – these early attempts were often custom builds within institutional EHR systems and largely did not leverage the nascent clinical research data standards simultaneously in development.

Later attempts at EHR-to-EDC eSource went a step further toward generalizability across systems, studies and institutions and developed[32–38] or utilized[12–18,24–27] data standards to facilitate consistent definition and format of the data. These second-generation EHR-to-EDC solutions significantly advanced EHR-to-research interoperability through the use of data standards.[3,5,7] Still, these implementations were few and, like early attempts, were confined to single-EHRs, single-studies (EDCs), or single-institutions.[11] Further, evaluation measures have infrequently included quality and cost outcomes and are difficult to compare across studies.[6,31]

For example, in 2009, the Munich Pilot,[13] utilized a pre-post design to (1) measure the impact of the integrated EHR–EDC system on trial quality, efficiency, and costs, and (2) characterize the workflow differences between 'traditional' trial management and the integrated EHR–EDC system. The Munich Pilot was conducted within a 19-patient investigator-initiated oncology trial, for which between 48 to 69 percent of the study data (more for some visits, less for others) were prepopulated, resulting in an almost five-hour reduction in data collection time.[13] Time parameters appear to have been manually logged, i.e., not blinded, and data quality was measured by the number of system or monitor generated queries.[13] Though only these weak endpoints were measured, the Munich Pilot demonstrated a statistically significant reduction in time for data collection activities.[13] However, there were too few data queries to assess this pilot study's impact on data quality.[13,39,40]

In 2014, Laird-Maddox, Mitchell, and Hoffman[16] implemented and evaluated a technical solution at Florida Hospital (Orlando, Florida), referred to as the Florida Hospital Cerner Discovere Pilot. Leveraging the IHE RFD standard and the HL7 Continuity of Care Document (CCD®), the objective of the study was to enable the electronic exchange of EHR data from the Florida Hospital Cerner Millennium EHR to the separate and independent, web-based Discovere EDC platform. The pilot reported minimal interruption of the EHR session and available data flow from the EHR to the study eCRF without manual reentry.[16] The investigators claimed improved data quality and reduced data collection time, but the results were not quantified.[16]

In 2015, at the University of Arkansas for Medical Sciences (UAMS), Lencioni and colleagues[17] implemented and evaluated an EHR-to-Adverse Event Reporting System (AERS) integration. The premise was to develop a tool to automate the detection of adverse events (AEs) using routine clinical data directly from the EHR using MirthConnect's web service, HL7® messages, and the IHE Retrieve Process for Execution (RPE) integration

profile.[17] The system integration software was developed to provide systematic surveillance and detection of AEs knowable from the health record including (1) lab related adverse events that are automatically generated based on study participant's lab results and (2) unscheduled visits. Although data quality was not assessed, implementation of the AERS system was associated with a reduction in "sponsor generated AE-related queries, and a staff-estimated 75% increase in lab-based AE reporting."[11,17] The system is still in use and demonstrates that it is possible for clinical research systems to interface directly with a site's EHR. Still, it is important to note that the outcomes assessed as part of this study were based on staff perceptions, and, again, this was a single-site implementation.

In 2017, at Duke University, Nordo and colleagues[18] developed and evaluated the RADaptor software, a standards-based, EHR-to-EDC solution. RADaptor was deployed for use at a single institution (Duke) for an OB/GYN registry. Similar to the Cerner Discovere Pilot, the RADaptor utilized the IHE RFD integration profile and compared data collection via the eSource approach versus the standard approach (traditional medical record abstraction). Through the use of RADaptor, a decrease in collection time (37%) and transcription errors (eSource, 0% versus non-eSource, 9%) was achieved.[18] Although these results are promising, the evaluation was performed on a single-site implementation; and, therefore, these findings lack generalizability.

