ORIGINAL RESEARCH

Automating Data Entry from Electronic Health Record to Electronic Data Capture Using a Trusted Cloud-Based Application in Multisite Cancer Clinical Trials

Keith Goodman[*,†], Chris Cook[*,†], Dani Weatherbee[*,‡], Sunita Yadav[§], Dinesh Pal Mudaranthakam[‖], Rachael Sexton[*,†], Nichole Mahaffey[¶], Leslie Garcia[¶], Teresa Ta[¶], Erin Cebula[**], Janie Smart[‖], Tanya Goudeau[‖], Sandy Annis[‖], Aubrey Gilmore[*,†], Daria Chugina[*,†], Pasarlai Ahmadzai[*,†], Antje Hoering[*,†] and Michael LeBlanc[*,‡]

Introduction: For more than two decades, researchers have sought to develop and to improve technologies to seamlessly move data from electronic health record (EHR) systems to study forms in electronic data capture (EDC) systems. The goal is to utilize advancing technology to improve study data accuracy and quality while decreasing the burden of data collection on busy clinical research professionals.

Objectives: This report discusses findings from the SWOG Cancer Research Network's use of a secure and trusted third-party cloud-based application as a technology link between EHR databases and the clinical trial EDC application. The objectives are to ease the burden on site staff and improve data accuracy and completeness.

Methods: Three SWOG sites used a cloud-based EHR-to-EDC application to enter study data for two follow-up tumor assessment case report forms for six patients each. This software-assisted method was compared to manual medical record abstraction. Time savings, error rate, and interrater reliability were measured.

Results: A comparison of the two methods demonstrated substantial time savings and improvements in data quality. This is especially true for data fields that can be automatically captured using the application for clinical trials using the software-assisted approach.

Conclusions: Using a secure and trusted cloud-based application to access the EHR to assist in data collection for clinical trials resulted in welcome time savings for clinical research professionals and higher data accuracy.

Introduction

Clinical trial complexity, including the amount of data collected, has continued to increase.[1] The SWOG Cancer Research Network, a National Cancer Institute (NCI)-supported National Clinical Trials Network (NCTN) group, has worked with a third-party vendor to evaluate and implement an automated method for gathering data from the electronic health record (EHR) for clinical trial research.[2] SWOG's long-standing goal is to automate data collection and processing to increase data quality and to lessen the burden on clinical research staff. EDC applications are currently the standard for data entry for trials, having replaced paper case report forms (CRFs), even though EDCs typically require duplicate entry of data already captured in source databases, such as EHR systems. Integrating EHR databases and the clinical trial EDCs would enable more accurate, efficient, and more affordable data collection.

This report provides an overview of technology, data standards and data quality; it examines the benefits and barriers of using a trusted cloud-based application to facilitate clinical trial data entry. It also describes the problem being addressed, reports on results, and

* SWOG Cancer Research Network Statistics and Data Management Center, Seattle, WA, USA

† Cancer Research And Biostatistics, Seattle, WA, USA

‡ Fred Hutchinson Cancer Center, Seattle, WA, USA

§ nCoup Inc., Fremont, CA, USA

‖ University of Kansas Cancer Center, Kansas City, KS, USA

¶ UC Davis Comprehensive Cancer Center, Sacramento, CA, USA

** University of Rochester Medical Center, Rochester, NY, USA

Corresponding author: Chris Cook (chrisc@crab.org)

Art. 1, page 2 of 16

Goodman et al: Automating Data Entry from Electronic Health Record to Electronic Data Capture Using a Trusted Cloud-Based Application in Multisite Cancer Clinical Trials

discusses the findings from evaluating an application that integrates site EHR systems with the SWOG EDC. While the clinical trial used in this evaluation is in multiple myeloma cancer, the concepts are likely to be generalizable to other diseases.

Background

Over time, technological advancements have led to data collection improvements in clinical trial research. In 2003, SWOG replaced paper forms with an online electronic data capture (EDC) application, which had been under development for several years.[3] According to a 2020 survey of 500 global organizations, most clinical operations professionals (91%) report that their organizations use electronic case report form (eCRFs) within EDC applications.[4] With the widespread adoption of EDC applications, emerging data standards and enhancements to EHR applications have expanded opportunities for digital data transfer from the EHR to the EDC.[5]

Feedback mechanisms to improve data quality have also been established. For example, electronic queries are used to verify data consistency. These queries are generated by data coordinators, who review EDC data at the sponsor's data management center and submit the queries to site clinical research professionals (CRPs), also known as clinical research coordinators and study coordinators. Data quality has also been improved through robust automated edit checks on the data fields to prompt CRPs to review specific entries if a value is outside of a normal range or otherwise requires review.

Clinical Trial Data in the EHR

Patient health information resides within the EHR, including integrated or separate databases with patients' medical history, diagnoses, treatments, medications, laboratory reports, test results, radiology images, doctor's notes and recommendations, and more.[6] These records are valuable sources of data for clinical trial research even though the systems are designed and focused on patient care and billing.[7,8] A significant amount of data in the EHR remain unstructured, meaning that they do not follow universal data standards and are not stored in a pre-specified format.

For most clinical trials, data are primarily extracted from the EHR via human interaction, manually or with assistance from a data management system, using a data entry process commonly called medical record abstraction (MRA).[3,9,10] CRPs access medical records in the EHR and then record that data in electronic eCRFs within the EDC application. Data may be structured or unstructured. Structured data may include any alphanumeric entries specific to the data field, such as laboratory test results, demographical information, vitals, and medications. Unstructured data types may include radiology, pathology and laboratory reports.

Data Standards and Technology

Guidelines and standards help to ensure that data gathered from the EHR is accurate and reliable for use in research. Before implementing a study using EHR data,

considered to be real-world evidence by the US Food and Drug Administration (FDA), researchers are directed to assess all potential data sources to determine the most reliable and relevant ones.[11] FDA guidance suggests that research answer clinical questions using real-world evidence available within health records.[11,12] Real-world evidence, defined by the FDA, is related to patient care information and health care delivery as gathered from EHR and other sources.[13,14]

The development of common data formats has improved the ability to utilize patient data across institutions and for research. Clinical Data Acquisition Standards Harmonization (CDASH), developed by the Clinical Data Interchange Standards Consortium (CDISC) and released in 2008, has been adopted for many clinical trials, including those sponsored under NCI grants.[15–17] Along with adding common data elements across trials, CDASH facilitates the reporting of trial outcome data to the FDA in a standardized format. While the CDISC CDASH standards contribute to a standard method for the collection of data in eCRFs, these standards were developed for clinical trials and do not extend to data within health records, the source of clinical trial data.

Another effort to standardize some oncology data elements is the Minimal Common Oncology Data Elements (mCODE™) Initiative, which began in 2018. The mCODE project focused on standardizing 90 data elements used in research and treatment related to oncology patients.[18,19] The initiative aimed to modify a set of identified unstructured data elements typically found in medical records into structured categories to be available within EHRs for future research. As work on mCODE standards continues, a related initiative, the ICAREdata® Project, aims to utilize mCODE data elements and other data in clinical research.[8,19] ICAREdata, which is a collaboration between the MITRE Corporation, a not-for-profit public interest group, and the Alliance for Clinical Trials in Oncology, a National Cancer Institute (NCI)-supported NCTN group, seeks to evaluate how well mCODE elements improve data for clinical trials.[8,20,21] These efforts show promise to improve interoperability to help researchers to utilize EHR data and to better serve patients, especially if EHR vendors and healthcare institutions adopt mCODE standards.[22]

Improvements in technology have been important to the improvement of the interoperability between systems, such as EHR and EDC applications. One of the most important advances in interoperability standards is the Health Level Seven International (HL7®) with Fast Healthcare Interoperability Resources (FHIR®). The HL7 organization was begun in 1987 with a mission to improve workflow and data exchange for health information.[23,24] Standards released under the leadership of HL7 improved the ability to extract information. The HL7 FHIR standards introduced a set of data elements, grouped in "resources," to provide data needed for healthcare in an international common data exchange standard.[25,26] As healthcare adoption of the HL7 FHIR standards increased, the National Institutes of Health (NIH) began to encourage their use in research.[27,28]

Goodman et al: Automating Data Entry from Electronic Health Record to Electronic Data Capture Using a Trusted Cloud-Based Application in Multisite Cancer Clinical Trials

Art. 1, page 3 of 16

Many EHR system vendors, such as Epic® and Oracle® Health, support using HL7 FHIR standards in research.[29] The readiness of sites to take advantage of HL7 FHIR for research is improving, yet there are barriers to implementing this technology. Eisenstein et al. (2023) surveyed 61 site clinical research coordinators, principal investigators, and informatics leaders at 23 pediatric research sites.[30] These sites conducted pharmacokinetic or pharmacodynamic studies to determine their readiness to utilize an application to move data from the EHR to eCRFs.[30] As of July of 2021, only ten of the 23 responding sites had FHIR technology in use or under development. The primary barriers to implementing HL7 FHIR were organizational priorities, structure, and limited resources.[30] Since then, more sites have adopted the newer technology, likely driven in part by the 21st Century Cures Act, which requires patient data interoperability using HL7 FHIR. Use of standardized health data classes and elements, known as United States Core Data for Interoperability (USCDI), were defined to facilitate data sharing.[31]

The HL7 FHIR standard is important for research because it can effectively access clinical trial data. In 2021, Garza et al. (2021) examined three studies in three therapeutic areas. The researchers determined that approximately half of the sought-after data elements were available via an HL7 FHIR interface to EHR data.[32] Results from these recent manuscripts provide optimism that data capture using HL7 FHIR will continue to advance research as sites can implement the technology.

