



REVIEW ARTICLE

Beyond EDC

Meredith Nahm Zozus*, Bill Sanns* and Eric Eisenstein†

Introduction: The last two decades have seen Electronic Data Capture (EDC) scale the adoption curve in the therapeutic development industry.

Objective: This synthesis documents the progression to EDC in clinical research towards improving the development, implementation, and use of emerging and similar technology in multicenter clinical studies.

Methods: The 158 full-text sources identified in the Good Clinical Data Management Practices (GCDMP) systematic literature search were used as the basis for this scoping review. A single reader evaluated this literature to determine rationale for and barriers to EDC implementation as well as remaining gaps in the use of EDC technology in clinical research. Quantitative evaluations were identified and summarized. Two reviewers examined the findings to identify omitted seminal work and alternate perspectives.

Results: Metrics and outcomes used in EDC evaluation are detailed. Three major shortfalls in current EDC implementation remain. Similarities and differences between EDC and new clinical study technology are highlighted.

Conclusions: The variation in outcome measures described in the limited accounts of EDC evaluation in the literature precluded meta-analysis. Though improved over the last two decades, EDC performance has seemingly plateaued. If successful, developing standards may significantly broaden opportunities for therapeutic development beyond those available in today's EDC. In addition, multiple adjunct new technologies have potential to edge past the current performance plateau. Of these, direct extraction and use of Electronic Health Record (EHR) data to populate study databases parallels EDC in terms of workflow disruption and technology interaction at clinical sites. Thus, similar adoption challenges are expected.

Keywords: Electronic Data Capture; EDC; Data Quality; Technology Adoption; Technology evaluation

Introduction

The first revolutionizing development in clinical research data collection and management was the use of structured paper data collection forms to facilitate consistent data collection across multiple clinical sites. Use of computers to organize, process, and store data was the second. Web-based Electronic Data Capture (EDC) was arguably the third. The progression to web-based EDC has taken 40 years. Multiple reasons have been articulated for the slow evolutionary development of data collection and management methods. However, neither the history of this important evolution in clinical research nor the reasons for the apparent protracted adoption have been systematically synthesized and reported.

Today, we stand at the cusp of the adoption of direct data acquisition from site EHRs in multicenter, prospective, longitudinal clinical studies. This new data acquisition option has been variously referred to as EHR

eSource, EHR2EDC, and EHR-to-eCRF data collection. Data collection such as this requires the ability to identify study data in EHRs, request the data from the EHRs, reformat the data for the study, and transfer the data into the study database.¹ The interoperability with clinical site EHRs required for direct data collection from EHRs is likely the next major advance in clinical research data collection and management. While varying workflows and data flows are being pursued, EHR-to-eCRF and EDC adoption are similar in that both involve significant technology, process, and behavior changes at clinical sites and study sponsors alike. Learning from EDC adoption experience will help EHR-to-eCRF adoption proceed with less risk and return value more quickly.

Background

Although more broad in literal meaning, the label Electronic Data Capture within the therapeutic development industry has historically referred to manual key entry, automated discrepancy identification, and manual discrepancy resolution. EDC enables these functions at geographically distributed sites using web-based software (i.e., an EDC system). The fundamental shift enabled with web-based EDC was decentralized entry

* Joe R. and Teresa Lozano Long School of Medicine, University of Texas Health Science Center, San Antonio, TX, US

† Emeritus Duke Clinical Research Institute, Duke University, Durham, NC, US

Corresponding author: Meredith Nahm Zozus (zozus@uthscsa.edu)

and centralized organization of data processing, updating, and storage. Core functionality available in most web-based EDC systems today includes the ability for a data manager to (1) design and maintain screens for data entry via the internet; (2) add and maintain univariate and complex multivariate rules to check for discrepant data; (3) develop rules for alerts and conditional form behavior such as adding fields or forms based on user entered data; (4) import and export data; (5) store and retrieve data; (6) implement and maintain role-based privileges; and (7) track and report status of data entry and processing. Some systems have additional functionality supporting randomization of study participants, assigning controlled terminology to data, and collection of data through patient-completed questionnaires. Support for specialized data collection and management have also been reported and include centralized image interpretation, classification of clinical events, management of serious adverse events in studies, and management of source data when special requirements^{2,3} are met.

Since EDC has largely replaced collection of structured data on paper forms, it is no surprise that EDC led the ranked-list of implemented data systems in the recent eClinical Landscape Survey in which all of the 257 eligible respondents reported use of web-based Electronic Data Capture.⁴ Responding companies reported managing 77.5% of their data volume in EDC systems.⁴ The most common types of data managed in EDC systems included eCRF data (100%), local lab data (59.5%), and Quality of Life data (59.5%). Companies also reported use of EDC systems to process Patient Reported Outcomes (ePRO) data (34.2%), Pharmacokinetic data (33.9%), and Biomarker data (28%).⁴

Methods

The systematic literature review supporting the revision of the Good Clinical Data Management Practices (GCDMP) EDC Chapters identified many articles of historical importance but of limited value to inform present EDC practices. However, the lessons learned from the adoption of web-based EDC may be helpful in the design, implementation, and evaluation of other technological innovations involving data collection and processing in clinical studies. Toward this objective, literature identified in the reviews supporting the recent revision of the three GCDMP EDC chapters was leveraged for the historical review reported here. The literature search criteria used and described in the recent GCDMP EDC chapter revision was executed on the following databases: PubMed (777 results), CINAHL (230 results), EMBASE (257 results), Science Citation Index/Web of Science (393 results), and the Association for Computing Machinery (ACM) Guide to the Computing Literature (115 results). A total of 1772 works were identified through the searches, which concluded on February 8, 2017. Search results were consolidated to obtain a list of 1368 distinct articles. The searches were not restricted based on the date of the work or publication. Two reviewers from the GCDMP EDC writing group screened each abstract to identify articles written in the English language and describe EDC implementation or use in clinical research. Disagreements

regarding whether an abstract met these criteria were adjudicated by the GCDMP EDC writing group. Forty-nine abstracts meeting inclusion criteria were selected for full text review. The selected works (mostly journal articles) yielded 109 additional sources. While still mostly journal articles, the works identified through references in the initial batch consisted of a higher proportion of works from trade publications than did the initial batch. Eighty-five works were identified as relevant to EDC. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram for the review is published in each of the three EDC GCDMP Chapters.

Three new articles were identified in the course of the work reported here, and one previously identified but heretofore irretrievable work was obtained for a total of 89 included sources in this review. The full text of each of the 158 identified works was read by a single reviewer to identify events, descriptions, or research results relevant to the history of EDC adoption, including rationales for and barriers to adoption of EDC for the work reported here. EDC benefits were identified and categorized according to the mechanism by which each added value to design, conduct, and reporting of clinical studies. Challenges encountered in EDC implementation were identified and partitioned into those that have largely been overcome versus those remaining today. The latter were enumerated as gaps in the use of EDC today. Information about evaluation design, implementation context, and outcome measures were abstracted from works reporting quantitative evaluations of EDC. Quantitative evaluations were categorized according to the research design, controls, and comparators employed in the evaluation of EDC. Metrics and outcome measures used in the quantitative evaluations were enumerated and classified by the construct assessed. Heterogeneity in reported evaluation methods and outcome measures precluded quantitative meta-analysis of the studies identified through the literature search. Lastly, similarities between EDC and emerging technologies, with particular emphasis on acquisition of data directly from site EHRs, were identified. The synthesis was reviewed by two reviewers working in clinical research informatics for the duration of the EDC adoption curve to identify and correct significant omissions and unsupported assertions.

Results

Remote Data Entry (RDE) as the Predecessor to Web-based EDC

The earliest use of computers in research at clinical sites likely occurred in 1963.⁵ The earliest reports of data collection via computers at clinical sites described locally developed systems.⁶ An example early implementation is described in "Development of a Computerized Cancer Data Management System at the Mayo Clinic."⁷

The foundations of distributed data entry in multicenter clinical studies originated before the advent of the internet. Starting in the early 1970s, sporadic reports of development and implementation of computer systems for Remote Data Entry (RDE) at clinical sites appeared in the literature. The earliest reported use of remote data entry in a multicenter clinical study was in the National

Institutes of Health (NIH)-sponsored Kidney Transplant Histocompatibility Study in 1973, in which two of the twenty-five sites piloted data entry via remote terminals connected by telephone lines to a mainframe at the data center.^{8,9}

In the mid-1980s RDE use expanded with availability of microprocessors.^{10–15} RDE use in multicenter studies was a sponsor or coordinating center driven process in which sites were provisioned with computers, data entry software, and a data transmission device such as floppy discs or a modem. Early RDE implementations distributed database management to the clinical sites with varying levels of responsibilities.¹⁶ Some implementations required sites to first complete paper CRFs and afterward perform first data entry with limited edit checks. Other implementations included centrally performed second entry or additional centrally run edit checks. While others relied upon single entry, with sites having varying amounts of local edit checking during or shortly after entry, and differing extents of batch or central batch edit checks afterwards.^{8,9} Early registries from the 1970s through the 1990s similarly employed systems where data were entered at clinical sites on thick- or full clients, i.e., software applications where most or part of the program was stored on the user's computer with subsequent transfer of data to a central database.

Early commercial systems based on a distributed model (transferring data by media or modem) started appearing in the 1990s.⁹ The earliest of these commercial systems required distribution of laptop computers to investigational sites. These off-line systems were vulnerable to local data loss and, due to the distributed nature, were more difficult to secure. The expense of hardware distribution to sites, especially in multiple countries, was cost prohibitive and logistically challenging. With few exceptions, RDE largely failed to show significant improvement over traditional paper-based data collection and central processing.¹⁷ Clinical data capture was not listed among major improvements in clinical trial procedures and implementation realized during the 1980s.¹⁸

Collen (1990),⁶ Lampe and Weiler (1998),¹⁹ Hyde (1998),²⁰ Kubick (1998),²¹ and Prove (2000)²² provide more complete historical descriptions of RDE. Reports of RDE were published up to the turn of the century^{11,23–28} when RDE ultimately gave-way to web-based approaches.