Around the same time, Matsumura and colleagues[24] published their work on the CRF Reporter, a tool developed to improve efficiency in clinical research by integrating EMRs and EDCs. According to the authors, CRF Reporter is an independent application that can be integrated with the EMR (an RFD-like integration). It offers dynamic, "progress notes" templates for clinical data entry that utilize Clinical Data Interchange Standards Consortium (CDISC) Operational Data Model (ODM) for configuring study metadata (data element mapping), which then allows for standardized transmission from CRF Reporter to the "CRF receiver" within the EDC or clinical data management system (CDMS).[24] The templates are created by end-users (clinicians/researchers) using the template module from within the CRF Reporter interface;[41,42] the functionality is similar to that of semi-structured clinical progress notes.[24,41,42]

A major limitation of this work is the lack of detail surrounding the evaluation (no mention of evaluative processes or measures) and the lack of evidence (no data) to support the claims noted within the manuscript. For example, while there is mention of the number of templates and data points registered and used at one site (Osaka University Hospital), there is no quantitative evidence to indicate whether or not the CRF Report has been useful (e.g., improved data quality, reduced site burden, or reduced costs). In their primary manuscript, Matsumura and colleagues claim that the system (1) "has been evaluated in detail and is regarded with high esteem by all of [clinical research coordinators] and doctors in

charge" and (2) "can save both labor and financial costs in clinical research;"[41] but, again, there is no supporting evidence to justify these claims, nor is there any explanation of any evaluative measures used to quantify success.

In 2019, Takenouchi and colleagues published their work on the SS-MIX2 (Standardized Structured Medical Information eXchange2) project. Similar to the Cerner Discovere and RADaptor projects, SS-MIX2 leveraged the IHE RFD integration profile to support data exchange from the EMR to the EDC. The SS-MIX2 project was inspired by the early work of Kimura and colleagues[43] on the SS-MIX (Standardized Structured Medical Information eXchange) project. The SS-MIX project focused explicitly on the exchange of EMR data between institutions for clinical purposes by offering a "standardized storage" infrastructure through use of the HL7® v2.5 messaging standard.[43] Takenouchi's SS-MIX2 project took Kimura's work a step further to extend beyond routine clinical care to include the clinical research use case (promoting exchange between the EMR and EDC). However, the primary technology remained centered on the HL7® v2.5 messaging standard.

SS-MIX2 utilized IHE RFD to display a CRF template within the EMR, referred to as the "EMR Stamp," which allowed the clinician to enter both routine clinical and research-related data directly into the patient's health record. Although the intent of the product was to be utilized across multiple sites and EMRs, this particular project was a single-site (Hamamatsu University Hospital), single-EMR (Translational Research Center for Medical Innovation (TRI) System) implementation. The authors mentioned the use of data elements specific to a type 2 diabetes observational study, but it did not appear that this work was done within the context of the actual study, only that the data elements were used for template creation and mapping and testing the system. It was also noted that, due to the limitations in the type of data collected, SS-MIX2 would be most applicable to "simpler protocols" (i.e., observational studies).[25] Takenouchi and colleagues also indicated that there was potential for "better cost-benefit performances," resource savings, and improved data quality[25] – referencing the results from the RADaptor project.[18] These statements, though, were not supported by any quantitative results and simply referencing another project's results is insufficient and non-transferrable. Similar to the CRF Reporter publication,[24] Takenouchi and colleagues[25] centered their writing on the development process and technical architecture, but not on the system evaluation.

Around the same time, Rocca and colleagues published a report describing an eSource implementation at the University of California San Francisco (UCSF), referred to as the OneSource Project.[26] The OneSource Project was a collaborative effort between UCSF and the FDA to develop a standards-based approach for automating EHR-to-EDC data collection for clinical research studies. This work utilized CRFs from an existing phase II clinical trial, but the implementation and evaluation were conducted