While data capture applications have traditionally worked best with structured data, efforts to extract data from unstructured text have seen some progress using tools related to machine learning or artificial intelligence, such as natural language processing (NLP). NLP uses algorithms to extract data from text in medical records and radiologic images.[33,34] NLP also uses intelligent character recognition to provide insight into the unstructured narrative and textual data in the EHR.[3] Research has demonstrated that NLP algorithms have the potential to be an effective tool in the extraction of patient diagnosis and treatment information, especially if the algorithms are configured to recognize data structures related to the data being sought.[35] The ability of NLP to extract clinical trial data may improve as advancements in machine learning and artificial intelligence advances are applied to the challenge.[36]

Data Quality

Data quality is an important element of the use of EHR data for research. Principal investigators are considered to be responsible for all facets of data quality which traditionally have been evaluated based on completeness, uniqueness, timeliness, accuracy, validity, and consistency.[37] Responsibilities also extend to patient safety. FDA regulations for recordkeeping and record retention place the responsibility on the principal investigator under the Code of Federal Regulations (CFR) Title 21, section 312.62.[38] An investigator's responsibilities are broad, including data quality maintenance through recordkeeping, oversight, training of study staff, patient safety and records, and regulatory compliance.[39] Investigators are guided by the International Conference on Harmonisation (ICH) Guideline for Good Clinical Practice (GCP) to safeguard patients and ensure scientific quality. Although these guidelines are not law, they provide best practices for the conduct of clinical trials.[40]

While regulation and GCP address the roles of the investigator and study team, technology standards address processes used to assist in the collection of data while maintaining quality and data integrity. The National Cancer Registrars Association (NCRA) acknowledges that the use of EHR software to abstract registry data through mapping must reduce time and effort, and automated processes must, at minimum, represent patient case data with fidelity and predictability.[41] The NCRA defines fidelity as the "degree to which the data represents the actual case history of the patient" and predictability as the "percent of occurrences in which the abstract will be accurate."[41]

Despite the emergence of standards and new technologies, data that reside in EHRs are not easily used in clinical trials without some degree of assistance. Data quality and the potential for error is dependent on the source of the data.[42] Key dimensions of data quality include accuracy, completeness, consistency, timeliness, relevance, granularity, specificity (no ambiguity), precision, and attribution.[42,43] The first two elements – accuracy and completeness – were called out in FDA guidance as being important to achieving reliable and relevant study results.[11] Traceability was also included as an important element for quality data. While it is difficult to account for all potential errors in the EHR or the clinical research database, established systematic processes are important in fostering a high level of data quality.[42] To this end, current clinical trial data reporting processes rely heavily on manual abstraction, recording, and curation by CRPs.

Manual MRA and data entry are the primary processes used for the transfer of EHR data to the EDC or data management system.[10] The typical MRA process is inefficient and requires significant time and effort.[2,5,43] Data within the EHR are often not found in one system but are instead located in interlocking or even disparate systems that contain both structured and unstructured information.[44,45] Manual processes require CRPs to access each of the systems to locate data, which they then use to fill out fields on eCRFs in the EDC. In many cases, structured information contained within the systems must be transformed in some way for entry into the study EDC. For example, adverse events represented in lab data often must be coded under Common Terminology Criteria for Adverse Events (CTCAE) classifications. Another example is unit conversion, such as when the medical record stores patient weight in kilograms while the eCRF requires pounds and ounces.

Observational and measurement differences by individuals may also contribute to data quality issues. It is recommended that site research teams have processes in place to assess and control for interrater discrepancies between staff at the same site and also between staff at different sites.[42] Measurement of agreement between two CRPs, known as interrater reliability (IRR), has been

Art. 1, page 4 of 16

Goodman et al: Automating Data Entry from Electronic Health Record to Electronic Data
Capture Using a Trusted Cloud-Based Application in Multisite Cancer Clinical Trials

effective in assessing data collected by two or more CRPs using MRA processes to collect the same data.[46–48] To improve data reliability and to decrease variability, manual processes must be combined with training.[48–50] Two common statistical measures, percent agreement and Cohen's kappa (k), are often used to evaluate IRR.[49,50] For example, percentage agreement as a measure of interrater reliability was evaluated by Zhao et al. (2022) along with several measurement indices for interrater reliability. The researchers found that percentage agreement held up well in determining true reliability.[50] The Cohen's kappa measurement is often preferred, when appropriate, as it considers the potential of agreement by chance.[49,50] On the other hand, Cohen's kappa places agreement measurements into broad categories that range from no agreement ($k = .20$ or less) to almost perfect agreement ($k = .90$ or higher). Some researchers have concluded that the Cohen's kappa categories should require higher scores to account for the high level of precision needed in measuring clinical data agreement.[51] Both percentage agreement and the kappa index have been widely used in health-related research.[49,52]

Manual MRA processes and its flaws can drive up costs and increase the possibility of errors.[53] With a focus on costs and data quality, automated data extraction and transfer to the EDC is viewed as having an economic benefit. Sundgren et al. (2021) examined costs associated with 21 oncology trials and projected potential cost savings of $15,000 per patient if 50% of all data fields could be automatically transferred from EHR to EDC versus manual entry.[5] Along with a cost-benefit advantage through efficiency, MRA assisted by direct data transfer from the EHR can also improve accuracy and timeliness.[9,53] The potential for time and cost savings, along with accuracy and timeliness, have contributed to the desire to find a secure and efficient method to transfer EHR data to the EDC.[54]

The Automatic Clinical Trial

Automatic transfer of medical record data from the EHR to the EDC across multiple sites using disparate EHR systems has been elusive.[3] The SWOG Statistics and Data Management Center (SDMC), co-located at the Fred Hutchinson Cancer Center and Cancer Research And Biostatistics, conducts multisite trials and seeks to utilize EHR-to-EDC technology to reduce data entry time and to improve data quality, accuracy and completeness. In 2013, the SWOG SDMC conducted a proof of concept study, named "The Automatic Clinical Trial," to examine a potential method to efficiently transfer data from EHR to EDC.[3] The SWOG SDMC worked with the UC Davis Comprehensive Cancer Center to transfer trial data for one study from the UC Davis EHR to eCRFs within an on-premise clinical trial management system (CTMS) known as Velos® eResearch.[3,55] The eCRFs were identical to those in the SWOG EDC. Results demonstrated that at least 30% of data fields could be directly transferred from the EHR to the CTMS where CRPs would review the data prior to transfer to the SWOG EDC. The use of CTMS technology showed promise in being able to reduce time and effort for CRPs and improving data quality at one

site, or multiple sites using the same system. However, extracting and saving EHR data locally would be inefficient for multisite trials as each site would need its own CTMS system with interoperability with the site's EHR and the central study database.

Some successful technology implementations have been used in research, but limitations have suppressed widespread use. For example, since 2012, Japan has used technology to gather research data directly from the EHR.[56–58] The system started with the use of a system similar to a CTMS and advanced to the use of a custom-built clinical data collection system that uses a custom CRF reporter system to transfer data to a common research database.[57,58] To facilitate the collection of medical images, a current version of the system utilizes HL7 FHIR interoperability to enable automated capture of medical images.[60] In 2021, use of the CRF reporter system was reported to be limited, as many hospitals in Japan had not implemented the CRF reporter or HL7 FHIR.[57]

In a 2017 proof of concept study, the Duke University Office of Research Informatics evaluated its own application, named RADaptor, that used an interoperability standard know as Retrieve Form Data Capture (RFD) to extract data from the EHR for a clinical trial. The Duke researchers found that the use of RADaptor to assist data capture resulted in a substantial 37% time savings for demographic data collection with application assistance with a 65% reduction in the total number of keystrokes and a 30% reduction in process motions such as scrolling.[59]

The use of EHR-to-EDC applications to support studies across multiple sites with a wide variety of EHR systems with HL7 FHIR interoperability have been rare. Garza et al. (2019) reviewed published manuscripts reporting on EHR-to-EDC data transfer and found 14 relevant articles that met criteria of being focused on direct EHR-to-EDC exchange in the context of a clinical study.[60] Eight of the 14 articles focused on single sites with a single EHR. Of the remaining six articles, four were related to the same European pilot project, named European EHR for Clinical Research (EHR4CR), and only one article focused on multiple sites and multiple EHR systems, the European TRANSFoRm project.[60,61]

In an article that updated the previous review of EHR-to-EDC articles, Garza et al. (2021) searched articles published between January 2018 and December 2020.[62] The researchers found 20 manuscripts that discussed 15 types of distinct interventions, three of which supported multisite research across multiple EHR systems using interoperability technologies that preceded HL7 FHIR. Of the other 15 articles, one reported on a pilot project that used HL7 FHIR interoperability but was limited to a single site with a single EHR.[62]

Based on lessons learned from SWOG's 2013 study and the development of emerging technologies, the SDMC considered how to improve efficiency for multisite trials at sites with different EHR systems through the use of a single cloud-based system, rather than individual on-premise systems, such as the CTMS system. SWOG researchers projected that a cloud-based research data

Goodman et al: Automating Data Entry from Electronic Health Record to Electronic Data Capture Using a Trusted Cloud-Based Application in Multisite Cancer Clinical Trials

Art. 1, page 5 of 16

Table 1: Criteria for Multisite Cloud-based Research Data Management System.