Web-based EDC

Three main advances in Internet technology made web-based EDC as we know it today possible: (1) internet enabled connectivity of geographically distributed sites, (2) executable code embedded within web pages, and (3) central database storage with decentralized retrieval of information via the internet.²⁹ Early use of web-based forms to collect data over the internet first started appearing in 1995 and continued increasing in frequency to the turn of the century.^{29–38} Several provided descriptions of early system architecture and design.^{29,34,36,39–41}

The first large web-based trial was initiated in 1997.³⁸ A fifteen-country rare disease trial reported in 1999²⁹ was likely initiated in the same time frame. The first surgical web-based trial was reported in 2002.⁴² These

early web-based EDC attempts used locally developed web-based systems. During this time period, early speculation and sporadic reports of cost and time savings associated with web-based EDC piqued the interest of the therapeutic development industry and the number of commercially available systems grew. Industry-sponsored studies primarily used commercially available systems. However, reports of locally developed web-based systems with varying combinations of functionality continued to be reported at academic centers through 2010.^{42,43}

Even though 86% of respondents in a 2002 survey of large pharmaceutical companies agreed that “new technologies would have great impact in the near future,” and 56% agreed that there was “already a high level of urgency within their organizations to adopt new technologies,” adoption of web-based EDC remained low at the start of the new millennium.⁴⁴ At the time, it was widely recognized that “reliance on traditional paper-based processes” in clinical trial data collection commonly incurred a three to four month delay before information became available.⁴⁵ Others reported longer lag times in data availability when comparing use of EDC electronic Case Report Forms (eCRFs) to using traditional paper-based data collection.^{46–48} In the same year, an Association of Clinical Research Professionals (ACRP) survey of 2,300 respondents found that 66%, 60%, and 50% of sites, CROs, and Sponsors, respectively, expected to adopt eCRFs within two years.⁴⁴

EDC adoption progressed more slowly than the early surveys predicted. Reported reasons included organizational lack of strategic planning, the customized requirements of each trial, the immaturity and fragmentation of commercially available EDC software, lack of scalability from pilots to enterprise adoption, and lack of addressing both process and organizational change.^{44,45,49} The lack of process change needed to implement EDC has been attributed to lack of leveraging of the technology to re-engineer clinical study processes, lack of organizational strategy and planning to move from pilots to enterprise adoption, and inability to measure success.⁴⁵ Further, failure to change historic delineations of job roles and responsibilities has been cited as a significant factor inhibiting process re-engineering.⁴⁹ Additionally inhibiting the introduction of new technologies was the potential of new technology to create unanticipated bottlenecks in processes. For example, new systems may be larger and more complex, increasing the need for training or the need for specialists.⁵⁰ More complex systems result in fewer team members understanding the workflow, data flow, or other operations of the system, increasing the potential for mistakes and longer times to troubleshoot problems. These situations would, in turn, require more specialists or increase the number of individuals needed to configure, test, implement, troubleshoot, and maintain systems.⁵⁰ EDC study start-up required more up-front work and data management staff.⁵¹ Lack of leadership mandate or support and lack of alignment throughout the organization have also been faulted for slower-than-projected adoption, as well as lack of delighting stakeholders (sites) with an improved trial experience.⁴⁹

By year 2001, only 5% of new clinical trials reported using EDC or Remote Data Entry.⁴⁹ In the same time frame, CenterWatch reported that only 16% of sites were required to use EDC (probably inclusive of RDE).⁵² In the concurrent literature others also lumped both RDE and web-based EDC under the label “EDC”.^{9,17,45} By year 2004, however, 70% of sites reported using an EDC system.⁵² El Emam et al. (2009) estimated that by the 2006–2007 time frame, 41% of Canadian trials were using EDC.⁵³ Varied EDC system designs and architectures were reported during this time. Examples of these include both client-server-based systems on laptops,⁵⁴ and wholly internet-based approaches developed for single studies,^{55–57} as well as use of commercial web-based EDC platforms.^{41,58–60}

Commercial EDC platforms were more heavily used by industry than academia. Disproportionately fewer studies were published by industry authors, however, because information about the methods and technology industry utilized were often considered proprietary. Thus, the earliest reports were predominantly from academic institutions and not reflective of industry practices. Broad adoption of web-based EDC at academic centers lagged in favor of basic general-purpose or office applications such as spreadsheets and inexpensive pseudo-relational databases.⁶¹ Delineated reasons for slower adoption in academia primarily stemmed from lack of motivation or resources to scale tools built for single studies, high cost of early commercial solutions compared to low-budget small investigator-initiated studies, and lack of institutional support to defray costs to academic investigators.^{62,63} Later, free solutions such as REDCap® filled the gap in academia by providing free publicly available software meeting the needs of academic researchers.⁶⁴ Multiple such in-house solutions were developed and used at academic institutions. Today, the REDCap® system is nearly ubiquitously adopted at clinically-oriented academic

institutions in the United States (and likely abroad) for EDC. An important difference between commercial and academic EDC platform usage is that although the technical controls requisite in Title 21 CFR Part 11 are often present in software developed and implemented at academic institutions, however, in our experience, they are not routinely locally validated to Part 11 standards. In contrast, commercial EDC platforms serving the therapeutic development industry meet these regulatory requirements for validation.

Perceived Benefits of EDC

Benefits of web-based EDC Implementation such as cost savings, more timely availability of data, increased accuracy, and fewer queries (**Tables 1 and 2**) have been widely touted.^{29,65–67} These and other benefits stem from the internet providing *both* (1) rapid and continuous central data acquisition from geographically distributed sites *and* (2) immediate and decentralized data use. Internet provided connectivity allowed common information system benefits such as shared data use, decision support, automation, and knowledge generation across geographically distributed sites and teams in multicenter studies. Thus, the value of EDC to an organization increases as more data, especially data traditionally managed in separate systems, become available via EDC systems.^{65,68} Echoing what process engineers have been saying for the past several decades, although web-based EDC implementations will improve most data collection scenarios, their real benefit arises when processes are re-designed to take advantage of having real-time information simultaneously available to all research team members.³⁷

Reported benefits of EDC are classified here according to mechanisms through which health information technology creates value.⁶⁹ In addition to the categorization from this review (**Tables 1 and 2**), early advantages and disadvantages of EDC have been particularly well articulated by Marks et al.³⁸

Table 1: Reported Benefits of EDC Over Paper and RDE Data Collection in Study Start-up.

Benefits During Study Start-up	Mechanism
Elimination of paper Case Report Form (CRF) printing and physical distribution ^{38,42,51,70}	A
Elimination of filing, storage, and retrieval tasks associated with paper CRFs and study documentation ^{38,66,71,72}	A, P
Making use of sites' existing “office grade” computers ^{29,38}	C, E
Making use of sites' existing “office grade” web-browser software in a platform independent manner. ^{29,38,42,73,74}	C
Elimination of time needed and error associated with manual generation of an annotated CRF through maintaining association of form fields with logical database storage location as a natural by-product of the database set-up ⁷⁵ .	A
Elimination of printing and physical distribution of study-related information, training materials, and job aids. ^{38,41,42,66,73}	A
Use of data from previous studies to identify high performing sites or inform planning with regards to volume, task time, and elapsed time expectations. ⁷⁶	KG, DS

A: Automation; C: Connectivity; DS: Decision Support; E: Making use of less expensive or pre-existing equipment; KG: Knowledge Generation from data such as from data mining; P: Physical space savings; R: Relocation of work tasks to more efficient, better skilled, or less expensive individuals.

Table 2: Reported Benefits of EDC Over Paper and RDE Data Collection in Study Conduct.

Benefits During Study Conduct	Mechanism
Online Treatment Allocation (Randomization) ^{38,51} and associated "Improvements in data validation at the input stage ensure that fewer invalid patients are initially signed up for the trial" ⁴⁵	A
Use of an information system instead of or to enforce manual processes:	
• Enables standardization and application of desired processes through workflow automation and control. ⁷⁷	A
• Reduces illegible fields or symbols that require interpretation. ^{38,78}	R
• Automates generation of process information such as date and time stamps and user identifiers for all actions performed in the system. ³⁸ Tracking becomes a by-product of a task being performed.	A
• Enables use of electronic signatures ⁵¹	A
Relocation of data entry to clinical sites:	
• Eliminates double-key entry (double data entry) related tasks ⁵¹	A, R
• Eliminates intermediate transcription steps, i.e., onto a paper form between original data collection and data that are used for statistical analysis. ^{38,42,50,66,70,72,75,79}	A
• Facilitates entry of data at sites by staff most familiar with the data, likely resulting in more accurate data input. ^{43,75,77}	R
• Facilitates entry of data closer (in time, space, and process steps) to the source where discrepancies can be identified and corrected before commitment to the database. Close proximity between data source and data processing may lead to more accurate data ^{43,75,77}	R
• Eliminates manual CRF retrieval, transmittal, and courier tracking. ⁴⁶	
Automation of discrepancy identification and relocation of discrepancy resolution to sites:	
• Enables immediate correction of missing, out of range, and inconsistent data identified during data entry. This can improve data quality <i>and</i> decrease time needed to clean data. ^{20,38,42,43,45,46,50,51,66,70-72,78,80}	A, R
• Decreases the overall number of queries. ^{38,41,51,70}	R
• Automates parts of the query process and decreases query turn-around time ⁷⁵	A
• Identifies discrepant data as data are entered, providing site personnel and monitors the opportunity for early and continuous learning and potentially decreasing the number of queries per patient as the study progresses. ⁴⁶ This effect was seen in one of two studies evaluated by Dimenas et al. ³⁷	A, R
• Identifies discrepant data as they are entered, which may prompt study coordinators and investigators to become more pro-active in reducing such error ⁶⁶	R
Centralization of information in a web-based system rather than on a site computer, as was the case with RDE, enables more comprehensive security and backup. ^{43,73}	E, A
Increasing information system and data access for clinical trial monitors	
• Facilitates focus on potential problems or identified risk by making information about the status of the trial and data at sites available to monitors prior to monitoring visits. More efficient monitoring reduces the number and length of monitor visits and associated follow-up. ^{38,50,51,65,66,70-72,75}	C, DS
• Facilitates remote performance of some monitoring tasks and decreases the number of monitoring days on-site. ^{46,65}	R, A
Workflow assistance for (1) monitor queries resulting from Source Document Verification (SDV) and (2) documentation of SDV in the EDC system may make SDV more efficient or enable capture of useful data not previously available, such as the error rate detected through SDV. ⁷⁵	A
Replacing batch processing with the continuous flow of data ¹⁷ while making information available to users around the globe facilitates control the study at all levels. ^{43,45,66,80}	
• Enables instantaneous decision support such as protocol prompts, reminders, and triggered alerts via real-time interactive capabilities. ^{29,38,43,46,51,70}	A, DS
• Enables immediate oversight and improved coordination of the study. ^{37,38,50,75,80} For example <ul style="list-style-type: none"> ◦ Real-time access to enrollment data and other trial progress indicators,^{38,42,43,51,70,71} ◦ Faster identification of process changes and training opportunities, ◦ Offering and tracking site incentives for recruitment⁴⁵ ◦ Faster and better data increase potential for faster and better decisions and ultimately the ability to more quickly discontinue futile studies and poorly performing compounds.^{38,45,74} 	C, DS
• Enables parallel rather than sequential processes such as waiting to submit data until the visits have been monitored; decreases cycle times ¹⁷	R, A

(Contd.)

Benefits During Study Conduct	Mechanism
<ul style="list-style-type: none"> · Decreases time lag between when data are collected at a study visit and when they are processed and used: <ul style="list-style-type: none"> ◦ Faster data cleaning and ultimately locking the database faster,^{46,51,70,75} and ◦ Shorter time to analysis, so that study results are available sooner.^{42,45,50,66,74} 	<p>A</p> <p>A, C, DS</p>
Overcomes many challenges of geographical dispersion in data collection and cleaning. ²⁹	C

A: Automation; C: Connectivity; DS: Decision Support; E: Making use of less expensive or pre-existing equipment; KG: Knowledge Generation from data such as from data mining; P: Physical space savings; R: Relocation of work tasks to more efficient, better, or less expensive skilled individuals

Barriers to and Unmet Potential of EDC

Barriers to use of web-based EDC that surfaced during the early phases of EDC adoption have largely been overcome. These included problems with hardware and internet provision, internet connectivity, as well as new roles and large learning curves for data management, monitoring, and site personnel. Other barriers to early EDC adoption that have now been largely overcome include management of system access and privileges, provision of technical support for site personnel, and privacy and confidentiality concerns due to transmission of patient data over the internet.^{37,41,49,50} The latter necessitated new technology and methods to secure systems open to or operating over the internet;^{3,73} vulnerability to local bandwidth and network traffic being outside the control of the study team; increased dependence on informatics and IT expertise and institutional processes, for approval to install software that abided by local-client requirements or allowed institutional data to be housed in Sponsor’s web-based systems.^{46,73} Early adopters identified inconsistency between institutional policies at sites for retention of paper copies and EDC-based electronic research documentation retention guidelines.⁴⁶ Early lack of enthusiasm of site personnel for performing data entry, perceived as “clerical”,⁹ was largely overcome as site personnel perceived reduction in query-related effort and other benefits related to EDC implementation.