within a test environment, and not within the context of an ongoing trial. Phase I of the OneSource demonstration project leveraged the HL7® CCD®, Clinical Data Interchange Standards Consortium (CDISC) Operational Data Model (ODM), and IHE RFD standards for the capture and transmission of clinical research data. The primary objective of Phase I was to "assess the utility of the standards-based technology"[26] through EHR-to-eCRF data element mapping using Clinical Data Acquisition Standards Harmonization (CDASH) terminology (similar to the data standards work done by Garza and colleagues[44–46] and to the RADaptor implementation work by Nordo and colleagues[18]). As part of Phase I, Rocca and colleagues[26] performed a gap analysis to identify the availability of CRF data elements within the EHR. However, the methods and results were not clearly articulated, and the sample (n = 31 data elements across 3 CRFs) was hyper-specific to the specific trial and much too small to be generalizable to other studies and other institutions. Moreover, neither the tool itself, nor the standards implemented, were truly evaluated as part of this work. Additionally, the standards implemented as part of Phase I – primarily HL7® CCD® and IHE RFD – have been identified in earlier works as outdated and/or lacking in content coverage.[11,18,46] Phase II was discussed briefly as proposed future work to extend OneSource by incorporating additional standards (i.e., HL7® FHIR®) and functionality (i.e., visualization systems integration); although, a timeline for this work was not provided.

Most recently, in 2020, Zong and colleagues[27] published preliminary results on a demonstration project using a FHIR-based mechanism, referred to as the "computational pipeline," for extracting EHR data for clinical research. The researchers leveraged an existing FHIR-based cancer profile (Australian Colorectal Cancer Profile, ACP) and the FHIR® Questionnaire and QuestionnaireResponse resources[47] to create a standards-based approach for automating CRF population from EHR data for cancer clinical trials.[27] For this particular use case, the researchers evaluated the accuracy of the tool in pulling data from the Mayo Clinic's Unified Data Platform (UDP), or clinical data warehouse, to a single CRF for 287 Mayo Clinic patients (1,037 colorectal cancer synoptic reports), with an average accuracy of 0.99.[27] Of all the publications identified through the systematic review, this is the only published study utilizing the FHIR® standard. However, the study was limited to a single-site, single-EHR implementation, and, like its predecessors, the findings lacks generalizability. Moreover, the mechanism developed did not connect to the EHR, but to a clinical data warehouse. The benefits to this use case are that Mayo's UDP provides access to "heterogeneous data across multiple data sources,"[27] including the EHR, which offers data seekers a single platform from which to extract data using FHIR®. Nevertheless, there are certain factors to consider when working with warehouse data. Typically, the source data that makes it into a clinical data warehouse

often undergo several rounds of transformation prior to reaching the warehouse. Thus, the warehouse is not always an exact copy of the source system(s); and, depending on the institution's procedures for pulling the data into the warehouse, there may be a lag between the data from the actual source system and the data in the warehouse. Further, data warehouses are developed internally, to address the specific data needs of a particular institution, and the results from one are not always generalizable to all. Zong and colleagues[27] also noted three limitations of their work: (1) The study was limited to a single, and relatively simple CRF; (2) the ACP (FHIR-based profile) was used to cover most of the data elements from their CRF, but it did not provide a comprehensive list of data elements typically required; and (3) much of the phenotypic cancer data required for research is often embedded in unstructured narratives, however, this study only focused on synoptic reports and structured data elements. This only further emphasizes the need for additional research in this area.

From here, the next step was determining the extent to which existing EHR data would be used in clinical research studies and, like the STARBRITE,[12] RADaptor,[18] and clinical pipeline[27] demonstration projects, evaluating existing or developing standards for their utility in secondary data use for clinical research. The clinical trial data collection use case has since become more fully developed as the direct use of EHR data via electronic extraction and semi- or fully-automated population of a study-specific eCRF within an EDC system has come to light.

4.2. Multi-Site, Single-EHR Implementations
The EHR4CR initiative, which began in 2011, was a multifaceted, collaborative project comprised of various phases.[19–22,30] Unlike its predecessors, the EHR4CR pilot study went beyond the single-site implementation approach, engaging with multiple academic medical hospitals across five European countries to install and evaluate the EHR4CR platform. Work is ongoing, but at the time of this review, the EHR4CR platform had not yet been evaluated within the context of a pragmatic trial.