Requirement	Description
Substantial Increased Efficiency	Return on investment for site research teams, including time savings in the form of a reduction in both manual MRA and resolution of inconsistencies or errors
Structured Data Import	Able to capture demographic, medical history, treatment, medication, lab results and notes, and other test results for each patient with minimal intervention from site staff
Unstructured Data View	Able to capture care provider notes and recommendations, and other text information for review by site staff
Broad EHR Support	Integrates with EHR systems from many manufacturers
Support for Multiple Data Transmission Standards	Compatible with systems using native HL7 FHIR standard, traditional HL7, and EHR vendor program interfaces
Trusted – Secure and Private and HIPAA[a] Compliant	Compliant with standards and regulations and the vendor may need to be prepared to enter into a HIPAA Business Associate Agreement with sites
Adaptable to Evolving Data Standards	Ability to adapt to data structure advancements, such as mCODE®, and other efforts towards data standardization
Optional Extendable Use of Data	Reuse of trial data optionally available for site use for other purposes, such as source data verification, audit support, patient-related care, research and cancer or other disease registry reporting

a. Health Insurance Portability and Accountability Act (HIPAA) of 1996 and HIPAA Privacy Rule.[63,64]

management system would need to meet several criteria (see **Table 1**) to be an effective tool for clinical trials. Most importantly, any system would need to provide substantial time savings for research staff at healthcare institutions as improved operational efficiency has been considered to be an important benefit of direct EHR-to-EDC data transfer.[54]

An effective system was expected to support both structured and unstructured data and work with many EHR systems using current interoperability technologies. It must also meet regulatory and security requirements from sites and from SWOG. Optimally, data within the research data management system would be able to be repurposed by sites to help to meet other needs, including FDA audits, cancer registry submissions, and source data verification.

Next Generation EHR-to-EDC Project

An application that meets SWOG requirements for a potential solution to automate EHR data transfer from the EDC emerged in a cloud-based software application named nCartes™ (nCartes Inc., Fremont, CA). *Cartes* means "maps" in French.[65] Hereafter, nCartes will be referred to as the "application" or the "EHR-to-EDC application." The application is designed to transfer data from the EHR and map to an EDC system. The application met the criteria, listed in **Table 1**, that the SWOG SDMC considered to be essential for a cloud-based multisite research data management system. In 2019, researchers at the University of Kansas Cancer Center worked with the application vendor to perform a proof-of-concept study in which an estimated 50% data entry time and cost savings was achieved using the application to transfer data from the EHR to the study EDC.[66]

Early pilot testing results of the application implementation with SWOG were encouraging. UC Davis timed an experienced CRP who entered 43 forms using the application. The evaluation was performed for fields that auto-populated from EHR data linked to the application. The task of reviewing the 43 forms for submission to the EDC took 100 minutes, or about 2 minutes and 20 seconds per form.[46] The data coordinator supervisor estimated an average time savings of 5 to 15 minutes, depending on form size and complexity, for forms auto-populated using the application. Data for 10 patients across three studies, a total of 93 eCRFs, were collected for the pilot study. Data collected via the application had approximately 1.5 fewer errors per patient then comparison data in the EDC. Fifteen data discrepancies were identified from a total 1,605 data points. Ten of these were related to lab fields.[67]

Within a year after this early evaluation, the EHR-to-EDC application had been used successfully to submit eCRFs in four SWOG studies. In mid-2020, a year after the initial single-site pilot, project teams at three sites worked with the vendor and the SWOG SDMC to design a larger evaluation project to see if use of the application could lead to time savings and improved data quality across multiple sites.

Methods

The goal of the multisite pilot was to compare efficiency of the application to compile and submit study data to the EDC system with the current practice of manual abstraction and data entry directly into EDC forms. Four outcome measures were chosen:

1) **Time savings.** Does the use of the application increase time efficiency for site CRPs versus manual MRA processes?
2) **Error rate comparison for all fields.** Is the data consistent when manual MRA and application-assisted entries for all structured and unstructured fields are compared with a gold standard?

3) **Error rate comparison for structured fields.** Is the data consistent when manual MRA and application automated entries for structured fields are compared with a gold standard?

4) **Interrater reliability.** To what extent do CRPs at the same site agree that they are observing the same data?

Time savings, or loss, was measured as the difference in minutes and seconds taken to enter data using the application to autofill some entries, subtracted from time taken to manually abstract and enter data in the same forms. Stopwatch applications on smart phones were used to measure time taken for each form. If interrupted, the CRP immediately paused the stopwatch and restarted it when able to return to the task.

Error rate was defined as the extent to which evaluation entries were consistent with real-world data represented by a gold standard. For this study, the *gold standard* was defined as data previously entered into the active SWOG study EDC which had been reviewed and validated by the SDMC. Error rate was calculated by dividing the number of identified errors by the number of total fields checked.[42] Percentage reduction in errors using completely manual MRA methods versus application-assisted abstraction was calculated. CRPs at the sites conducted the initial data entry. Personnel of the application vendor performed data comparisons. Errors were verified against the EHR source and tracked in spreadsheets. At the conclusion of the evaluation, discrepancies found in the SWOG study EDC were reported to data coordinators at the SDMC.

Structured data included lab data and date fields available in a standard format that the EHR-to-EDC application was able to automatically capture. *Unstructured data* included all other fields which CRPs must abstract from medical, lab and imaging notes. *Source data* refers to original EHR data.

Manual MRA methods were traditional processes in use by the CRPs. There was no interaction with the EHR-to-EDC application.

Application assistance refers to CRP abstraction of unstructured patient data that was provided within the application interface. It also included the automated abstraction of structured data by the application and transfer to the EDC.

Error rate comparison for all fields. This error rate comparison included all fields, structured and unstructured, and compared the error rate between CRPs using traditional manual MRA processes with the error rate when CRPs used application assistance. Using the application, CRPs selected data from unstructured information, such as from medical records and lab notes, and used application tools to copy it to the appropriate fields. Fields with structured data from labs and the medical record were automatically populated by the application.

Error rate comparison for structured fields. This error rate comparison focused on structured data fields. It involved the comparison of the error rates of CRPs using manual MRA processes versus the use of the application, which automatically collected and transferred structured data from labs and the medical record.

Interrater reliability (IRR). Differences between responses from CRPs at the same site responding to the data may be related to human bias.[42,53] At Sites 2 and 3, percent agreement measures were used to better understand the consistency, or agreement, between the two CRPs.[46,48–50]

Study, Sites, Sample, and Setup

Ten SWOG oncology studies were available to each site with a focus on these cancer types: genitourinary (3 studies), multiple myeloma (1), lymphoma (2), lung (1), leukemia (1), breast (1) and malignant solid tumor (1). Sites implemented those studies in which they had patients enrolled. The multiple myeloma clinical trial was chosen for this evaluation as participating sites had patients active in that study. The study's follow-up tumor assessment form was chosen for the evaluation as the form was required for all patients upon the first visit and alternating visits thereafter. The form was also a good fit for evaluation as it included both structured and unstructured data fields. Three sites that had implemented the EHR-to-EDC application and the myeloma study agreed to participate in the evaluation. For anonymity, sites were randomly assigned a number using an established online list randomizer application.[68]

A convenience sample of six patients per site was used, as one site had exactly six on the multiple myeloma study. Six patients were selected randomly from those enrolled in the study at the other two sites. Because the CRPs needed to find time to work on this evaluation, the use of six patients was also convenient in limiting their time spent.