Most early barriers to EDC have been reduced or overcome through improvements in technology, process, and implementation or through evolution of trial personnel perceptions and acceptance over time. For example, consensus has been reached on interpretation and implementation of Title 21 CFR Part 11.⁴⁹ Today the regulation is understood as a requirement and organizations have procedures in place to ensure compliance. Though largely overcome today, the cumulative impact of these challenges has hampered, obstructed, and slowed implementation and adoption of EDC.⁶⁵

Today, despite all the progress made, EDC has still not reached full potential. For example, EDC was supposed to eliminate recording data on paper study forms at clinical sites. Early reports documented that using a paper form as an intermediate step between the source and the EDC system “doubled the data entry workload at the sites and also increased the monitor’s workload”.³⁷ However, use of paper forms at sites in this way has perpetuated at some sites^{59,67,78,81,82} indicating that further process improvements or better integration into site-specific workflow and data flow are needed. Early EDC

adoption was hindered by transiency of internet resources (e.g., controlled terminologies), web based knowledge sources, and interfaces used for web-based information exchange.⁷³ Transiency was a particularly difficult problem given increased desire to receive and use data from one system to another, which subsequently intensified need for system interoperability, technology reliability, and standards for data exchange. Though in some cases the transiency has been overcome, significant interoperability challenges remain.

Early Site-user Evaluations of EDC

Early EDC adopters were concerned about site acceptance of new technology and process changes, in particular the aforementioned relocation of data entry to clinical sites.⁹ Three site investigator surveys were reported following EDC pilots. Dimenas (2001) reported that 77% of investigators and 74% of monitors “found the workload to be reasonable”.³⁷ Similarly, 71% of 107 participants at the final investigators’ meeting indicated that the web-based EDC system offered definite advantage over other alternatives.³⁷ Litchfield et al. (2005) reported equivocal results with 57% of investigators indicating that setting up the EDC study sites took either a little more or much more time than past, non-EDC paper studies.⁴⁷ Half of the respondents reported that monitoring visits took more time, with only 28% reporting monitoring visits taking less time than in paper studies.⁴⁷ Similarly, Litchfield et al. go on to report that 42% perceived CRF completion in the EDC study as easier than their experience on past paper studies, while 28% perceived it as more difficult.⁴⁷ Half of the responding site investigators in this survey reported the perception that the number of queries was slightly fewer and their handling easier with EDC, while 7% perceived the number of queries to be much fewer.⁴⁷ Thirty-six percent of the responding site investigators in the Litchfield survey perceived the handling of queries to be more difficult with EDC than their experience on past paper studies.⁴⁷ Investigators perceived a worsening in increased time (1) to set up the study (58%) and (2) to complete the CRF (50%).⁴⁷ However, in the same survey, 36% of the internet sites thought use of EDC reduced the time and costs associated with the trial overall.⁴⁷ Eight of the 14 centers (57%) perceived that use of EDC for clinical trial data recording was better or much better than conventional systems, though 28% still thought it was worse or much worse.⁴⁷ A majority (71%) of the responding site investigators indicated they would prefer to use EDC over paper CRFs for future studies.⁴⁷ In a separate, brief site investigator survey conducted in

2005 following a pilot EDC study, respondents reported data quality (67%), data entry (78%), and workload (59%) to be an improvement over paper.⁶⁶ Overall, site investigators in the reported survey studies responded favorably to EDC.

Quantitative Evaluations of EDC

Today EDC is largely accepted as having cost, time, and quality advantages over data collection on paper.^{33,38,46,48,57,58,65,66,72,73,76,78,83,84} While early direct comparisons between EDC and paper were quite favorable toward EDC in terms of reduction of query volume and data collection time, only a few were published. Mitchel (2006) lamented that while it was clear that there were theoretical advantages of internet-based trials over traditional paper-based clinical trials, there was a paucity of evidence and that data were needed regarding the efficiencies of data entry, trial monitoring, and data review.⁷² Twelve studies reporting quantitative evaluation of EDC were identified in the literature search.

Of the twelve identified EDC evaluations (**Table 3**), only Litchfield et al. (2005) employed a randomized, controlled design.⁴⁷ Ten of the remaining eleven studies were observational. Six of the observational studies employed a comparator such as data double or single entered centrally from paper forms on the same study or on a comparable but different study.^{48,58,70,85–87} Four other of the observational EDC evaluations reported operational metrics including cycle times, number of data changes, and discrepancy rates from source-to-EDC audits from actual studies but with no comparator.^{37,59,72,81} The one

remaining report made comparisons between EDC and paper data collection, but did not describe the methods in sufficient detail to classify the research design.⁴⁵ Surprisingly, given the desire to retain the best clinical sites and sensitivity to site burden in clinical research,⁸⁸ no formal workflow or usability studies of EDC technology were found among published EDC evaluations.

Due to the small number of evaluations documented in the published literature, we do not know whether the results from the evaluations are representative of most EDC evaluations. It is likely, based on the identified studies, that most were observational in design and used existing organizational, operational metrics similar to those listed in **Table 4** as the basis of their appraisal. Cost, quality, and time metrics were variously used in the evaluations (**Table 4**). Only one evaluation reported metrics in all three categories. The cost, quality, and time outcome measures are summarized to inform design of future evaluation of data collection and management technology. Since the observed operational metrics varied between studies and were not well specified, further synthesis is dubious.

Data Accuracy in EDC Evaluations

The fundamental belief that EDC improves data accuracy is widely held and reflected both in the literature and practice. Yet, in preparing this review, only four comparisons of clinical trial data accuracy between EDC and traditional paper-based data collection in clinical studies were found.^{59,70,81,86} All four identified studies compared data initially recorded on paper forms to the same data subsequently entered in the EDC system. Thus,

Table 3: Quantitative Studies Evaluating EDC.

Chronological List of Identified Studies Quantitatively Evaluating EDC	Randomized Controlled Evaluation	Observational With A Comparator	Observational Without A Comparator
Banik and Mochow, 8th Annual European Workshop on Clinical Data Management, 1998 ⁵⁸		X	
Green, Innovations in Clinical Trials, 2003 ⁴⁸		X	
Mitchel et al., Applied Clinical Trials, 2001 ⁸⁵		X	
Dimenas et al., Drug Information Journal, 2001 ³⁷			X
Spink, IBM Technical Report, 2002*, ⁴⁵			
Mitchel et al., Applied Clinical Trials, 2003 ⁷⁰		X	
Litchfield et al., Clinical Trials, 2005 ⁴⁷	X		
Meadows, Univ. of Maryland at Baltimore Dissertation, 2006 ⁸⁶		X	
Mitchel et al., Applied Clinical Trials, 2006 ⁷²			X
Nahm et al., PLoS One, 2008 ⁵⁹			X
Mitchel et al., Drug Information Journal, 2011 ⁸¹			X
Pawellek et al., European Journal of Clinical Nutrition, 2012 ⁸⁷		X	

Randomized Controlled Evaluation: a research design where the experimental units (in this case data) are randomly assigned to an intervention (in this case EDC) versus some other data processing method as a control. Observational With A Comparator: a research design where there is no prospectively assigned intervention and no control, but data were collected for EDC and some other data processing method to indicate if one or the other is associated with a better outcome. Observational Without A Comparator: a research design where operational metrics were collected and reported for EDC use but where the same or similar metrics were not also observed for a different data processing method; studies using this design are commonly referred to as descriptive studies.

*There is insufficient detail in the Spink (2002) IBM Technical Report to classify the research design.

Table 4: Metrics Used to Evaluate EDC.

Metrics	DQ	Time	Cost
Query rate (number of queries per subject, form, page, or variable) ^{37,45,47,48,86,87,89}	X		X
Query rate at the time of data entry (number of queries per subject, form, page, or variable) ⁸⁵	X		X
Number of queries generated by the monitoring group ⁸⁵	X		X
Percentage of queries according to subgroups: missing, out-of-range, inconsistent, or invalid data ^{45,89} reason for change, e.g., data entry error, additional information, other ⁸¹ univariate and multivariate discrepancies ⁸⁶ detectable versus not detectable by additional, after entry, rule-based checks ⁸⁷	X		
Percentage of discrepancies resolved ⁸⁶	X		
Percentage of queries requiring clarification ⁴⁵	X		X
Number of (subjects, forms, pages, data elements, data values) requiring a modification ⁷²	X		X
Percentage of data (subjects, forms, pages, data elements, data values) requiring a modification ⁴⁵	X		X
Change in measures of data element central tendency, before and after cleaning ⁸¹	X		
Change in measures of data element dispersion, before and after cleaning ⁸¹	X		
Error rate (number of values in error/number of values assessed) ^{59,70}	X		
Percentage of invalid enrolled subjects ⁴⁵	X		X
Number of days between patient visit and data entry ^{37,47}		X	
Number of days from data entry to query resolution ³⁷		X	
Number of days from query generation to query answered ³⁷		X	
Number of days from query answered to query resolved ³⁷		X	
Number of days from query generation to resolution ^{37,47}		X	
Number of days between data entry and final modification ⁷²		X	
Number of days between data entry and queries resolved ³⁷		X	
Number of days between entry and clean data ⁷²		X	
Number of days between entry and form review by the CRA in the field ⁷²		X	
Number of days between data entry and form review by the in-house data reviewers ⁷²		X	
Number of days between data entry and form review by CDM ⁷²		X	
Number of days between Patient's last visit to Patient locked ³⁷		X	
Number of days between the First Patient First Visit (FPFV) to the date of the last data change ⁴⁷		X	
Trial duration ⁴⁸		X	
Number of days between Last Patient Last Visit (LPLV) and clean file ^{37,47,48}		X	
Cost of raising and resolving a query ⁴⁵			X

³⁷ Dimenas et al. Drug Information Journal, 2001. ⁴⁵ Spink IBM Technical Report, 2002, ⁴⁷ Litchfield et al. Clinical Trials, 2005. ⁴⁸ Green, Innovations in Clinical Trials, 2003. ⁵⁹ Nahm et al., PLoS One, 2008. ⁷² Mitchel et al., Applied Clinical Trials, 2006. ⁸¹ Mitchel et al., Drug Information Journal, 2011. ⁸⁵ Mitchel et al., Applied Clinical Trials, 2001. ⁸⁶ Meadows, University of Maryland at Baltimore Dissertation, 2006. ⁸⁷ Pawellek et al. European Journal of Clinical Nutrition, 2012. ⁸⁹ Takasaki et al., Journal of Nutritional Science and Vitaminology, 2018.

all four evaluation studies measured transcription fidelity from paper forms to the EDC system.