In 2017, Ethier and colleagues published on the European TRANSFoRm Project. The TRANSForm Project was conducted from 2010–2015 as a pilot study to design and evaluate the implementation of a new, technical architecture for the Learning Health System (LHS) in Europe.[48] The TRANSFoRm eSource method and tools were formally evaluated using a mixed-methods study of the TRANSFoRm technology as a nested, cluster randomized trial embedded fully within an RCT (700 subjects, 36 sites, five countries and five different EHR systems). The primary objective for the study was to evaluate clinical trial recruitment, comparing eSource to traditional methods.[23] The study's sample size estimate assumed the TRANSFoRm system would increase subject recruitment by 75% (RR 1.75), from 20% in the control arm to 35% in the intervention arm.[23] In the actual study, average recruitment rates were 43% and 53%,

respectively in the TRANSFoRm and control arms with a large between-practice range in recruitment rates. Most importantly, TRANSFoRm was able to demonstrate "that implementation of EHR-to-EDC integration can occur within an RCT's start-up timeline."[11,23]

In 2018, Devine and colleagues published their work on the Comparative Effectiveness Research and Translation Network (CERTAIN) Validation Project. The intent of the project was to develop and validate a semi-automated data abstraction tool for clinical research. The abstraction tool, herein referred to as CERTAIN, leveraged HL7® standard messages (although the version was not fully specified) to semi-automate the data abstraction process from disparate EHR systems to a central data repository (CDR), with the intent of using the consolidated data from the CDR for quality improvement (QI) and clinical research purposes. Four unaffiliated sites installed the commercial CDR product to retrieve and store data from their EHRs (one Cerner, one Siemens, and two Epic sites) using a centralized federated data model.[28] The data elements selected as part of the validation project were specific to those collected for Washington State's Surgical Care Outcomes and Assessment Program (SCOAP) QI registry. A two-step process[28,49] was used to select "which EHR data sources to ingest and which SCOAP data elements to automate," resulting in five domains of interest – ADT (admission, discharge, and transfers), laboratory services, dictations, radiology reports, and medications/allergies – across three different forms – abnormal/oncologic, non-cardiac vascular, and spine.[28] Then, a comparison was performed to assess data availability within the various EHRs and accuracy between the manual and automated methods for case identification – both processes occurring in parallel. The validation was conducted in three phases: ingestion, standardization, and concordance of automated versus manually abstracted cases.[28]

The CERTAIN Validation Project went beyond many of the previous studies (both SS-SE and MS-ME) in that the researchers conducted an evaluation of the source data to support their use case (across multiple EHRs). Through this work, Devine and colleagues[28] identified significant gaps in the EHR with regards to the availability of the data in a format conducive to automated extraction techniques (structured versus unstructured/requiring NLP versus requires "additional resources"). The evaluation pertaining to the EHR-to-EDC data collection mechanism (CERTAIN) found lower accuracy rates in identifying SCOAP cases through use of their semi-automated tool versus traditional manual approach. It is unclear as to whether the standard of choice was a factor; the authors noted the use of "HL7 standard messages", but did not provide detail as to the specific standard or the version implemented.

Both EHR4CR and TRANSFoRm have built on earlier successes and have expanded system functionality beyond that of the aforementioned first- and second-generation EHR-to-EDC eSource implementations. Moreover, each solution was designed to support multicenter clinical trials, leveraging data exchange standards and advancing the field by more consistently evaluating quality and efficiency. However, EHR4CR was reliant on sites having clinical data warehouses with relevant EHR data and has not yet been evaluated for data collection in a multicenter clinical trial; and while the TRANSFoRm project was conducted in the context of an ongoing clinical trial, only 26 data elements were extracted, and the evaluation conducted on the implementation was only focused on the impact on recruitment rate, for which it failed to show efficacy.