One of the three evaluation sites completed implementation of the EHR-to-EDC application to establish interoperability using HL7 FHIR before the evaluation. The other two sites had completed HL7 interoperability and were working to implement FHIR. Data mapping between the source EHR and the study forms was completed using USCDI 1.0 and 2.0 standards, which result in consistent source data across sites.[31] Data mapping for individual studies must be validated for each site to address nuances in how and where individual sites store data. To avoid negative impacts from data quality and accuracy, data must be collected systematically and must be harmonized for accurate use in research.[69] Data captures and transfers from multiple EHR sources were checked when acquired by the application and again when transferred from the application to the study EDC. The vendor team worked with CRPs and other content experts at the sites to set up and to verify the accurate acquisition and transmission of data from EHR sources to the application.

Prior to the evaluation, facsimiles of all the myeloma study eCRFs were created in the EHR-to-EDC application to simulate the study in the SWOG study EDC. This was an automated process using an export of the forms that are set up within the EDC system being used for the SWOG study. Each site established a connection with

Goodman et al: Automating Data Entry from Electronic Health Record to Electronic Data Capture Using a Trusted Cloud-Based Application in Multisite Cancer Clinical Trials

Art. 1, page 7 of 16

the myeloma study within the EHR-to-EDC application using HL7 FHIR or a current version of HL7 to enable the transmission of data from the site to the application. CRPs were then able to review and manage information from within the application.

Data for structured fields was automatically mapped to the application's eCRFs. CRPs reviewed the pre-filled data and made changes as needed. For example, if data from an earlier lab test needed to be used as a result of a damaged sample. The CRP could choose the correct blood test results from a pick list available within the application. Unstructured data from the EHR, found in visit or progress notes and in lab or image reports, were also available to the CRPs via application search tools for each form. To collect data from text notes, the CRPs read the notes rendered in the application and then manually populated the remaining unstructured eCRF fields.

To evaluate the EHR-to-EDC application, an evaluation EDC was emulated using a facsimile of the follow-up tumor assessment form within a separate instance of the application. In a real study, outside this controlled evaluation, completed forms would be submitted through the EHR-to-EDC application to the study EDC system electronically via established connections.

CRPs at each site were chosen based on practical considerations. While the use of team members with knowledge of the study protocol was preferred, workload and availability required the use of CRPs with different levels of experience and of familiarity with the protocol. Sites participating in the evaluation were asked to assign one junior-level and one senior-level CRP. Inclusion criteria for CRPs were based on training, years of experience, and availability. Due to staffing restrictions, Site 1 had one junior-level CRP available for the evaluation. Sites 2 and 3 had both a junior- and senior-level CRP available. CRPs were trained in clinical trial work and in the use of the application. They were also trained in the use of EHR data systems at their site to abstract data for the trials used in the evaluation.

Data Collection

CRPs at each site and data coordinators at the SDMC performed quality assurance testing on all study forms, including the follow-up tumor assessment form, within the EHR-to-EDC application. Once the sites and the vendor were satisfied with the eCRFs, the forms were released and the sites began entering patient data using the application. Data collection was conducted by each CRP over two days. On the first day, the CRP entered data for six patients manually directly into the tumor assessment form. On the second day, the same CRP entered data for the same six patients using application-assisted data collection. CRPs selected the two-day period based on their availability. CRPs utilized two versions of the form for each of the six patients. After data collection and processing for the myeloma trial, patient data was entered into form version 1 and then into form version 2. For ease of execution of the evaluation study, both were rendered within the EHR-to-EDC application:

· Form version 1 was an empty form that mimicked the actual EDC follow-up tumor assessment form. The CRP used manual MRA processes to enter data for each patient by referencing and manually abstracting EHR and lab data within those individual data systems and then manually entering the data into the form as is traditionally done in EDC.
· Form version 2 also mimicked the actual EDC form; however, structured data available for the patient was auto-populated using application automation. The CRP referenced clinical notes rendered within the EHR-to-EDC application to enter data into the unstructured fields. The CRP also reviewed the automatically entered structured data fields.

Nine structured data elements were collected on each of two follow-up tumor assessments for each patient (see **Table 2**). There were 11 structured fields on the tumor assessment form; however, the sites did not conduct lab tests for the structured fields for Immunoglobulin D and E serum levels. The remaining nine structured fields comprised 17% of the 53 total fields on the tumor assessment form. The EHR-to-EDC application was able to automatically collect and transfer data for the nine fields.

Table 2: Follow-up Tumor Assessment Form Fields Available for Automated Capture.

Structured field names available	Included in evaluation
Calcium, serum	Yes
Serum calcium date	Yes
Urine volume	Yes
Urine total protein	Yes
Kappa free light chain	Yes
Lambda free light chain	Yes
Immunoglobulin A (IgA), serum	Yes
Immunoglobulin D (IgD), serum	No
Immunoglobulin E (IgD), serum	No
Immunoglobulin G (IgG), serum	Yes
Immunoglobulin M (IgM), serum	Yes

For the six patients at each site, the CRPs entered 636 total fields into the two follow-up tumor assessment forms. All intended fields were captured and evaluable for each form. **Figure 1** shows the total fields in red and the fields available (3,180) for automatic abstraction (540) in blue. CRPs entered 3,180 data fields into the evaluation EDC twice, once using manual MRA and once using the application. All fields, on 60 forms, were used to calculate time savings and interrater reliability. For error rate calculations, five forms at Site 1 and two forms at Site 3 could not be used as they were not previously entered into the SWOG study EDC. The 53 forms used for comparisons comprised 2,809 fields.
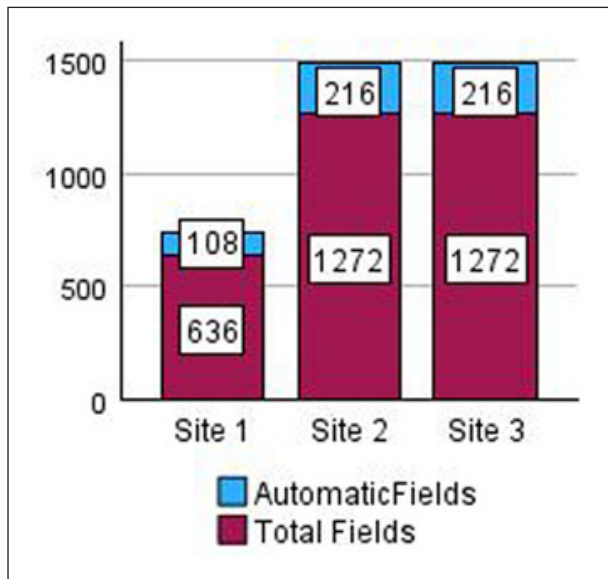
Art. 1, page 8 of 16

Goodman et al: Automating Data Entry from Electronic Health Record to Electronic Data Capture Using a Trusted Cloud-Based Application in Multisite Cancer Clinical Trials

Figure 1: Summary of Overall Total and Automatic Fields by Site.

Results

These results report on the outcome measures for time savings, error rate comparisons, and accuracy of the automated capture and transfer process.

Time Savings Using Assisted Data Entry

Time savings was measured as the difference in time taken to enter data using manual MRA versus the time taken to enter the same data with assistance from the EHR-to-EDC application, including automated entry of nine fields. An average time savings of 2 minutes and 40 seconds was observed for entry of the two follow-up tumor assessment forms (see **Table 3**). This represents a 36% average time saving per form.

Time savings for the individual follow-up forms are shown in **Table 4**. The first follow-up tumor assessment form demonstrated time savings of 1 minute and 31 seconds using the application. The time savings for the second follow-up form was 1 minute and 9 seconds. In most cases, fewer fields were entered on the second form for a patient.

Statistical Analysis

For analysis, data was reviewed and formatted within the spreadsheets before being transferred to a statistical application, where it was checked for consistency and accuracy. Statistical analyses were accomplished using IBM® SPSS® (Statistical Product and Service Solutions) Statistics (Ver. 29). Descriptive statistics were used to report the evaluation results. Descriptive tables and figures were used to display results. Percent agreement was used as the IRR measurement. This decision was supported by McHugh (2012), who noted that percentage agreement may be safely used to determine interrater reliability in situations in which raters are well trained and are not guessing.[49]

Error Rate

Two outcome measures were sought regarding the error rate. One measured the error rates for entry of all fields, unstructured and structured, comparing outcomes for CRPs using manual MRA with no assistance from the application. The other measure was with CRPs entering unstructured data fields using EHR data displayed within the application, plus automatic capture and transfer of nine structured data fields by the application. A total of 47 forms, available for comparison in the SWOG study EDC, were completed across the sites.

When CRPs used manual MRA to enter all fields on the forms, 163 errors were recorded for the 2,809 total fields, a 5.8% error rate (see **Table 5**). When CRPs used the

Table 3: Time Savings (Seconds) for the Combined Two Follow-up Tumor Assessment Forms.

Data Entry Method, Two Forms Per Patient (N = 30)	Min Seconds	Max Seconds	Mean Seconds	Mean Mins/Secs	Std. Dev.
Manual MRA	247	889	446	7 mins, 26 secs	163
Application Assisted	108	547	286	4 mins, 46 secs	102
Time Savings			160	**2 mins, 40 secs**	

Table 4: Time Savings (Seconds) for Entry of Each of the Follow-up Tumor Assessment Forms.