These studies all suffer from methodological weaknesses that limit their usefulness in assessing the accuracy of data entered at clinical sites via EDC. Only two of the studies^{59,70} reported a discrepancy or error rate. In a third evaluation study, Staziaki et al. (2016) employed a crossover experimental design to compare EDC with data collection via key-entry into a spreadsheet.⁶¹ However, the sample size was too small to assess data accuracy.

Thus, the Staziaki does not appear in **Table 4**. In a fourth evaluation study, Takasaki et al. (2018) reported data quality assessment for a rehabilitation nutrition study.⁸⁹ They enumerated missing, out of range, and inconsistent data as detected by rule-based data quality assessment, i.e., query rules. Four errors were detected in the data entered for 797 patients with data type errors, out of range errors, input errors, and inconsistency all assessed. The objectives of the Staziaki et al. and Takasaki et al. studies were not EDC evaluation.

The paucity of evaluative results regarding data quality is striking. Multiple authors (**Table 4**) have reported counts and rates of data discrepancies identified through rule-based data quality assessment, also called edit checks or query rules. Data discrepancy reports such as these are often and unfortunately confused with data accuracy. Discrepancies identified by rule-based methods may serve as an indicator of data accuracy since it is reasonable to infer that data discrepancies (detected by rules) are a by-product of actual errors in the data – and vice versa. However, (1) rules, real-time or otherwise, miss conformant but inaccurate data and (2) data discrepancies identified through rules are dependent on the number of rules, the logic of the rules, and the data elements to which the rules are applied. These aspects vary from study to study and, as such, so do the number of identified discrepancies and the percentage of fields, forms, patients, etc. with data discrepancies. In other words, the number of discrepancies for two data sets of equal accuracy may vary based on the rules used. Thus, neither the number nor rate of rule-identified data discrepancies provide a reliable estimate or basis for comparison of data accuracy across studies. Rule-identified data discrepancies, while easy to measure, are only a broad indicator- but not a measure of data accuracy.

Accuracy of a data value can only be determined by comparison with the true value, which in most cases is not known. Redundancy-based methods of identifying data discrepancies, i.e., comparing the data to an independent source of the same information, are more comprehensive than rule-based methods in that they detect all divergences from the comparison values. For example, in-range but wrong data will be detected by redundancy-based methods to the extent that the redundant data are themselves accurate. As such, where good sources of comparison can be found, redundancy-based methods as proposed by Helms (2001) are better indicators of data accuracy than rule-based methods.⁹ Sending two samples from the same time-point, patient, and instance of sample collection (a split sample) to two independent labs is an example of a redundancy-based method; as is comparing data collected for a study back to the original recording of the information, i.e., the source, as is done in clinical trial source data verification processes.⁹⁰ To the extent that redundancy-based methods cover all data elements of interest and employ an independent and more “truthy” comparator, they get closer to actually assessing accuracy than rule-based methods. Comparison to the source is still not a complete measure of accuracy because errors in the source are usually not known and not measured in the comparison. In the reports of EDC data quality assessment, only one, Nahm (2008), compared EDC data to the source; however, the source was a structured form completed during or after the patient visit and may have still contained errors itself.⁵⁹

Claims that EDC increases data accuracy (**Table 1**) are likely based in first principles, mainly that (1) the “forced rigor” of structured fields, limited response options, and real-time checks for or prohibition of missing or inconsistent data during entry decrease errors and (2) entering data closer in space and time to the original

capture allows correction where there is recollection or an original recording. While the forced rigor has been associated with lower error rates,⁹⁰ these likely true claims have not been proven in a generalizable way. Further, error rates measured from data entry were an order of magnitude lower than those measured between collected data and medical record source documents from which they were abstracted. Unfortunately, the only appropriate conclusion from this review regarding evaluation of data quality in clinical studies using EDC is that little is reported in the literature about the accuracy of data captured through EDC.

Discussion

EDC Adoption

In the last decade, the therapeutic development industry has made extensive use of EDC. In 2008, the CenterWatch EDC adoption survey of investigative sites reported that 99% of sites were using EDC in at least one of their trials, with 73% of trial sites utilizing EDC for at least one-fourth of their studies.⁵² In the same survey, 36% of sites reported using some type of EDC including interactive voice randomization, web-based data entry, fax-based data entry, or electronic patient diaries for at least one-half of trials they conducted.⁵² Only 2% of responding sites, however, indicated they no longer collected any case report form data on paper.⁵² A decade later in the 2018 eClinical Landscape Survey, 77.5% of eligible respondents reported managing CRF data in the primary EDC system.⁴ Today with most traditional CRFs entered and cleaned over the internet, the primary mechanism for collection and management of CRF data is web-based EDC.

The time period from realization that clinical study data could be collected electronically at sites and immediately available to the study team until full adoption seems long. The time period from the earliest report of RDE (in 1973), until the most recent adoption reports, spans four decades. However, as seen in **Figure 1**, the EDC adoption curve (bold blue line) is comparable with most contemporary new technology adoption rates.⁹¹ The entry tail of the adoption curve may have been more protracted for EDC than those for most other reported technologies. However, early data before the inflection point are not provided for most other innovations on the graph and such a comparison is not possible. Given complexities involving the regulated international nature of therapeutic development, the primacy of human subject protection concerns (including protection of individual health information), the challenges of crossing organizational boundaries for EDC implementation, and the sociotechnical aspects of EDC changes wrought in workflow and information flow (for sites and sponsors alike), it would not be surprising if the approach to EDC adoption were slower than for other innovations.

Lingering Challenges With EDC

Fully adopted does not necessarily mean fully matured. Today, web-based EDC is a valuable member of an ecosystem containing multiple data sources and information systems supporting clinical trial operations. The functionality in web-based EDC systems has become

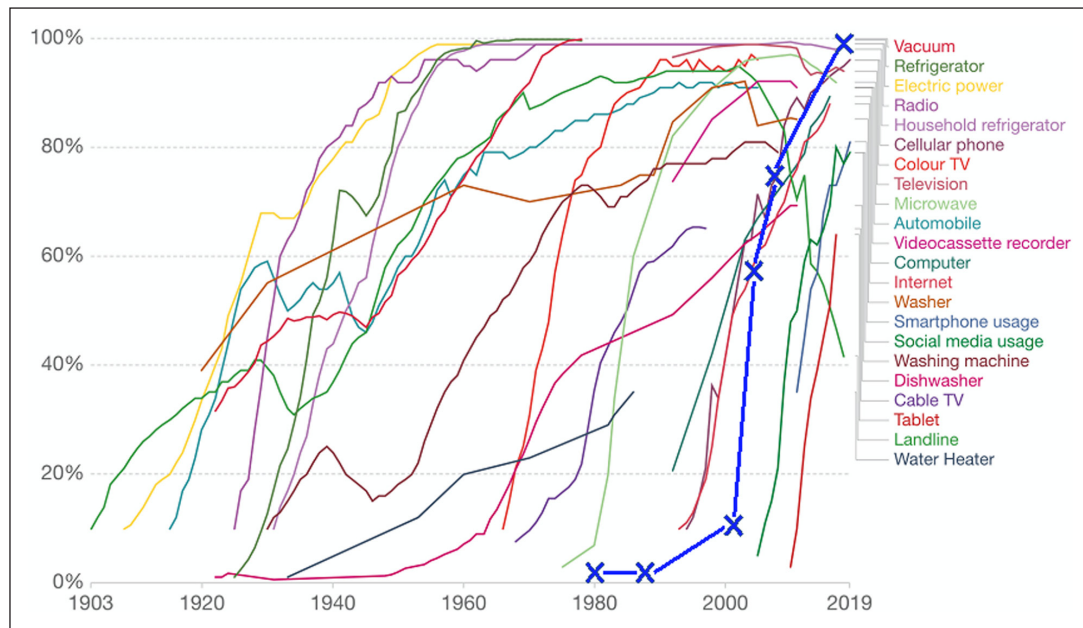


Figure 1: Estimated EDC Adoption Curve Superimposed on US Technology Adoption Data.

Source: US Technology adoption in US Households data and image from Ritchie and Roser.⁹² The data sources from which the image was compiled are listed at <https://ourworldindata.org/technology-adoption#licence>. The image was reproduced and adapted here under the creative commons license. EDC adoption data were obtained from the studies referenced here^{4,52,53} and superimposed in bold blue “X’s” on the image. Point estimates were made by the authors averaging percent adoption of sites and trials where both existed at the same timepoint. Thus, the actual EDC adoption curve may lie several years or percentage points in either direction. Some regions of the world likely experienced significantly shifted curves from others.

fairly stable. However, as others^{67,76,78} have noted there are multiple opportunities for software vendors and organizations to move past the current state and realize significant EDC powered improvements in clinical research data collection and management.

Limitations of EDC technology itself (1) and the extent to which we exploit it (2 and 3) remain today and include:

1. Lack of interoperability
 - a. Lack of interoperability between EDC systems and the ever-increasing myriad of non-EDC data sources in clinical studies such as medical devices, ePRO systems, and site EHRs.⁶⁰ These include moving beyond site-visit-based studies to support direct to consumer studies and pragmatic studies, in which some or all baseline and outcome data may come from EHRs, devices, or direct patient input.
 - b. Lack of interoperability between EDC systems and institutional infrastructure systems such as research pharmacy systems, electronic Institutional Review Board (eIRB) systems, and sponsor and site-based Clinical Trial Management Systems (CTMSs).⁶⁰
2. Lack of processes re-design to leverage EDC technology more fully⁸⁰ such as routinely requiring daily data entry, review, and feedback to decrease cycle times and error impact,^{17,72,93} providing automated alerts and decision support, and employing surveillance analytics that continuously monitor data to detect process anomalies.
3. Lack of formal error rate estimation for data collected through EDC systems.

These areas are large reservoirs of untapped potential, and at the same time remain an impediment to advancing the conduct of clinical studies.

Limitations 1 and 2

The ability to accomplish the re-engineering needed to advance the conduct of clinical studies is partially dependent on, and greatly augmented by, interoperability with other systems. Some of the earliest articles about EDC conceptualize EDC systems as a hub, and describe functionally as integrating operational and clinical study data from multiple sources.^{16,21,38,76,84,94} However, this comprehensive integration of infrastructure and study data sources has not come to fruition. Lack of ability to integrate data from disparate sources in EDC systems (limitation 1) was reported early in the EDC adoption curve.^{43–45,74,95,96} Multiple recent reports provide examples of custom EDC integration with other systems for individual studies after not finding available solutions.^{55,60,67,71,97,98} Today, integration of data from other clinical or operational systems with EDC remains largely a custom and point-to-point endeavor. Reports of integration and interoperability remain a challenge in most organizations, with 77% of companies represented in the 2018 eClinical Landscape survey reporting challenges loading data into primary EDC systems.⁴

Interoperability-related EDC limitations are significantly exacerbated by the increasing volume and variety of new and external data sources in clinical studies. The 2018 eClinical Landscape Survey indicated that in trials the following non-CRF data are frequently collected and managed: central lab data, local lab data, quality of life data, ePRO data, pharmacokinetic data, biomarker data, pharmacodynamic data, electronic Clinical

Outcome Assessment (eCOA) data, medical images, genomic data, and mobile health data.⁴ The 2015 and 2018 Clinical Data Management job analysis survey corroborated the increasing use of these types of non-CRF data within clinical trials.^{99,100} Respondents to the 2018 eClinical Landscape Survey reported an average of 4.2 to 6.5 computer systems used in clinical trials, with more than two-thirds of respondents reporting they will increase the number of data sources over the next three years.⁴ Others, e.g., Lu (2010),⁶⁰ Howells (2006),⁶⁸ Brown (2004),⁸⁴ and Comulada (2018),⁹⁸ report similarly high numbers of computer systems used in studies and Handelsman (2009)¹⁰¹ the need for integration. At the same time, additional alternate real-world data sources are aggressively being pursued.^{102–104}

These non-CRF data were reported as managed in the primary EDC system by fewer than five percent of the eClinical Landscape Survey respondents.⁴ The increase in non-CRF data^{99,100} coupled with management of non-CRF data in systems other than EDC systems, indicates that EDC systems may play an increasingly smaller but still essential role in data collection and management. A shrinking role would limit EDC systems' ability to advance study conduct, management, and oversight. Today, there are few commercial solutions for integrating all study data during the study. Lack of interoperability to accommodate these external data sources^{21,76} leaves organizations to integrate data by building their own clinical study data integration hubs, use general commercial tools built for other industries, or do without needed data integration during studies (and integrate data only for statistical analysis post data collection).