4.3. Limitations
The limitations of this review are as follows. Although robust databases were utilized to identify the publications, relevant manuscripts were difficult to find. Further, much of the work done in the standards development and implementation space is not often published in peer reviewed journals, but rather as white papers or within project-specific wikis or project management platforms. Thus, there is the potential that some relevant work may have been missed. Lastly, the methods used for search and review of the literature could have been strengthened by utilizing a second reviewer throughout the review process, primarily to assist with screening full-text articles.

5. Conclusion
From these results, several things become apparent. First, given the long-standing desire to leverage EHR data for secondary use and streamline data collection for clinical research, there are currently few examples of direct EHR-to-EDC eSource implementations in the published literature. Second, of those publications that are available, the majority were single-site, single-EHR implementations that (1) were not evaluated within the context of an ongoing clinical trial, (2) referenced older (or no) data exchange standards, and/or (3) were inconsistent in measuring and evaluating the implementation. As such, the scalability and generalizability of these attempts comes into question. Of the three multi-site, multi-EHR implementations presented, two referenced older standards and one did not specify the version used. A fourth implementation was identified, but has not yet published results, and while it was said to be part of an ongoing clinical trial, there is no information yet available on whether or not the implementation was carried out accordingly, nor on the measures (and results) used to evaluate the implementation. The results of the literature review only further emphasize the observation by Kim, Labkoff, and Holliday[10] that the clinical trial data collection use case continues to be the most difficult and least demonstrated. Thus, additional work is critically needed in this area to address the gaps identified from the literature. We acknowledge the ongoing work to address existing interoperability and data exchange issues in clinical research; but the results of these efforts do not appear to be readily available to the broader scientific community. Therefore, we encourage those working in this area to publish their findings.

Additional File

The additional file for this article can be found as follows:

· **Appendix A.** Summary of the Literature (S = single-site/-EHR, M = multi-site/-EHR). Adapted from Garza and colleagues, 2019. DOI: https://doi.org/10.47912/jscdm.66.s1

Abbreviations

Term* indicates further explanation of the term below

CRF = Case Report Form [for case reporting in clinical trials]
CRF Reporter is an independent application that can be integrated with the EMR (an RFD*-like integration)
HL7® v2.5 = Health Level Seven
HL7® CDA® = Clinical Document Architecture
IHE = Integrating the Healthcare Enterprise [independent non-profit membership organization]
RFD = IHE RFD = Retrieve Form for Data Capture – provides a method for gathering data within a user's current application to meet the requirements of an external system
HL7® FHIR® = Fast Healthcare Interoperability Resources
HL7® CCD® = Continuity of Care Document
CDISC = Clinical Data Interchange Standards Consortium (CDISC)
CDISC ODM = CDISC Operational Data Model (ODM)
CDMS = clinical data management system
CDASH = CDISC's Clinical Data Acquisition Standards Harmonization establishes a standard way to collect data consistently across studies and sponsors so that data collection formats and structures provide clear traceability of submission data into the STDM*
STDM = Study Data Tabulation Model
ACP = Australian Colorectal Cancer Profile

Funding Information

Competing Interests

The authors have no competing interests to declare.

Author Contributions

All authors contributed materially to study design. MYG and MZ independently reviewed the titles and abstracts for inclusion. MYG reviewed the full articles for inclusion and analyzed the results. MYG wrote the article. All authors reviewed the manuscript and contributed to revisions. All authors reviewed and interpreted the results and read and approved the final version.