Data Entry Method	Min	Max	Mean	Mean	Std. Dev.
Follow-up Form 1 (N = 30)					
Manual MRA	120	582	253	4 mins, 13 secs	114
Application Assisted	41	344	162	2 mins, 42 secs	77
Time Savings			91	**1 min, 31 secs**	
Follow-up Form 2 (N = 30)					
Manual MRA	88	336	193	3 mins, 12 secs	66
Application Assisted	44	224	124	2 mins, 3 secs	56
Time Savings			69	**1 min, 9 secs**	

Goodman et al: Automating Data Entry from Electronic Health Record to Electronic Data Capture Using a Trusted Cloud-Based Application in Multisite Cancer Clinical Trials

Art. 1, page 9 of 16

Table 5: All Fields Error Rate Reduction Using Manual MRA vs. Application Assistance.

All Fields	Total Errors	Total Fields Checked	Error Rate (%)	Error Rate 95% CI (%)	Error Rate Reduction (%)
Manual MRA	163	2,809	5.8	(4.07,7.19)	—
Application Assisted MRA	34	2,809	1.2	(0.72,1.68)	79.0

Table 6: Structured Fields Error Rate Using Manual MRA vs. Application Assistance.

Structured Fields	Total Errors	Total Fields Checked	Error Rate (%)	Error Rate 95% CI (%)	Error Rate Reduction (%)
Manual MRA	65	477	13.6	(9.12,18.66)	—
Application Automated MRA	0	477	0.00	(0.00,0.01)	100.0

application, total errors were reduced to 34, an error rate of 1.2%. The error rate reduction with assistance from the EHR-to-EDC application was 79%.

For a limited subset of 477 structured data fields, automatically captured and transferred from EHR data, there were 65 errors recorded for manual MRA data entry, a 13.6% error rate; there were zero errors when using the application to automatically enter the data (see **Table 6**). The results confirm that use of the EHR-to-EDC application resulted in an error rate reduction. Several unit conversion errors were observed using manual MRA. These errors are not unusual.[42] When using the application, which processes unit conversions automatically, there were no errors for structured fields.

Interrater Reliability

An analysis of IRR was performed to examine the extent to which CRPs consistently collected identical patient data. The results of analyses are presented for Sites 2 and 3, which each had two CRPs, one junior and one senior, entering the same data for six patients. (Site 1 was not included as it had only one CRP available for the evaluation.) Percentage agreement was determined to be the most appropriate calculation.[49] The IRR analysis focused on the structured data fields as these fields are representative of the entire set of data.

The structured data fields observed for six patients at each site contained a total of 108 fields per site, see **Table 7**. The two CRPs at Site 2, using manual MRA, collected identical data 86.1% of the time. When using manual MRA, the two CRPs at Site 3 collected identical

data 77.8% of the time. When using the EHR-to-EDC application, which automatically collected and transferred the structured data, there was 100% agreement.

Discussion

The purpose of this study was to explore if use of an application to collect EHR data for clinical trials would provide time savings and improved data quality in the operation of a clinical trial. Both were confirmed in this evaluation. In addition, results demonstrated other potential benefits of the application.

Time Savings Using Automatic and Assisted Data Entry

Time savings for members of clinical research teams is an important benefit of EHR-to-EDC technology. Clinical research teams are often overburdened with work and have experienced high turnover and increased stress since the COVID-19 pandemic.[70] In a survey of SWOG clinical research leaders, 80% of respondents reported that their site has experienced workforce shortages.[70] Survey respondents provided ranked reasons for high attrition rates that extend beyond the challenges of functioning during a pandemic. Reasons provided included pay levels, job opportunities, flexible work environment and general burnout.[71]

Using the EHR-to-EDC application to reduce data entry time has been welcomed by leadership and research personnel at participating SWOG sites. In this evaluation, the estimated time savings of 1 minute and 20 seconds, or 36%, for one follow-up tumor assessment form can be very meaningful over the entire sample and visit schedule

Table 7: IRR Percentage Agreement for Structured Fields.

Structured Fields	Cases	Agree	Disagree	Percentage Agreement
Site 2				
Manual MRA	108	94	15	86.1
Application Assisted	108	108	0	100.0
Site 3				
Manual MRA	108	84	24	77.8
Application Assisted	108	108	0	100.0

Art. 1, page 10 of 16

Goodman et al: Automating Data Entry from Electronic Health Record to Electronic Data Capture Using a Trusted Cloud-Based Application in Multisite Cancer Clinical Trials

of a clinical trial. A single patient in the multiple myeloma study could receive treatment and generate forms for up to seven years from registration. Treatment forms could be filled out approximately every two months for a total of about 45 cycles. Over seven years, expanding the time savings of 1 minute and 20 seconds to the full 45 cycles of one repeating form would become 60 minutes per patient for that one form alone.

Substantial time savings would be possible for all forms over the entire life of the trial. This was found to be particularly relevant in this evaluation as time savings results included both the automatic extraction and transfer of structured data fields. This allows an estimate of total fields that could accrue for all forms for a single patient who continues without progression through the full follow-up period of the multiple myeloma study (up to 15 years). **Table 8** shows the number of data fields for all forms over the life of the trial. There are approximately 56 forms with 800 data fields. Some forms are used during treatment, and others for follow up. Other forms are used multiple times during and after treatment. Out of the total 11,514 total data fields that might be collected over the life of the trial, the application could assist with 8,652 fields (75%) of which 1,475 (13%) structured data fields could be auto-populated.

Considering that the application assists with both structured (automatic) and unstructured fields (assisted), time savings can be predicted for all potential forms for one patient over the entire 15-year term of the myeloma trial. **Table 9** shows estimated time savings of 70 seconds per form based on 88% for fields being assisted by using the EHR-to-EDC application. This time savings covers CRP MRA processes; however, time savings combined with improved data quality also decreases time needed to process data corrections through other elements of the clinical trial. Estimated time savings were extrapolated for all forms based the results from this evaluation using the follow-up tumor assessment form.

In the RADaptor pilot study, the substantial 37% time savings for demographic data collection with application assistance was similar to the 36% time savings observed in this evaluation.[59] While not directly comparable with this evaluation, the RADaptor results confirmed early on that applications that utilize interoperability technology can improve the MRA process.

An evaluation by Garza et al. (2024) utilized the same EHR-to-EDC application used in this evaluation. The researchers compared error rates when using traditional manual MRA processes with data collection using assistance from the EHR-to-EDC application. Following the evaluation, participants estimated the potential for up to 50% in time saved for combined MRA, data entry and quality control activities.[72]

Error Rate

Error rate measures accuracy – consistency with the gold standard – the extent to which entries match a real-world state of the data.[42] For all structured and unstructured

Table 8: Total Data Fields Per Patient Over the Full Study, Up to 15 Years.

Form Types	Total Forms	Total Data Fields	Application Assisted Fields Total	Assisted Fields (%)	Automatic Capture Fields[a]	Auto Fields (%)
Setup: Demographic and registration forms	35	448	0	0%	0	0%
On-study: Vital status, baseline, initial treatment	17	297	278	94%	81	27%
Repeated on-study: On-treatment forms[b]	308	5,236	4,048	77%	697	13%
Repeated follow-up forms[c]: Follow-up forms	64	5,533	4,326	78%	697	13%
Totals	424	11,514	8,652	75%	1,475	13%

a. Fields that the EHR to EDC application can auto-populate.
b. Repeated on-treatment forms include vital status, tumor assessment, treatment, and adverse events (AE) forms.
c. Repeated follow-up forms include vital status, late AE, and expedited reporting.

Table 9: Estimated Times Savings for One Patient Over the Full Study.

Form Types	Total Forms	Total Fields	Assisted Total	Assisted Fields (%)	Estimated Seconds Times Savings	Total Time Savings
Estimated Time Savings Using One Form for One Patient[a]						
Follow-up Assessment Form	1	54	53	99%	80	1 min., 20 secs.
Estimated Time Savings for All Forms (424) for One Patient[b]						
All Forms Over 15-year Term	424	11,514	10,127	88%	70	8 hrs., 15 mins.

a. Tumor follow-up assessment form.
b. All forms for one patient who continues through the full follow-up without progression.

Goodman et al: Automating Data Entry from Electronic Health Record to Electronic Data Capture Using a Trusted Cloud-Based Application in Multisite Cancer Clinical Trials

Art. 1, page 11 of 16

fields, use of the EHR-to-EDC application was associated with a 79% decrease in errors versus manual MRA. This improvement was likely related to the direct mapping to the needed data in the EHR and the ability for CRPs to access unstructured EHR data within the application. Rather than repeatedly changing focus between the EHR and the EDC, CRPs used the evaluated application to view and lookup data and apply them to the eCRFs in the application.