The cost of not having non-CRF data integration with EDC is high. Without it, organizations largely lack comprehensive information in near real-time needed to better manage and conduct studies. Mitchel et al. (2006) and Summa (2004) illustrate the importance of near real-time data collection and review to identify problems before they affect multiple subjects and to help finalize data collection and processing sooner.^{17,93} Wilkinson et al. (2019) report increases over the past decade in cycle times to build the study database, input post-visit data, and lock study databases as well as more variability in data handling cycle-times.⁴ The increased number of non-CRF data sources used in studies, in the absence of supporting data integration, is a possible contributor for the lengthening cycle times.¹⁰⁵ Similarly, the burden on clinical investigational sites associated with data collection remains a significant concern.¹⁰⁶ Distribution of mobile devices, training patients in their use, and manually tracking their use in studies increase site burden. The absence of supporting integration also results in boluses of after-the-fact questions when data are later integrated and reconciled. The trend of getting studies into production later, providing data later, and encountering barriers to timely reconciliation all affect trial execution. Based on the most recent survey,⁴ lack of data integration with EDC systems remains a significant obstacle to optimal study conduct and management.

As the adage goes, "you can't manage what you can't measure," and without integrated data, we can't measure,

much less use the data to re-engineer processes. As time and cost demands on clinical studies intensify, solving the non-CRF data integration problem and gaining the ability to leverage data in near-real-time to manage studies should be one of, if not the prime target for study risk and cost reduction.

Fortunately, focused efforts toward creating standards to support interoperability within clinical research have recently intensified. Industry, federal, patient, professional, and academic organizations and the Clinical Data Interchange Standards Consortium (CDISC) have joined with Health Level Seven (HL7) to create an organization within HL7 called VULCAN (www.hl7.org/vulcan). Vulcan's chief mandate is to accelerate development and adoption of the Fast Healthcare Interoperability Resources (FHIR®) data standards in clinical research. The FHIR® standards are now being widely adopted in healthcare in many countries, and provide a mechanism to extract data from site EHR systems. The HL7 FHIR® standards could be further developed toward use cases for broader information exchange between research and healthcare, between organizations working together to conduct studies, and for better information exchange within research sites. Thus, the HL7 FHIR® standards may play a key role in advancing beyond our current state.

Using the EDC system as the original recording of study data, i.e., the source, or to receive an electronic copy of the source are two special cases of interoperability likely to grow in the near future. Two early authors, Mitchel (2010, 2013, 2014) and Vogelson (2002), promoted use of EDC as eSource. Mitchel (2010, 2013, 2014) developed, implemented, and evaluated EDC as eSource.^{78,93,107} An early survey revealed that 22% of respondents were entering data directly into the EDC system.⁴⁵ The survey likely overstated what would be accepted today as eSource use of an EDC system since it pre-dated public guidance¹⁰⁸ by four years and regulatory guidance² by a decade. Where data are not documented in routine care, such as in the case of commercial sites that do not provide care outside the context of clinical studies or studies that require data not collected or not documented in routine care, EDC eSource will likely have a secure niche. However, sites leveraging mainstream EHR systems to provide routine care to large patient populations may be better supported by extracting EHR data using FHIR® standards to prepopulate the eCRF. The literature emphasizes, however, that not all data are available through EHRs and that data availability will vary across sites^{109,110} meaning that those capable of EHR-to-eCRF interoperability will still require the ability to enter some data, source or otherwise, into the EDC system. For these reasons, sponsors and sites will likely be best served by EDC systems that support both electronic acquisition of available EHR eSource data as well as entry of the original data into EDC systems.

EDC Limitation 3

Finding only one instance of comprehensive, i.e., source-to-EDC, data accuracy assessment in the literature suggests that data accuracy was not a significant concern in initial EDC evaluations. Lack of routine measurement and

reporting data accuracy in practice for data collected via EDC⁵⁹ is a significant but correctable oversight. The error rate for source document verified data could be calculated with functionality and metadata in many EDC systems today. In the absence of a measured data entry error rate, and without leveraging opportunities to calculate an error rate from source document verification, EDC has decreased the knowledge about data accuracy available to trialists and regulators. Reasons behind not using EDC systems to assess data accuracy likely include (1) Methods for data accuracy measurement are poorly understood by those outside data management and statistics; (2) Data accuracy measurement from SDV processes aided by EDC systems would require documenting the fields for which SDV is performed and each data error identified; and (3) Monitoring and SDV procedures are usually outside the control of data management and statistical team members. Merely using available functionality to mark fields on which SDV was performed, using EDC system metadata to estimate a source-to-EDC error rate from SDV, and reporting the error rate would be a significant improvement in the rigor of data collected via EDC.

“There is a major difference between a process that is presumed through inaction to be error-free and one that monitors mistakes. The so-called error-free process will often fail to note mistakes when they occur”¹¹¹ Knowing the accuracy of data *during* a study makes intervention and prevention of future errors possible. Knowing the accuracy of data *from* a study is necessary to demonstrate that data are capable of supporting the study conclusions. Without such comparisons we do not know if EDC data are capable of supporting study conclusions. Lack of data accuracy assessment remains a major shortfall in our use of EDC technology today.

Lessons Learned from EDC Adoption

We have learned from the history of EDC adoption that non-technical factors significantly impeded adoption in the therapeutic development industry.^{48,77,80} For example, companies demurred from significant role re-definition and therefore missed significant re-engineering opportunity.^{80,101} In clinical informatics, it is commonly held that successful adoption of new information systems is “about 80% sociology, 10% medicine, and 10% technology”.¹¹² Thus, to the extent that individuals, groups, and organizations will interact with new technology, needs assessment, design, development, dissemination, and implementation should equally account for human and sociotechnical factors and potential barriers.

The EDC adoption story demonstrates that “beneficial effects are obtained when the ways of thinking and working are changed to take advantage of the opportunities arising from having information online”.³⁷ Web-based EDC offers the ability to centralize information while decentralizing its use, making information and information products (such as decision support and automation) available to everyone on the study team simultaneously and in real-time. This enables study teams to do things not possible, or at least unwieldy, prior to web-based EDC technology. As evidenced by the aforementioned limitations 1 and 2,

implementing new technology without transformative process change to leverage these opportunities has not and will not yield the expected improvement.^{48,60,67,77,80,94,113}

In addition to the process re-engineering needed to realize the benefits of new technology, infrastructure that comprise Quality Management Systems such as technical, managerial, and procedural controls need to be adjusted to the new technology. Roles and responsibilities have to be adjusted to the new working processes.^{60,67,77,80,94} Individuals in affected roles need training and time to adjust and gain experience with the new technology.^{60,67,77,94} The implemented processes and software need to be monitored in order to ensure expected performance in local contexts.^{74,75,77,80} This capacity building and infrastructure development is a project unto itself and should be managed as such, separately from evaluation pilots and stabilized before using new technology in routine operations.⁸⁰ Even then, new technology faces Solow’s Paradox; i.e., that IT investments aren’t often or immediately evident as increased productivity.¹¹⁴ Hypothesized reasons why immediate benefit is not perceived with new technology include lack of accounting for work redistribution, lack of accounting for work to meet needs for increased explicitness needed for automation and decision support, and failure to leverage and integrate the new technology with pre-existing technology, infrastructure, and processes. Looking back on EDC adoption, we can see their mark. The difficulty prospectively conceptualizing and valuing opportunities made possible by new technology adds another reason why increased productivity and return on investment (ROI) often isn’t immediately evident. Such ROI projections should account for increasing value of information technology to an organization as more data, especially data traditionally managed in separate systems, become available for use.^{65,68} Data sharing from early pilots of new technology may help overcome the paradox through more accurate prediction as will methods that take into account likely causes of the IT productivity paradox.

In major re-engineering endeavors, large companies with existing and stable infrastructure (including roles, responsibilities, procedures, and technology in place) tend to have more inertia and are slower to change.¹⁷ On the other hand, companies without legacy infrastructure are usually more nimble in implementing beneficial change.¹⁷ Thus, large organizations need clear strategy, strong leadership, meticulous goal alignment, and thorough understanding of how new technology will mesh with pre-existing infrastructure in order to re-engineer at a pace similar to their smaller and newer competitors.²⁵

In the authors’ experience many organizations piloted one or more EDC systems. As evidenced by this literature review, very few of these pilot projects published results, and even fewer published in peer-reviewed scientific literature. Practices regarding information sharing in therapeutic development have evolved over the last three decades and pre-competitive information sharing now occurs much more frequently than it did in the past. Examples include initiatives such as TransCelerate

Biopharma (<https://transceleratebiopharmainc.com/>), the Clinical Trials Transformation Initiative (CTTI, <https://www.ctti-clinicaltrials.org/>), the Society for Clinical Data Management (SCDM) eSource Consortium (<https://scdm.org/esource-implementation-consortium/>), projects undertaken by members of European Federation of Pharmaceutical Industry Associations (EFPIA), and the VULCAN Accelerator in HL7 (<http://www.hl7.org/vulcan>). Through initiatives such as these, future technology evaluations are expected to be more collaboratively undertaken and published.

The history of EDC as told through the literature is that, with one exception,⁴⁷ the EDC evaluations that were published leveraged observational methods and empirical data without the support of experimental controls. In many cases the research was done without contemporaneous comparators. This review found no record of ongoing productivity monitoring past initial evaluations. Further, variability in the outcome measures used in the reported studies precluded formal meta-analysis to synthesize evaluation results (**Table 4**). A general lack of sharing evaluation results and the heterogeneity of evaluation outcomes, along with the aforementioned barriers all likely contributed to the “perpetual piloting” phenomenon mentioned in the literature, and lengthened the EDC’s adoption curve. More rigorous evaluation on consistent outcome measures, along with earlier information sharing will likely benefit the industry in future new technology adoptions. Collaborative evaluation could further decrease uncertainty earlier in and shorten the adoption curve.