References

1. **Forrest WHJ, Bellville JW.** The use of computers in clinical trials. *Brit J Anaesth.* 1967; 39(4): 311–319. DOI: https://doi.org/10.1093/bja/39.4.311
2. **Collen MF.** Clinical research databases–a historical review. *J Med Syst.* 1990 Dec; 14(6): 323–44. DOI: https://doi.org/10.1007/BF00996713
3. **Safran C, Bloomrosen M, Hammond WE,** et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association white paper. *J Am Med Inform Assoc.* 2007 Jan–Feb; 14(1): 1–9. DOI: https://doi.org/10.1197/jamia.M2273
4. **Goodman K, Krueger J, Crowley J.** The automatic clinical trial: leveraging the electronic medical record in multisite cancer clinical trials. *Curr Oncol Rep.* 2012 Dec; 14(6): 502–8. DOI: https://doi.org/10.1007/s11912-012-0262-8
5. **Eisenstein EL, Nordo AH, Zozus MN.** Using medical informatics to improve clinical trial operations. *Stud Health Technol Inform.* 2017; 234: 93–97.
6. **Raman SR, Curtis LH, Temple R,** et al. Leveraging electronic health records for clinical research. *Am Heart J.* 2018 Aug; 202: 13–19. DOI: https://doi.org/10.1016/j.ahj.2018.04.015
7. **Nordo AH, Levaux HP, Becnel LB,** et al. Use of EHRs data for clinical research: historical progress and current applications. *Learn Health Syst.* 2019 Jan 16; 3(1): e10076. DOI: https://doi.org/10.1002/lrh2.10076
8. **Gersing K, Krishnan R.** Clinical computing: clinical management research information system (CRIS). *Psychiatr Serv.* 2003 Sep; 54(9): 1199–200. DOI: https://doi.org/10.1176/appi.ps.54.9.1199
9. **Murphy EC, Ferris FL, 3rd, O'Donnell WR.** An electronic medical records system for clinical research and the EMR EDC interface. *Invest Ophthalmol Vis Sci.* 2007 Oct; 48(10): 4383–9. DOI: https://doi.org/10.1167/iovs.07-0345
10. **Kim D, Labkoff S, Holliday S.** Opportunities for electronic health record data to support business functions in the pharmaceutical industry – a case study from Pfizer, Inc. *J Am Med Inform Assoc.* 2008 Jun; 15(5): 581–584. DOI: https://doi.org/10.1197/jamia.M2605
11. **Garza M, Myneni S, Nordo A,** et al. eSource for standardized health information exchange in clinical research: a systematic review. *Stud Health Technol Inform.* 2019; 257: 115–124.
12. **Kush R, Alschuler L, Ruggeri R,** et al. Implementing single source: the STARBRITE proof-of-concept study. *J Am Med Inform Assoc.* 2007 Sep–Oct; 14(5): 662–73. DOI: https://doi.org/10.1197/jamia.M2157
13. **Kiechle M, Paepke S, Shwarz-Boeger U,** et al. EHR and EDC integration in reality. *Applied Clinical Trials Online*; 2009 Nov. Available from: https://www.appliedclinicaltrialsonline.com/view/ehr-and-edc-integration-reality. Accessed September 2018.
14. **El Fadly A, Lucas N, Rance B,** et al. The REUSE project: EHR as single datasource for biomedical

research. *Stud Health Technol Inform.* 2010; 160(Pt 2): 1324–8.

15. **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 Dec; 44(Suppl 1): S94–102. DOI: https://doi.org/10.1016/j.jbi.2011.07.007

16. **Laird-Maddox M, Mitchell SB, Hoffman M.** Integrating research data capture into the electronic health record workflow: real-world experience to advance innovation. *Perspect Health Inf Manag.* 2014 Oct 1; 11(Fall): 1e.

17. **Lencioni A, Hutchins L, Annis S,** et al. An adverse event capture and management system for cancer studies. *BMC Bioinformatics.* 2015; 16 Suppl 13(Suppl 13): S6. DOI: https://doi.org/10.1186/1471-2105-16-S13-S6

18. **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 Jul; 103: 89–94. DOI: https://doi.org/10.1016/j.ijmedinf.2017.04.015

19. **Beresniak A, Schmidt A, Proeve J,** et al. Cost-benefit assessment of the electronic health records for clinical research (EHR4CR) European project. *Value Health.* 2014 Nov; 17(7): A630. DOI: https://doi.org/10.1016/j.jval.2014.08.2251

20. **Doods J, Bache R, McGilchrist M,** et al. Piloting the EHR4CR feasibility platform across Europe. *Methods Inf Med.* 2014; 53(4): 264–8. DOI: https://doi.org/10.3414/ME13-01-0134