The EHR-to-EDC application made no errors in the extraction and transfer of the nine structured data fields on each form (477 fields in total), a notable improvement from the 65 errors (13.6%) observed when using manual MRA. The result is comparable to that found by Garza et al. (2024), with a decrease in errors from 13 to zero when using the same application on structured data fields.[71] Extrapolating the observed difference in error rate across all forms for the entire length of the study indicates that a substantial improvement in data quality is possible.

Cost Reduction
Time savings and fewer errors reduce the cost of a clinical trial. Prevented errors mean that fewer data queries will be generated and that will save time for those working in data collection, data management, data analysis, and elsewhere downstream.[9] While positively impacting the quality of the trial, prevented errors may also be consequential with respect to both real and elapsed time that would otherwise be spent addressing the errors. Managing and responding to data quality queries can require a lot of time throughout the lifecycle of the trial.

This evaluation showed that the EHR-to-EDC application could automatically collect and transfer 13% of structured data fields to the study EDC. Further, for 75% of data fields for which unstructured data was the source, the evaluated application could assist through convenient access to visit notes, progress notes, lab data, and image reports. Data to fill these unstructured fields was abstracted from within the application, saving time compared to manually finding the needed note or report, manually abstracting the data, and manually entering the data.

If 50% of all fields for a clinical trial could be automatically populated, Sundgren et al. (2021) predicted that substantial cost savings could be achieved.[5] An estimated value of time saved for one patient throughout the 15-year study, using the $60 estimated hourly rate by Sundgren et al. (2021) and a time savings of 8 hours and 15 minutes equates to $495 per patient.[5] **Table 9** shows projected time savings for one patient for the full 15-year duration of the trial without progression.

The data accuracy improvements, especially for structured data for which the application-assisted error rate was zero, are particularly valuable. Such data quality improvements are likely to materially reduce the downstream time and cost of source data verification for both sponsors and sites and also potentially reduce elapsed time in trial fulfillment. Data regarding downstream benefits such as these were not quantified in this research.

Interrater Reliability
The IRR percentage agreement measures demonstrated that the use of the evaluated application substantially improved agreement between the two CRPs at Site 1 and Site 2 (where two CRPs were available). The increased IRR observed with use of the EHR-to-EDC application reinforces the improvement of error rate results. As W. Edwards Deming (2018) stated, quality is defined as being "on target with minimum variation".[73] Use of the EHR-to-EDC application to increase IRR represents a notable improvement in the MRA process, which translates to improved agreement between the CRPs at each of the two sites that were evaluated.

Other Observations, Uses, and Technology Evolution
In this evaluation, data quality improvements and time savings have been demonstrated to be benefits for sites to use a trusted cloud-based application to assist with MRA data collection for clinical trials. Time savings may also occur for data management offices as a decrease in data review and query generation and curation is likely.

The greatest time savings and quality improvements seen in this evaluation related to structured data fields, for which the application was able to automate entry for nine (17%) of the form fields. Most oncology studies involve a greater proportion of lab-related data and other such structured data. As a result, CRPs are using the application for other SWOG trials to complete a considerably greater proportion of data entry, which is resulting in materially greater total and percentage time savings. This was also observed in other evaluations.[59,72]

Within SWOG, the EHR-to-EDC application is in production at 12 sites. Sites have been processing data related to six SWOG studies. The application has assisted MRA for more than 160 patients, including more than 3,500 case report forms. In December 2024, five additional sites signed on for a total of seventeen. Four more SWOG trials are being added to the application for a total of fourteen.

As EHR-to-EDC applications are adopted, other benefits will be identified and implemented. For example, the sites can choose to use clinical trial data available in the application to meet other reporting needs, such as audit reports or redacted reports for sponsors. CRPs at participating SWOG sites estimate that automated redacted reports could save 10 to 15 minutes per patient. Further evaluation for this use is merited.

Accuracy of the EHR-to-EDC application's automated data capture and transfer process is a measure of data quality, as defined by Zozus et al. (2023), who noted that it is made up of two components, "representational inadequacy" and "degradation or loss of information".[42] Zozus et al. (2023) also describe "representational inadequacy" as the introduction of inaccuracies through the process of data mapping, the acquisition of data from other sources for the trial record, and inaccuracies made in collecting other data.[42] With zero errors observed for structured data fields that were automatically handled by the application, the application's capture and transfer process has been observed to be accurate.

Art. 1, page 12 of 16

Goodman et al: Automating Data Entry from Electronic Health Record to Electronic Data Capture Using a Trusted Cloud-Based Application in Multisite Cancer Clinical Trials

During the evaluation and review process of the 2,491 fields of evaluation data, 20 errors were discovered in the SWOG study EDC. Following the completion of the evaluation, each of these errors was confirmed to be incorrect when compared with EHR source data. These errors were reported to data coordinators at the SWOG SDMC who generated site queries for CRPs to review and correct following established data quality procedures.

Medication information retrieval is ripe for improvement. While medical record vendors have been adding HL7 FHIR connectivity to their systems, some pharmacy and medication tracking systems have not yet implemented this connectivity, or, in some cases, organizations have not yet upgraded to versions that provide integration. As medication information becomes more easily accessible, it will be easier to collect the data, and the value of that information for clinical trials will increase.[30]

Standardization of clinical data and assigning standardized codes to clinical data will continue to improve the utility of EHR data for use in research.[8,39] The work of the mCODE initiative and similar programs will continue to make EHR data more meaningful and accessible for use in cancer clinical trials. These standards and new technologies may improve the ability to integrate EHR-to-EDC applications and to provide improvements in efficiency and data quality. Added benefits of data standardization efforts will improve the ability to identify patients who can benefit from clinical trials and to tailor treatments based on real-world results of ongoing trials.

Future Research
To fully understand the benefits and challenges of the use of EHR-to-EDC technology to assist in data collection and submission, future research should compare manual and assisted data entry across a larger sample size, across multiple therapeutic areas, and across multiple studies conducted by multiple sponsors. It would also be important to examine the benefits and barriers for implementation of this technology at healthcare institutions of different sizes and settings.

A digital divide exists in which hospitals that treat a marginalized population may not have access to interoperable systems.[55] Conversely, technologies like HL7 FHIR bundled with newer applications may eventually make it easier to use new technologies effectively.[45,74] Research into disparities in digital systems access and use would help to increase understanding as to how to best support all facilities.

Resource availability is another potential barrier. Not all centers may be able to participate in clinical trials due to availability of staff, EHR technology, and cost. While the use of technology has promise to provide a return on investment, resource scarcity, cost of contracting, and security concerns remain potential barriers. Further research in this area may identify how best to help these institutions.

MRA error rates are a challenge to the quality of data collected for clinical trials. Training and quality control checks can lower MRA error rates.[10,46] Further research

could examine the potential of the application to be used to assist with quality control checks and to provide feedback to CRPs that could be used to improve data quality.

Limitations
As a small observational evaluation there are multiple limitations.

Time and resources limited the size of the evaluation in terms of sites, form types, and number of participating CRPs. In an ideal situation, the order of entry would have been randomized.[75] Parallel data entry would have been done both within the evaluated application with assisted automatic entry of available data and manually within the EDC system.[75] Because a second copy of the EDC study database would have required additional effort, a separate set of data entry screens in the evaluated system was used to emulate the EDC forms for unassisted data entry. Any differences between these screens and the study EDC system would offer an alternate explanation for discrepancies and weaken the conclusions. The sample size was limited to six patients at each site. One site had only one CRP available for the evaluation, while the other two sites used two CRPs each.

Junior and senior CRPs collected data for the evaluation at two of the three sites. Junior-level staff are typically trained to work on Phase II and III studies, while senior-level staff typically work on more complex early-stage Phase I studies. The senior-level CRPs were less familiar with the Phase III study selected for the evaluation and might have taken additional time and/or made more errors than the junior-level CRPs more familiar with the trial.

While each site followed the same methods for the evaluation, local site factors may have influenced the results. In an ideal situation, data would be collected with the CRP able to focus completely on the tasks; however, interruptions did occur, as is likely to also occur in a real-world setting. Another potential moderating variable could be related to the CRP's familiarity with the type of cancer (multiple myeloma) being investigated. CRP experience might have introduced a variance in data quality, as might other personnel-related factors such as fatigue, skill, previous knowledge and ability to focus.[42,53] CRPs were well-trained in their job duties, and they were trained in the use of the EHR-to-EDC application prior to the evaluation. That training might have helped to limit some of these potential moderating factors.

Data from a single multiple myeloma clinical trial was examined. While the results are not generalizable, this multiple myeloma trial shares data in common with many oncology trials, and results may be applicable to other therapeutic areas with studies based on similar data.

Variability in data mapping conducted at each site may have influenced the results.

Usability feedback collected from participants consisted of unstructured interviews. Depending on the number of evaluators, structured interviews and/or use of an academic user experience and usability questionnaire would improve the effectiveness of usability feedback.