There have been three major paradigm shifts in clinical research data management: (1) the use of structured forms for data collection; (2) the advent of clinical data management systems in the late 1990s, in which data

were entered, imported, integrated, stored, cleaned, coded, and otherwise processed; and (3) web-based EDC that decentralized data entry, cleaning, and use. While we argue that, though the potential of the latter is not yet fully realized, all three innovations have positively increased the ability to plan, conduct, and manage clinical studies, increased the types and number of studies that we can conduct, and improved the documentation, if not the quality of the study data. Three major limitations remain with EDC technology: (1) interoperability deficits, (2) unexploited process re-engineering potential, and (3) lack of error rate estimation. Significant untapped potential exists for the use of automation and immediate information availability to support, make, and act on study management decisions in real-time such as following-up on data and resolving discrepancies within a day of their commission and detecting and intervening in operational anomalies like protocol violations and non-compliance immediately. The potential gains are magnified when all sources of data on a clinical study are centrally available through interoperability to signal detection algorithms and for decision support.

Due to these limitations, in terms of diffusion of innovation,^{115,116} our current plateau on the innovation “S-curve” falls short of what could be achieved with current EDC technology (**Figure 2**).

Advances in processing and use of different types of data (beyond those obtained from CRFs) when integrated into EDC have the potential to further, and sustain, an upward performance trajectory in EDC effectiveness, adoption, and function. There are many technological advances on the horizon including new data sources, advanced ways to extract information from data such as image processing and natural language processing, more advanced ways

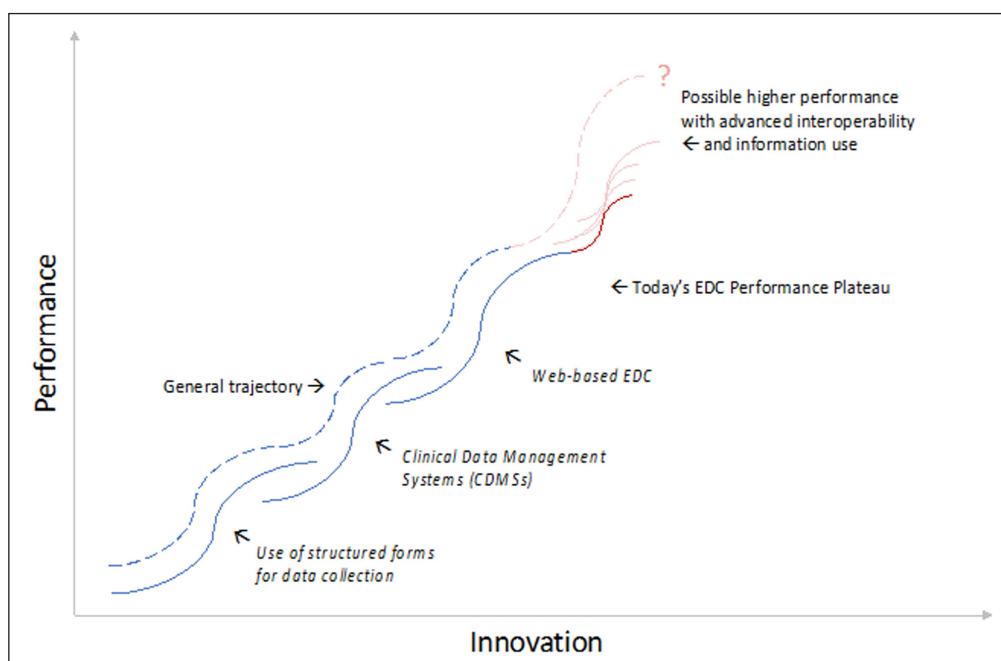


Figure 2: Major Events on the Trajectory of Innovation and Performance Improvement in Data Collection, Management, and Use in Clinical Research.

Diagram adapted from: Clayton M. Christensen, “Exploring the limits of the Technology S-Curve. Part 1: Component Technologies”. *Production and Operations Management* 1, no. 4, (Fall 1992) 340.

to generate knowledge from information (such as data mining and machine learning), and advanced ways to apply new knowledge to augment human performance such as use of artificial intelligence for signal detection and decision support. Likely similar advances will positively impact clinical study design, conduct, oversight, and reporting. Parallel advances in data standards such as those currently pursued through the HL7 VULCAN accelerator for clinical research could exponentiate performance gains through agreement on, and availability of more precise data definition and mechanisms for data exchange. If fully pursued to support clinical research use cases, FHIR® standards will unlock data not previously available (or not previously computationally accessible) and will enable new uses of operational data such as computationally aligning EHR data to a study schedule of events, or enabling new opportunities to shorten therapeutic development such as seamless conversion of open-label extension studies to EHR- or claims-based post-market registries.

Many organizations will pilot and eventually adopt new data sources and novel ways of collecting, processing, and using data in clinical studies. To gain value, organizations will need to integrate them into existing processes or re-engineer existing processes to optimally exploit them. For example, use of artificial intelligence to detect operational anomalies is easy to implement separate from existing data collection and processing pipelines. However, with a recent and notable exception, none as of yet have integrated such data into study data processes or at-scale, routine use by clinical trial teams. The EDC adoption story indicates that each advance may experience a protracted entry to the adoption curve for similar reasons experienced by EDC and will likely have the characteristic 15–20 years¹¹⁷ to reach maturity and widespread adoption as experienced by many other recent technological advances (**Figure 1**). Each of these advances has the potential to sustain performance increases in clinical research moving us beyond today's EDC (**Figure 2**). While only wild speculation, the next major paradigm change in data collection, management, and use in clinical studies may be comprehensive adaptation and adoption of the FHIR® standards in clinical research. Application of the FHIR® approach to enable seamless exchange of data within clinical site facilities, between clinical sites and study teams, and among organizations working together to conduct clinical studies would end the current siloed state.

Beyond EDC

What can we learn from the history of EDC adoption to help us move beyond today's EDC, to derive greater benefit from new technology, and to do so faster?

In the beginning, EDC processes and technology were new to site investigators, site staff, sponsors, and regulators. EDC adoption had a pervasive impact across processes and roles at sites, sponsors, and regulators. The broad impact involved new technology, new processes, new tasks, new skills, and involvement of new roles and affected new ways of working for individuals, groups, and organizations. Examples include the following:

1. EDC changed the data collection and submission workflow at clinical sites.
2. EDC shifted work to, and required new competencies of investigators and site staff, such as entering study data and reporting software problems.
3. EDC forced site staff to fundamentally change their thought processes used in data collection; for example, paper forms were often used as cognitive aids encoding CRF completion instructions and in some cases serving as worksheets and check lists supporting systematic and complete data collection. EDC challenged but did not completely overturn this practice.
4. EDC changed the data review and auditing process for sponsors and regulators.
5. EDC shifted new work to, and some existing work away from data management.
6. EDC added steps to site start-up, trial monitoring, and study management such as a) obtaining access to and training on EDC software, b) using EDC software to document SDV, and c) needing to have edit checks and workflow ready prior to the start of enrollment.
7. The automated alerts and dynamic form behavior available in today's EDC systems requires Data Managers to be skilled at workflow analysis and process design.
8. EDC requires the involvement of new roles (and people) at trial sites such as the addition of information technology support staff.
9. Using computer systems to provide automation such as generating additional pages and forms and communicating data discrepancies directly and immediately to sites required increased explicitness, detail, and precision in study specifications. For example, query wording had to be written so that it did not require manual editing or customization because sites saw the queries immediately. In general, increasing the level of automation increases the explicitness required to program computers to do operations previously handled by humans do operations previously handled by humans. In this way computer systems in increasing the explicitness required (and often not previously undertaken) are perceived as increasing rather than decreasing work, i.e., the IT Paradox.

It is not evident from the reviewed literature that the breadth and depth of these fundamental shifts were expected or clearly articulated at the onset of EDC. Similarly, it is also not evident from the reviewed literature that the new possibilities offered by the increased information content and availability offered by EDC were recognized by, clearly (i.e., mechanistically) articulated by, or exploited by early adopters. These are now better articulated in the Good Clinical Data Management EDC chapters. Additionally, the increased explicitness spurred by EDC has offered new possibilities, such as automated generation of forms only used under special conditions (e.g., an early withdraw form), and automated detection of protocol violations, alerts and other events. Other benefits

derived from increased protocol specificity, and EDC use in general, include gaining tracking information as a by-product of work tasks and the ability of geographically distributed teams to simultaneously access, review, and respond to data and alerts on data when entered. Though speculative, the lack of broad awareness of these new opportunities and the cost or added value of each, contributed to delay in EDC adoption.

Many emerging advances apply narrowly to one type of data or another. For example, natural language processing generates or extracts structured information from free text. Similarly, artificial intelligence operates over existing data and extends the possible uses of the data through automation or decision support. Though these certainly offer new possibilities for data use, they are focused on narrow use cases. In contrast; a new universe of opportunities may be opened by comprehensive and easily implementable data definition and exchange standards that bridge existing and previously computationally impermeable boundaries such as those between trial sites, healthcare facilities, sponsors, central labs, core labs, and central reading centers participating in study conduct, and those offering new sources of data such as healthcare claims data or data from medical devices. The data standards that would allow this, albeit slow in coming and with much investment remaining, offer new opportunities for gaining value from data use, like aqueducts through a nation of information deserts.

One area of information exchange already opened by the HL7 FHIR® standards is the direct extraction of data from Electronic Health Records (EHRs) and transmission of the data on an ongoing basis to a study EDC system. Though individuals and organizations have pursued direct use of EHR data in longitudinal studies for decades, early demonstrations were limited by use in single EDC systems, single EHRs, and single-site studies, use of older or no standards, and were largely conducted outside the context of an ongoing clinical trial.¹¹⁸ Different from most other study data sources, direct EHR-to-EDC data collection shares many of the fundamental shifts seen with EDC in that roles, processes, information flow, and needed skills are impacted for sites, sponsors, and regulators alike at the level of individuals, groups, and organizations. Like EDC, direct data collection from EHRs offers possibilities not available today such as decreasing data collection burden on sites, detecting and correcting data quality problems at the source, extending an unbroken chain of traceability back to the exact data value in the source, extending safety surveillance for years into the future, and assessing generalizability of study results. Before these and other opportunities not yet conceived can be pursued, we must climb the adoption curve. The cost and time pressures in therapeutic development today will likely not withstand another two- to four-decade wait. The similarities between EDC and direct EHR data collection are quite striking. Thus, the lessons learned from EDC implementation, adoption and scale-up offer knowledge and guidance toward a more direct and streamlined testing, evolution, adoption, implementation, and optimization of the EHR as an important data source in the development, testing, and monitoring of new therapeutics.

Limitations

Though extensive attempts were made to identify all seminal events and relevant evaluations in the history of EDC, few were published in peer-reviewed literature. Many pertinent and important articles and papers may have been missed because they were published in outlets not indexed or preserved for academic retrieval. This review completely misses the likely substantial work undertaken and communicated only within organizations. Additionally, one individual extracted information for this historical review. Although the articulation was independently reviewed by the two co-authors, the initial extraction is subject to human error and bias associated with first author's single reviewer synthesis. Commentary on the article is encouraged to counter any bias or missed information relevant to this review. In this vein, materials used in this review will be made available for re-interpretation and analysis.