21. **De Moor G, Sundgren M, Kalra D,** et al. Using electronic health records for clinical research: the case of the EHR4CR project. *J Biomed Inform.* 2015 Feb; 53: 162–73. DOI: https://doi.org/10.1016/j.jbi.2014.10.006

22. **Dupont D, Beresniak A, Sundgren M,** et al. Business analysis for a sustainable, multi-stakeholder ecosystem for leveraging the electronic health records for clinical research (EHR4CR) platform in Europe. *Int J Med Inform.* 2017 Jan; 97: 341–352. DOI: https://doi.org/10.1016/j.ijmedinf.2016.11.003

23. **Ethier JF, Curcin V, McGilchrist MM,** et al. eSource for clinical trials: implementation and evaluation of a standards-based approach in a real world trial. *Int J Med Inform.* 2017 Oct; 106: 17–24. DOI: https://doi.org/10.1016/j.ijmedinf.2017.06.006

24. **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. *Stud Health Technol Inform.* 2017; 245: 516–520.

25. **Takenouchi K, Yuasa K, Shioya M,** et al. Development of a new seamless data stream from EMR to EDC system using SS-MIX2 standards applied for observational research in diabetes mellitus. *Learn Health Syst.* 2018 Nov 15; 3(1): e10072. DOI: https://doi.org/10.1002/lrh2.10072

26. **Rocca M, Asare A, Esserman L,** et al. Source data capture from EHRs: using standardized clinical research data; 2019 Oct. Available from https://www.fda.gov/science-research/advancing-regulatory-science/source-data-capture-electronic-health-records-ehrs-using-standardized-clinical-research-data. Accessed October 2020.

27. **Zong N, Wen A, Stone DJ,** et al. Developing an FHIR-based computational pipeline for automatic population of case report forms for colorectal cancer clinical trials using electronic health records. *JCO Clin Cancer Inform.* 2020 Mar; 4: 201–209. DOI: https://doi.org/10.1200/CCI.19.00116

28. **Devine EB, Van Eaton E, Zadworny ME,** et al. Automating electronic clinical data capture for quality improvement and research: the CERTAIN validation project of real world evidence. *eGEMS.* 2018 May 22; 6(1): 8. DOI: https://doi.org/10.5334/egems.211

29. **Griffon N, Pereira H, Djadi-Prat J,** et al. Performances of a solution to semi-automatically fill eCRF with data from the electronic health record: protocol for a prospective individual participant data meta-analysis. *Stud Health Technol Inform.* 2020 Jun 16; 270: 367–371.

30. **Beresniak A, Schmidt A, Proeve J,** et al. Cost-benefit assessment of using electronic health records data for clinical research versus current practices: contribution of the electronic health records for clinical research (EHR4CR) European project. *Contemp Clin Trials.* 2016 Jan; 46: 85–91. DOI: https://doi.org/10.1016/j.cct.2015.11.011

31. **Nordo A, Eisenstein EL, Garza M,** et al. Evaluative outcomes in direct extraction and use of EHR data in clinical trials. *Stud Health Technol Inform.* 2019; 257: 333–340.

32. **Nahm N, Brian BM, Walden A.** HL7 version 3 domain analysis model: Cardiology, Release 1 V3: Cardiology: Cardiology (acute coronary syndrome) domain analysis model, R1. Health Level Seven (HL7) International; 2008.

33. **Nahm M, Walden A, Diefenbach J,** et al. HL7 version 3 standard: Public health; Tuberculosis domain analysis model, Release 1. Tuberculosis DAM R1. Health Level Seven (HL7) International; 2009.

34. **Hurrell M, Monk T, Hammond WE,** et al. HL7 version 3 domain analysis model: Preoperative anesthesiology, Release 1, HL7 informative document: HL7 V3 DAM ANESTH R1-2013. A Technical Report prepared by Health Level Seven (HL7) International and registered with ANSI: 5/31/2013; 2013.