Goodman et al: Automating Data Entry from Electronic Health Record to Electronic Data Capture Using a Trusted Cloud-Based Application in Multisite Cancer Clinical Trials

Art. 1, page 13 of 16

Conclusion

This study was an important step for the SWOG SDMC to verify that use of an application for automated and assisted MRA was associated with time savings for CRPs and improved data quality. The results were encouraging. SWOG plans to continue to expand the number of sites that use the evaluated application and the number of studies for which the application is available. EHR-to-EDC technologies like the evaluated application will continue to evolve and to become even more effective in supporting research. Additional evaluative studies will help research organizations, and those who depend on resulting data, to make decisions regarding adoption of EHR-to-EDC technology.

Acknowledgements

Competing Interests

One author who contributed to the manuscript and evaluation is employed by nCoup, Inc., owners of nCartes, Inc.

References

1. **Getz KA, Campo RA.** New benchmarks characterizing growth in protocol design complexity. *Ther Innov Regul Sci.* 2018; 52(1): 22–8. DOI: https://doi.org/10.1177/2168479017713039

2. **Blanke CD.** "Data Capture with nCartes: A Leap Day Opportunity." *The Front Line.* SWOG Cancer Research Network. Published February 16, 2024. https://www.swog.org/news-events/news/2024/02/16/data-capture-ncartes-leap-day-opportunity.

3. **Goodman K, Krueger J, Crowley J.** The automatic clinical trial: leveraging the electronic medical record in Multisite Cancer Clinical Trials. *Curr Oncol Rep.* 2012; 14(6): 502–508. DOI: https://doi.org/10.1007/s11912-012-0262-8

4. **Veeva 2020 Unified Clinical Operations Survey Report.** *Veeva Systems.* https://www.veeva.com/wp-content/uploads/2020/09/Veeva-2020-Unified-Clinical-Operations-Survey-Report.pdf. Accessed November 11, 2024.

5. **Sundgren M, Ammour N, Hydes D, Kalra D, Yeatman R.** Innovations in data capture transforming trial delivery. *Appl Clin Trials.* 2021; 30(7/8): 16–20. https://www.appliedclinicaltrialsonline.com/view/innovations-in-data-capture-transforming-trial-delivery. Accessed November 11, 2024.

6. What is an electronic health record (EHR)? HealthIT.gov. Published September 10, 2019. https://www.healthit.gov/faq/what-electronic-health-record-ehr. Accessed November 11, 2024.

7. **Mc Cord KA, Ewald H, Ladanie A,** et al. Current use and costs of electronic health records for Clinical Trial Research: A descriptive study. *CMAJ Open.* 2019; 7(1). DOI: https://doi.org/10.9778/cmajo.20180096

8. **Bertagnolli MM, Anderson B, Quina A,** et al. The electronic health record as a clinical trials tool: opportunities and challenges. *Clin Trials.* 2020; 17(3): 237–242. DOI: https://doi.org/10.1177/1740774520913819

9. **Cheng AC, Banasiewicz MK, Johnson JD,** et al. Evaluating automated electronic case report form data entry from electronic health records. *J Clin Transl Sci.* 2022; 7(1). DOI: https://doi.org/10.1017/cts.2022.514

10. **Garza MY, Williams T, Myneni S,** et al. Measuring and controlling medical record abstraction (MRA) error rates in an observational study. *BMC Med Res Methodol.* 2022; 22(1). DOI: https://doi.org/10.1186/s12874-022-01705-7

11. Real-world data: Assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products guidance for industry. *U.S. Food and Drug Administration.* https://www.fda.gov/media/152503/download. Accessed November 11, 2024.

12. Framework for FDA's Real-world Evidence Program. *U.S. Food and Drug Administration.* https://www.fda.gov/media/120060/download. Accessed November 11, 2024.

13. Framework for FDA's real-world evidence program. *U.S. Food and Drug Administration.* December 2018. https://www.fda.gov/media/120060/download. Accessed November 11, 2024.

14. **Concato J, Corrigan-Curay J.** Real-world evidence – where are we now? *N Engl J Med.* 2022; 386(18): 1680–1682. DOI: https://doi.org/10.1056/NEJMp2200089

15. **Clinical Data Acquisition Standards Harmonization (CDASH).** *Clinical Data Interchange Standards Consortium (CDISC).* https://www.cdisc.org/standards/foundational/cdash. Accessed November 11, 2024.

16. **Metadata services for cancer research.** *National Cancer Institute Center for Biomedical Informatics and Information Technology.* https://datascience.cancer.gov/resources/metadata. Accessed November 11, 2024.

17. **Hemkens LG.** Commentary on Bertagnolli et al.: Clinical trial designs with routinely collected real-world data—issues of data quality and beyond. *Clin Trials.* 2020; 17(3): 247–250. DOI: https://doi.org/10.1177/1740774520913845

18. **Minimal Common Oncology Data Elements (mCODE).** *Health Level 7 (HL7) International.* https://confluence.hl7.org/display/COD/mCODE. Accessed November 11, 2024.

19. **Osterman TJ, Terry M, Miller RS.** Improving cancer data interoperability: the promise of the Minimal Common Oncology Data Elements (mCODE) initiative. *JCO Clin. Cancer Inform.* 2020; 4: 993–1001. DOI: https://doi.org/10.1200/CCI.20.00059

20. **ICAREdata: expanding oncology clinical research capabilities project.** *Brigham Clinical & Research News.* https://bwhclinicalandresearchnews.org/2021/08/06/icaredata-expanding-oncology-clinical-research-capabilities. Accessed December 7, 2024.

21. Fighting Cancer with standard health records. Published February 7, 2019. https://www.mitre.org/news-insights/employee-voice/fighting-cancer-standard-health-records. Accessed November 11, 2024.

22. **Ross JS, Dhruva SS, Shah ND.** Commentary on Bertagnolli et al.: Leveraging electronic health record data for clinical trials—a brave new world. *Clinical Trials* (London, England). 2020; 17(3): 243–246. DOI: https://doi.org/10.1177/1740774520913850

23. **Macumber C.** What is HL7 plus introduction to product lines. *HL7.org.* https://blog.hl7.org/what_is_hl7_and_introduction_to_products_fhir_cda_v2. Accessed November 11, 2024.

24. **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. DOI: https://doi.org/10.1002/lrh2.10076

25. FHIR fact sheets. HealthIT.gov. https://www.healthit.gov/topic/standards-technology/standards/fhir-fact-sheets. Accessed November 11, 2024.

26. Welcome to FHIR®. HL7.org. http://hl7.org/fhir/index.html. Accessed November 11, 2024.

27. **Fast Healthcare Interoperability Resources (FHIR) standard.** National Institutes of Health. Published July 30, 2019. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-19-122.html. Accessed November 11, 2024.

28. **Accelerating clinical care and research through the use of the United States Core Data for Interoperability (USCDI).** *National Institutes of Health.* Published July 30, 2020. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-20-146.html. Accessed November 11, 2024.

29. What is FHIR? *The Office of the National Coordinator (ONC) for Health Information Technology.* https://www.healthit.gov/sites/default/files/2019-08/ONCFHIRFSWhatIsFHIR.pdf. Accessed November 11, 2024.

30. **Eisenstein EL, Zozus MN, Garza MY,** et al. Assessing clinical site readiness for electronic health record (EHR)-to-electronic data capture (EDC) automated data collection. *Contemp Clin Trials.* 2023: 128. DOI: https://doi.org/10.1016/j.cct.2023.107144

31. **Accelerating clinical care and research through the use of the United States Core Data for Interoperability (USCDI).** *National Institutes of Health.* Published July 30, 2020. https://www.healthit.gov/isp/united-states-core-data-interoperability-uscdi. Accessed November 11, 2024.

32. **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.* 2021; 2020: 472–481. Published 2021, Jan 25. DOI: https://doi.org/10.3233/SHTI210188

33. **Finney Rutten LJ, Ruddy KJ, Chlan LL,** et al. Pragmatic cluster randomized trial to evaluate effectiveness and implementation of enhanced EHR-facilitated cancer symptom control (E2C2). *Trials.* 2020; 21(1). DOI: https://doi.org/10.1186/s13063-020-04335-w

34. **Thompson J, Hu J, Mudaranthakam DP,** et al. Relevant word order vectorization for improved natural language processing in electronic health records. *Scientific Reports.* 2019; 9(1): 9253. DOI: https://doi.org/10.1038/s41598-019-45705-y

35. **Chowdhary KR.** Natural language processing. In: Chowdhary, KR, ed. *Fundamentals of Artificial Intelligence.* New Delhi: Springer Nature India; 2020: 603–649. DOI: https://doi.org/10.1007/978-81-322-3972-7_19

36. **Zozus MN, Sanns W, Eisenstein E.** Beyond EDC. *Journal of the Society for Clinical Data Management.* 2021; 1(1): 6, pp. 1–22. DOI: https://doi.org/10.47912/jscdm.33

37. **Wand Y, Wang RY.** Anchoring data quality dimensions in ontological foundations. *Communications of the ACM.* 1996; 39(11): 86–95. DOI: https://doi.org/10.1145/240455.240479

38. Investigator recordkeeping and record retention, 312.62. Code of Federal Regulations, Title 21. https://www.ecfr.gov/current/title-21/chapter-I/subchapter-D/part-312/subpart-D/section-312.62. Accessed November 11, 2024.