Conclusions

In this comprehensive review of the EDC literature, the large number of articles identified through references rather than through indexed literature search is striking. This likely indicates that organizational leaders and practitioners pursued EDC pilots and adoption in absence of the extant knowledge at the time – or based on anecdote. This likely slowed EDC adoption. The synthesis of EDC benefits and barriers presented here may inform future evaluation of technology for use in clinical studies. Literature synthesis, as is currently being done in the Good Clinical Data Management Practices (GCDMP) and pre-competitive information sharing common today should benefit those employing new technology or methods in the design, conduct, and reporting of clinical studies.

Based on the study designs employed in the reviewed EDC evaluation articles, very few EDC evaluations were conducted with rigorous designs capable of supporting causal inference that EDC technology directly brought about positive change in quality, cost, or time metrics. Disparate metrics reported in the EDC literature further impede progress by precluding comparisons and quantitative synthesis. We do conclude, however, that the published evidence supports the finding that EDC facilitates faster acquisition of data with fewer discrepancies. The evidence also indicates that significant room exists for decreasing data collection cycle-times, and that this can be achieved via as-soon-as-first-possible data entry (or transfer) of data and review of that data. The evidence does not support claims that overall data accuracy is improved. The dearth of measurement and reporting of source-to-EDC data accuracy is quite surprising, especially when calculation of such can be directly supported by EDC technology today. This constitutes a significant oversight by organizations conducting clinical studies. The most important question with respect to use of any data, and especially that used in regulatory decision-making, is whether the data are of sufficient quality to support intended decisions. This has not been proven in a generalizable way for EDC. We surmise that this lapse is fueled by the faulty perception that errors in the source cannot be detected or corrected or by the mistaken belief

that translation of source data into EDC is seamless and highly accurate. Lack of EDC data quality assessment, in-particular accuracy measurement in clinical studies, should be immediately remediated.

The review identified multiple future directions for moving beyond today's EDC, including: (1) Providing easy, real time, and seamless acquisition and integration of clinical data; (2) Providing easy, real time, and seamless interoperability with other operational information systems used in clinical studies; (3) Measuring the accuracy of study data, (4) Supporting study conduct and management through automation and decision support, and (5) Re-thinking and moving existing boundaries in therapeutic development through real-time exchange of computationally accessible data. Potential examples of the latter include clinical verification of direct-to-consumer data and broader early use of therapeutics enabled by direct EHR and claims data acquisition and surveillance. In most other industries new or faster availability of information has opened up entirely new opportunity. We are only beginning to use information technology such as EDC and available data such as routine care and claims data to benefit development of new therapeutics and protect the health of the public that uses them.

Appendix: Description of Quantitative Evaluations of EDC

Evaluation 1

Banik and Mochow, presented in 1998.

During the 1998 8th Annual European Workshop on Clinical Data Management,⁵⁸ Banik and Mochow reported on their comparison of EDC to traditional paper data collection in a study conducted at Bayer Vital GmbH & Co. The EDC evaluation compared two similar studies from the same drug development program sponsored by Bayer Vital GmbH & Co. One study employed traditional paper data collection, and the other implemented web-based EDC.⁵⁸ The results were subsequently published by Green (2003).⁴⁸ Banik and Mochow measured a 30% reduction in trial duration, an 82% reduction in queries, an 86% reduction in query resolution time (attributed to "use of immediate edit checks that are not possible with paper"), a 43% reduction in time to database lock and a 9% increase in the number of evaluable patients.^{48,58} This was the earliest identified quantitative evaluation of EDC.

Evaluation 2

Green, published in 2003

Green reported results from a Gilead Sciences study using the same EDC system deployed by Banik and Mochow (1998). Green reported a 75% reduction in query rate and a 45% reduction in time to database lock.⁴⁸ However, details of the comparator, traditional paper data collection with central data entry, were not provided. It is not clear whether the metrics from the comparator used by Banik and Mochow (1998) were used, whether a separate and comparable study to the Gilead trial was used, or whether the data for the comparison were from data processed via traditional paper data collection on the same study in different sites, different patients, or in the same patients but in parallel to the EDC processes.

Evaluation 3

Mitchel et al., published in 2001

A third early evaluation comparing EDC versus traditional paper data collection was reported by Mitchel et al. (2001). In the evaluation, three similar studies using an observational cohort design were examined.⁸⁵ The first study (study 1) was performed with traditional paper CRFs and a rule-based data cleaning system. The second study (study 2) used a CRO-developed, Web-based data collection system with no rule-based edit or logic check functions. The third study (study 3) employed the same web-based data collection system with full rule-based edit and logic check functions. In this comparison, EDC with full rule-based edit and logic check functions achieved a 63% decrease in queries at the time of data entry and a 65.5% decrease in the queries generated by the monitoring group compared with the traditional paper data collection.⁸⁵

Evaluation 4

Dimenas et al., published in 2001

In the fourth EDC evaluation identified by this review, Dimenas et al. (2001) observed operational metrics from two EDC pilot studies.³⁷ In the two pilots, 69% and 54% of visits were entered the same or next day, and 23% and 24% of queries were resolved the same or the next day with the average time from query generation to resolution of 18 and 17 days for the two pilot studies.³⁷ Last Patient Last Visit (LPLV) to clean file on average was 14 and 20 days, respectively.³⁷

Evaluation 5

Spink, published as an industry white paper in 2002

A fifth report of EDC evaluation provided metrics from ten phase three studies conducted over a three-and-a-half-year period and involving 6,700 subjects. The report was made via an industry white paper by Spink (2002).⁴⁵ Neither therapeutic area nor the evaluation design were described. Spink reported a 50% decrease between EDC and traditional paper data collection in the percentage of invalid enrolled subjects, an 80% decrease in the cost of raising and resolving a query, a 95% decrease in the number of queries per subject, a 95% reduction in the percentage of data requiring correction, a 100% decrease (from 48% paper to 0% with EDC) in the percentage of queries caused by missing data, an 86% reduction in the percentage of queries caused by inconsistent data, an almost complete reduction in the percentage of queries caused by out-of-range data (from 8% with paper to 0.1% with EDC), a 100% reduction (from 6% with paper to 0% with EDC) in the percentage of queries requesting clarification, and a 50% decrease in the percentage of queries caused by invalid data.⁴⁵

Evaluation 6

Mitchel et al., published in 2003

In a sixth relevant evaluation identified by the review, Mitchel et al. (2003) reported an overall error rate of 0.41% in EDC generated data, after detecting 950 errors in 229,152 fields. This observational evaluation compared single entry from forms into an EDC system with double

data entry of the same data in a 124-subject trial conducted at 15 clinical sites.⁷⁰ Though the aforementioned evaluations reported substantial reductions in queries, Mitchel et al. (2003) was the first EDC evaluation reporting measures of data accuracy.

Evaluation 7

Litchfield et al., published in 2005

In the seventh and most rigorous of the aforementioned evaluations identified for this review, Litchfield et al. (2005) directly compared EDC and paper data collection in a cluster-randomized experiment in which investigational sites were randomized to EDC or paper data collection. They report that the time from the last patient completing the study to the release of the database was shorter in the EDC sites (33 rather than 48 days, a 31% decrease) in spite of the much larger number of patients in the internet group.⁴⁷ The study reported no appreciable difference in time from the first patient first visit (FPFV) to the date of the last data change between the two arms of the study.⁴⁷ However, site differences (site at which the data were entered) in lag time to database release was highly significant, indicating that the differences in times were due to site effects rather than group effects.⁴⁷ The majority of EDC data were entered within a few days after the visit, with 90% of the data entered within three weeks after a study visit. This was a stark contrast to the paper group where data entry took up to six months.⁴⁷ Though there were more queries in the EDC group (11.4 queries per patient with EDC versus 1.4 in the paper group) on average, the query volume was not perceived as significantly higher. The study authors attributed this to the fact that queries in the EDC group were posed at the time of entry when correction could be made immediately, whereas paper queries came after-the-fact.⁴⁷ In the same study, queries resulting from programmed validation checks were resolved on average 13.2 days faster while clinical, and presumably manual, queries took on average 3.7 days longer.⁴⁷ Similar data entry time site effects were also reported in an observational analysis of studies using Direct Data Entry, i.e., using the EDC system as the source, reported almost a decade later.^{93,107,113}

Evaluation 8

Meadows, published as a Doctoral Dissertation in 2006

The eighth evaluation identified by the review and reported by Meadows (2006) used secondary analysis of existing observational data to assess EDC. To conduct the evaluation, Meadows compared data entered at clinical sites, from paper CRFs, into an EDC system to data from sites submitting paper forms for data entry at a central data center.⁸⁶ The study found a significantly higher proportion of forms with rule-detected discrepancies (queries) for paper-based forms as compared to EDC (46.5% vs. 31.7%), with the odds of having an error from one to two times higher for the paper process.⁸⁶ The average rate of both univariate and multivariate discrepancies was greater for paper-based forms than EDC forms.⁸⁶ Meadows found a statistically significant higher proportion of resolved discrepancies for EDC, as compared to paper-based forms (62% vs. 48%); i.e., discrepancies on EDC forms

were 2.2 times more likely to be resolved compared to discrepancies on paper forms.⁸⁶ This varied across the analyzed form types from 1.32 to 3.76 times greater likelihood for discrepancy resolution for EDC forms as compared to paper.⁸⁶

Evaluation 9

Mitchel et al., published in 2006

A contemporaneous observational case study conducted by Mitchel et al. (2006) reported metrics obtained from the first year of data collection on a 170 patient prostate cancer trial conducted using web-based EDC.⁷² They reported that 85.9% of all forms did not require any data modification. Of the forms evaluated, concomitant medication and adverse event forms required more data corrections (29% and 29.8%, respectively, required modification). Visit date and demographics forms required little correction with EDC (99.3% and 96%, respectively, requiring no modification).⁷² When forms required modification, 76%–99.5% across form types from concomitant medications (76%) to the visit date form (99.5%) had the final modification within 30 days of data entry.⁷² The high percentage of clean data early in the study was noted as a clear advantage of EDC.⁷² Regarding monitoring, 69%–83% of all forms were reviewed by the CRA in the field within 60 days of data entry or final form modification.⁷² Further, 9.8% of all forms were reviewed by the in-house data reviewers on the same day the monitor reviewed the form, i.e., immediately after the monitor completed SDV.⁷² In general, 88%–97% of all forms were reviewed by the clinical data manager within 60 days of form review by the clinical trial monitor, facilitating early and ongoing “by patient” locking of the database.⁷²

Evaluation 10

Nahm et al., published in 2008

In the tenth identified study quantitatively evaluating data quality, Nahm, et al. (2008) reported observational metrics from completed source-to-database audits of 24 sites participating in four EDC trials.⁵⁹ All trials and sites audited used study-standardized, paper worksheets as source documents for capturing trial data. Data from these CRF-like worksheets were single-entered by site staff into an EDC system with extensive on-screen checks.⁵⁹ The average error rate across all four trials assessed was 14.3 errors per 10,000 fields, with a 95% Confidence Interval (averaged across audit Confidence Intervals) of 12–39 errors per 10,000 fields.⁵⁹ This compared favorably with a contemporaneous pooled analysis of data discrepancy and error rates measured for single and double entered data in other studies,¹¹⁹ indicating that single-entry at sites with intensive on-screen edit checks can produce data quality comparable to centrally, double entered data.