35. **Zozus M, Younes M, Walden A.** Schizophrenia standard data elements. Health Level 7, Release 1. Domain analysis model (DAM) document of the HL7 verion 3 domain analysis model: Schizophrenia, release 1 – US Realm informative document: HL7 V3 DAM SCHIZ, R1. A Technical Report prepared by Health Level Seven (HL7) International and registered with ANSI: 10/26/2014; 2014.

36. **Zozus M, Younes M, Kluchar C,** et al. Major depressive disorder (MDD) standard data elements.

Health Level 7, Release 1. Domain analysis model (DAM) document of the HL7 verion 3 domain analysis model: Major depressive disorder, release 1 – US Realm informative document: HL7 V3 DAM MDD, R1. A Technical Report prepared by Health Level Seven (HL7) International and registered with ANSI: 10/26/2014; 2014.

37. **Zozus M, Younes M, Walden A.** Bipolar disorder standard data elements. Health Level 7, Release 1. Domain analysis model (DAM) document of the HL7 verion 3 domain analysis model: Bipolar disorder, release 1 – US Realm informative document: HL7 V3 DAM BIPOL, R1. Health Level Seven (HL7) International; 2016.

38. **Zozus M, Younes M, Walden A.** Generalized anxiety disorder standard data elements. Health Level 7, Release 1. Domain analysis model (DAM) document: Generalized anxiety disorder, release 1 – US Realm informative document: HL7 V3 DAM GAD, R1. Health Level Seven (HL7) International; 2017.

39. **Nahm M.** Measuring data quality. *Society for Clinical Data Management*; 2012.

40. **Zozus MN, Pieper C, Johnson CM,** et al. Factors affecting accuracy of data abstracted from medical records. *PLoS One.* 2015 Oct 20; 10(10): e0138649. DOI: https://doi.org/10.1371/journal.pone.0138649

41. **Matsumura Y, Hattori A, Manabe S,** et al. Interconnection of electronic medical record with clinical data management system by CDISC ODM. *Stud Health Technol Inform.* 2014; 205: 868–72.

42. **Matsumura Y, Hattori A, Manabe S,** et al. A Strategy for reusing the data of electronic medical record systems for clinical research. *Stud Health Technol Inform.* 2016; 228: 297–301.

43. **Kimura M, Nakayasu K, Ohshima Y,** et al. SS-MIX: A ministry project to promote standardized healthcare information exchange. *Methods Inf Med.* 2011; 50(2): 131–9. DOI: https://doi.org/10.3414/ME10-01-0015

44. **Garza M, Del Fiol G, Tenenbaum J,** et al. Evaluating common data models for use with a longitudinal community registry. *J Biomed Inform.* 2016 Dec; 64: 333–341. DOI: https://doi.org/10.1016/j.jbi.2016.10.016

45. **Garza M, Seker E, Zozus M.** Development of data validation rules for therapeutic area standard data elements in four mental health domains to improve the quality of FDA submissions. *Stud Health Technol Inform.* 2019; 257: 125–132.

46. **Garza MY, Rutherford M, Myneni S,** et al. Evaluating the coverage of the HL7 FHIR standard to support eSource data exchange implementations for use in multi-site clinical research studies. *AMIA Annu Symp Proc;* 2020 Nov. DOI: https://doi.org/10.3233/SHTI210188

47. **Health Level Seven (HL7).** HL7 FHIR Release 4. Health Level Seven; 2019. Available from: https://hl7.org/fhir/R4/index.html. Accessed November 2019.

48. **i-HD. TRANSFoRm.** 2016. Available from: https://www.i-hd.eu/index.cfm/resources/ec-projects-results/transform/. Accessed September 2018.

49. **Capurro D, Yetisgen M, van Eaton E,** et al. Availability of structured and unstructured clinical data for comparative effectiveness research and quality improvement: a multisite assessment. *eGEMS.* 2014 Jul 11; 2(1): 1079. DOI: https://doi.org/10.13063/2327-9214.1079