39. **Feehan AK, Garcia-Diaz J.** Investigator responsibilities in Clinical Research. *Ochsner Journal.* 2020; 20(1): 44–49. DOI: https://doi.org/10.31486/toj.19.0085

40. ICH E6 (R2) good clinical practice guidance for industry: Investigator responsibilities — protecting the rights, safety, and welfare of study subjects. Good Clinical Practice Network. https://ichgcp.net/4-investigator. Accessed November 11, 2024.

41. **NCRA policy statement on monitoring changes in cancer registry operations.** National Cancer Registrars Association. https://www.ncra-usa.org/Portals/68/PDFs/Informatics/Monitoring%20Changes%20in%20Cancer%20Registry%20Operations_2017.pdf. Accessed November 11, 2024.

42. **Zozus MN, Kahn MG, Weiskopf NG.** Data quality in clinical research. In: Richesson, R, Andrews, JE, Hollis KF (eds.), *Clinical Research Informatics.* 3rd ed. Cambridge, MA: Springer; 2023: 169–198. DOI: https://doi.org/10.1007/978-3-031-27173-1_10

43. **Weiskopf NG, Weng C.** Methods and dimensions of electronic health record data quality assessment: Enabling Reuse for Clinical Research. *J Am Med Inform Assoc.* 2013; 20(1): 144–151. DOI: https://doi.org/10.1136/amiajnl-2011-000681

44. **Russell LB, Huang Q, Lin Y,** et al. The electronic health record as the primary data source in a

Goodman et al: Automating Data Entry from Electronic Health Record to Electronic Data Capture Using a Trusted Cloud-Based Application in Multisite Cancer Clinical Trials

Art. 1, page 15 of 16

pragmatic trial: A case study. *MDM*. 2022; 42(8): 975–984. DOI: https://doi.org/10.1177/0272989 X211069980

45. **Mc Cord KA, Hemkens LG.** Using electronic health records for clinical trials: Where do we stand and where can we go? *CMAJ*. 2019; 191(5). DOI: https://doi.org/10.1503/cmaj.180841

46. **Yawn BP, Wollan P.** Interrater reliability: completing the methods description in medical records review studies. *Am J Epidemiol*. 2005; 161(10): 974–7. DOI: https://doi.org/10.1093/aje/kwi122

47. **Nurjannah I, Siwi SM.** Guidelines for analysis on measuring interrater reliability of nursing outcome classification. *Int J Res Med Sci*. 2017; 5(4): 1169–75. DOI: https://doi.org/10.18203/2320-6012.ijrms 20171220

48. **Liddy C, Wiens M, Hogg W.** Methods to achieve high interrater reliability in data collection from primary care medical records. *Ann Fam Med*. 2011 Jan–Feb; 9(1): 57–62. DOI: https://doi.org/10.1370/afm.1195

49. **McHugh ML.** Interrater reliability: The kappa statistic. *Biochem Med* (Zagreb). 2012; 22(3): 276–282. DOI: https://doi.org/10.11613/BM.2012.031

50. **Zhao X, Feng GC, Ao SH, Liu PL.** Interrater reliability estimators tested against true interrater reliabilities. *BMC Med Res Methodol*. 2022; 22(1): 232. Published Aug 29 2022. DOI: https://doi.org/10.1186/s12874-022-01707-5

51. **de Vet HCW, Mokkink LB, Terwee CB,** et al. "Clinicians are right not to like Cohen's κ." *BMJ*. 2013 Apr 12; 346(f2125): 1–7. DOI: https://doi.org/10.1136/bmj.f2125

52. **Sim J, Wright CC.** The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005; 85(3): 257–268. DOI: https://doi.org/10.1093/ptj/85.3.257

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

54. **Nordo A, Eisenstein EL, Garza M, Hammond WE, Zozus MN.** Evaluative outcomes in direct extraction and use of EHR data in clinical trials. *Stud Health Technol Inform*. 2019; 257: 333–40. PMID: 30741219.

55. eResearch CTMS. https://www.wcgclinical.com/technologies/eresearch-ctms. Accessed November 11, 2024.

56. **Yamamoto K, Yamanaka K, Hatano E,** et al. An eClinical trial system for cancer that integrates with clinical pathways and electronic medical records. *Clinical Trials*. 2012; 9(4): 408–417. DOI: https://doi.org/10.1177/1740774512445912

57. **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: 516–20. DOI: https://doi.org/10.3233/978-1-61499-830-3-516

58. **Manabe S, Takeda T, Hattori A,** et al. Practical use of a multicenter clinical research support system connected to electronic medical records. *Comput Methods Programs Biomed*. 2021; 210: 106362. DOI: https://doi.org/10.1016/j.cmpb.2021.106362

59. **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. DOI: https://doi.org/10.1016/j.ijmedinf.2017.04.015

60. **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. PMID: 30741183.

61. **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; 106: 17–24. DOI: https://doi.org/10.1016/j.ijmedinf.2017.06.006

62. **Garza M, Myneni S, Fenton SH, Zozus MN.** eSource for standardized health information exchange in clinical research: a systematic review of progress in the last year. *JSCDM*. 2021: 1(2). DOI: https://doi.org/10.47912/jscdm.66

63. Health Information Privacy. HHS.gov. https://www.hhs.gov/hipaa/index.html. Accessed November 11, 2024.

64. The HIPAA Privacy Rule. HHS.gov. https://www.hhs.gov/hipaa/for-professionals/privacy/index.html. Accessed November 11, 2024.

65. nCartes EHR to EDC. https://ncartes.ncoup.com. Accessed November 11, 2024.

66. **Mudaranthakam DP, Thompson J, Streeter D,** et al. Connecting the supply chain. Poster presented at: *Association of American Cancer Institutes Annual Meeting*; July 12–14, 2019; Chicago, IL. https://www.aaci-cancer.org/Files/Admin/CRI/2019-Poster-65.pdf. Accessed November 11, 2024.

67. **Cook C, Weatherbee D, Smith A,** et al. A more efficient approach to clinical trial data collection: The SWOG-nCartes pilot collaboration. Poster presented at: *Society for Clinical Trials Annual Meeting*; May 15–18, 2022; San Diego, CA. https://www.crab.org/wp-content/uploads/2024/01/2022-SCT-Cook.pdf. Accessed November 11, 2024.

68. **Haahr, M.** True random number service. *Random.org List Randomizer*. https://www.random.org. Accessed November 11, 2024.

69. **Kahn MG, Callahan TJ, Barnard J,** et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *eGEMs*. 2016; 4(1): 18. DOI: https://doi.org/10.13063/2327-9214.1244

70. **Sun G, Dizon DS, Szczepanek CM,** et al. Crisis of the clinical trials staff attrition after the COVID-19 pandemic. *JCO Oncology Practice*. 2023; 19: 533–535. DOI: https://doi.org/10.1200/OP.23.00152

71. **Dizon DS, Szczepanek CM, Petrylak DP,** et al. National impact of the COVID-19 pandemic on clinical trial staff attrition: results of the SWOG Cancer Research Network Survey of Oncology Research Professionals. *J*

Art. 1, page 16 of 16

Goodman et al: Automating Data Entry from Electronic Health Record to Electronic Data Capture Using a Trusted Cloud-Based Application in Multisite Cancer Clinical Trials

Clin Oncol. 2022; 40: 11049–11049. DOI: https://doi.org/10.1200/JCO.2022.40.16_suppl.11049

72. **Garza MY, Spencer C, Hamidi M,** et al. Comparing the accuracy of traditional vs. FHIR®-based extraction of electronic health record data for two clinical trials. *Stud Health Technol Inform.* 2024: 316: 1368–1372. DOI: https://doi.org/10.3233/SHTI240666

73. **Deming WE.** *Out of the Crisis.* Cambridge (MA): The MIT Press; 2018 (1982). DOI: https://doi.org/10.7551/mitpress/11457.001.0001

74. **Everson J, Patel V, Bazemore AW, Phillips RL.** Interoperability among hospitals treating populations that have been marginalized. *Health Services Research.* 2023; 58(4): 853–864. DOI: https://doi.org/10.1111/1475-6773.14165

75. **Zozus MN, Choi BY, Garza MY,** et al. Collaborative program to evaluate real world data for use in clinical studies and regulatory decision making. *AMIA Jt Summits Transl Sci Proc.* 2023 Jun 16: 632–41. eCollection 2023. PMID: 37350921.