Evaluation 11

Mitchel et al., published in 2006

In the eleventh evaluation of EDC identified for this review, Mitchel et al. (2011) reported operational metrics from a multicenter clinical trial investigating the efficacy and safety of a new treatment in 492 randomized men.⁸¹ The trial was conducted using EDC and data were transcribed

from paper source documents to the EDC systems.⁸¹ Of the 2,584 data changes, 71.1% were designated due to data entry errors, 18.8% due to additional information, and 10.1% due to other reasons. The data were also analyzed by form. While the means of trial variables did not change appreciably from before to after data cleaning, in all cases the estimate of the standard deviation was smaller after cleaning than before cleaning.⁸¹

Evaluation 12

Pawellek et al., published in 2012

The last evaluation identified was reported by Pawellek et al. in 2012.⁸⁷ Use of a commercial EDC system in a multi-center double-blind randomized clinical trial conducted in eleven centers across five European countries was evaluated.⁸⁷ The EDC software was pre-loaded on laptop computers provided to sites. Early visits were conducted using paper forms, presenting the opportunity for comparing observational pretest and posttest metrics to identify differences associated with the use of EDC at one site. Following EDC data collection, plausibility of single-anthropometric values, as well as changes in the values between two study time points, were checked independently of the EDC process. These checks identified data anomalies in 14.6% of visits documented by EDC compared to 35.6% of visits documented with paper-based CRFs (Chi-squared test, $P < 0.001$).⁸⁷ Overall, 44.0% of all data anomalies detected by the independent data management checks were detectable by the automatic checks implemented in the eCRF.⁸⁷ Pawellek et al. concluded that the need for after-trial plausibility checks of anthropometric data was significantly reduced for eCRF-collected data compared to data collected on paper and that, “the planning and implementation process before starting the trial is more time-consuming” for studies collecting data via EDC than via paper forms.⁸⁷

Competing Interests

The authors are collaborating on the development of open source software to extract data directly from EHRs for multicenter clinical studies.

References

1. **Zozus MN, Topaloglu U, Collins C**, et al. Requirements for data acquisition and use of electronic health record (EHR) data during multicenter clinical studies. *Therapeutic Innovation & Regulatory Science (TIRS)*. 2020; In press.
2. **Food and Drug Administration**. Guidance for industry: electronic source data in clinical investigations. In: U.S. Department of Health and Human Services, ed. September 2013.
3. Electronic Records; Electronic Signatures. In: Food and Drug Administration, U.S. Department of Health and Human Services, ed. 21. 1997.
4. **Wilkinson M, Young R, Harper B, Machion B, Getz K**. Baseline Assessment of the Evolving 2017 eClinical Landscape. *Ther Innov Regul Sci*. 2019; 53(1): 71–80. DOI: <https://doi.org/10.1177/2168479018769292>
5. **Forrest WH, Jr., Bellville JW**. The use of computers in clinical trials. *Br J Anaesth*. 1967; 39(4): 311–319. DOI: <https://doi.org/10.1093/bja/39.4.311>
6. **Collen MF**. Clinical research databases—a historical review. *J Med Syst*. 1990; 14(6): 323–344. DOI: <https://doi.org/10.1007/BF00996713>
7. **Bill J, Anderson R, O’Fallon J, Silvers A**. Development of a computerized cancer data management system at the Mayo Clinic. *Int J Biomed Comput*. 1978; 9. DOI: [https://doi.org/10.1016/0020-7101\(78\)90054-5](https://doi.org/10.1016/0020-7101(78)90054-5)
8. **Helms RW**. *Entering data from remote terminals in clinical centers using IBM’s OS/TSO in the Kidney Transplant Histocompatibility Study*. Chapel Hill, NC: University of North Carolina; 1973. Technical report 007.
9. **Helms R**. Data quality issues in electronic data capture. *Drug Information Journal*. 2001; 35: 827–837. DOI: <https://doi.org/10.1177/009286150103500320>
10. **Black D, Molvig K, Bagniewska A**. A distributed data processing system for a multicenter clinical trial. *Drug Information Journal*. 1986; 20: 83–92. DOI: <https://doi.org/10.1177/009286158602000113>
11. **Prud’homme GJ, Canner PL, Cutler JA**. Quality assurance and monitoring in the Hypertension Prevention Trial. Hypertension Prevention Trial Research Group. *Control Clin Trials*. 1989; 10(3 Suppl): 84S–94S. DOI: [https://doi.org/10.1016/0197-2456\(89\)90044-5](https://doi.org/10.1016/0197-2456(89)90044-5)
12. **Neaton JD, Duchene AG, Svendsen KH, Wentworth D**. An examination of the efficiency of some quality assurance methods commonly employed in clinical trials. *Stat Med*. 1990; 9(1–2): 115–123; discussion 124. DOI: <https://doi.org/10.1002/sim.4780090118>
13. **Hilner JE, McDonald A, Van Horn L**, et al. Quality control of dietary data collection in the CARDIA study. *Control Clin Trials*. 1992; 13(2): 156–169. DOI: [https://doi.org/10.1016/0197-2456\(92\)90021-Q](https://doi.org/10.1016/0197-2456(92)90021-Q)
14. **Higgins SB, Jiang K, Plummer WD, Jr**, et al. Pivot/Remote: a distributed database for remote data entry in multi-center clinical trials. *Medinfo*. 1995; 8 Pt 2:1097.
15. **Stone EJ, Osganian SK, McKinlay SM**, et al. Operational design and quality control in the CATCH multicenter Trial. *Prev Med*. 1996; 25(4): 384–399. DOI: <https://doi.org/10.1006/pmed.1996.0071>
16. **McFadden ET, LoPresti F, Bailey LR, Clarke E, Wilkins PC**. Approaches to data management. *Control Clin Trials*. 1995; 16(2 Suppl): 30S–65S. DOI: [https://doi.org/10.1016/0197-2456\(94\)00093-1](https://doi.org/10.1016/0197-2456(94)00093-1)
17. **Summa W**. Electronic data capture: automated management of clinical trial data. *Pharmind*. 2004; 66(5a): 623–630.
18. **Simon R**. A decade of progress in statistical methodology for clinical trials. *Stat Med*. 1991; 10(12): 1789–1817. DOI: <https://doi.org/10.1002/sim.4780101203>
19. **Lampe AJ, Weiler JM**. Data capture from the sponsors’ and investigators’ perspectives: balancing quality, speed, and cost. *Drug Information*

- Journal*. 1998; 32: 811–886. DOI: <https://doi.org/10.1177/009286159803200403>
20. **Hyde AW**. The changing face of electronic data capture: from remote data entry to direct data capture. *Drug Information Journal*. 1998; 32: 1089–1092. DOI: <https://doi.org/10.1177/009286159803200429>
 21. **Kubick WR**. The elegant machine: applying technology to optimize clinical trials. *Drug Information Journal*. 1998; 32: 861–869. DOI: <https://doi.org/10.1177/009286159803200402>
 22. **Prove J**. Challenges and solutions for the use of remote study monitoring in a transcontinental project. *Drug Information Journal*. 2000; 34: 121–127. DOI: <https://doi.org/10.1177/009286150003400117>
 23. **Kronmal RA, Davis K, Fisher LD, Jones RA, Gillespie MJ**. Data management for a large collaborative clinical trial (CASS: Coronary Artery Surgery Study). *Comput Biomed Res*. 1978; 11(6): 553–566. DOI: [https://doi.org/10.1016/0010-4809\(78\)90034-4](https://doi.org/10.1016/0010-4809(78)90034-4)
 24. **Mitchell HE**. Distributed data management and processing of multicenter data. *American Statistical Association*; August, 1984; Philadelphia, PA.
 25. **Waldron HA, Cookson RF**. Use of a viewdata system to collect data from a multicentre clinical trial in anaesthesia. *Br Med J (Clin Res Ed)*. 1984; 289(6451): 1059–1061. DOI: <https://doi.org/10.1136/bmj.289.6451.1059>
 26. **Santoro E, Nicolis E, Franzosi MG, Tognoni G**. Internet for clinical trials: past, present, and future. *Control Clin Trials*. 1999; 20(2): 194–201. DOI: [https://doi.org/10.1016/S0197-2456\(98\)00060-9](https://doi.org/10.1016/S0197-2456(98)00060-9)
 27. **Bagniewska A, Black D, Molvig K, et al**. Data quality in a distributed data processing system: the SHEP Pilot Study. *Control Clin Trials*. 1986; 7(1): 27–37. DOI: [https://doi.org/10.1016/0197-2456\(86\)90005-X](https://doi.org/10.1016/0197-2456(86)90005-X)
 28. **Pogash RM, Boehmer SJ, Forand PE, Dyer AM, Kunselman SJ**. Data management procedures in the Asthma Clinical Research Network. *Control Clin Trials*. 2001; 22(6 Suppl): 168S–180S. DOI: [https://doi.org/10.1016/S0197-2456\(01\)00170-2](https://doi.org/10.1016/S0197-2456(01)00170-2)
 29. **Hollingsworth RA, Hay C, Richards B**. An Internet implementation of an international clinical study. *Stud Health Technol Inform*. 1999; 68: 528–531.
 30. **Kiuchi T, Kaihara S**. Automated generation of a World Wide Web-based data entry and check program for medical applications. *Comput Methods Programs Biomed*. 1997; 52(2): 129–138. DOI: [https://doi.org/10.1016/S0169-2607\(96\)01793-2](https://doi.org/10.1016/S0169-2607(96)01793-2)
 31. **Kiuchi T, Ohashi Y, Konishi M, Bandai Y, Kosuge T, Kakizoe T**. A World Wide Web-based user interface for a data management system for use in multi-institutional clinical trials—development and experimental operation of an automated patient registration and random allocation system. *Control Clin Trials*. 1996; 17(6): 476–493. DOI: [https://doi.org/10.1016/S0197-2456\(96\)00104-3](https://doi.org/10.1016/S0197-2456(96)00104-3)
 32. **Keim E, Sippel H, Eich HP, Ohmann C**. Collection of data in clinical studies via Internet. *Stud Health Technol Inform*. 1997; 43 Pt A: 57–60.
 33. **Kelly MA, Oldham J**. The Internet and randomised controlled trials. *Int J Med Inform*. 1997; 47(1–2): 91–99. DOI: [https://doi.org/10.1016/S1386-5056\(97\)00091-9](https://doi.org/10.1016/S1386-5056(97)00091-9)
 34. **Workman R, Beatty E, Workman D**. Internet based data collection and analysis. *Stud Health Technol Inform*. 1998; 51: 182–185.
 35. **Kuchenbecker J, Dick HB, Schmitz K, Behrens-Baumann W**. Use of internet technologies for data acquisition in large clinical trials. *Telemed J E Health*. 2001; 7(1): 73–76. DOI: <https://doi.org/10.1089/153056201300093976>
 36. **Wubbelt P, Fernandez G, Heymer J**. Clinical trial management and remote data entry on the Internet based on XML case report forms. *Stud Health Technol Inform*. 2000; 77: 333–337.
 37. **Dimenas E, Johansson D, Palmblad M, Wrangstadh M**. Clinical Operations Online (COOL)—A World Wide Web-based approach to running clinical trials: Results from two international multicenter gastrointestinal trials. *Drug Inf J*. 2001; 35: 745–753. DOI: <https://doi.org/10.1177/009286150103500313>
 38. **Marks RG, Conlon M, Ruberg SJ**. Paradigm shifts in clinical trials enabled by information technology. *Stat Med*. 2001; 20(17–18): 2683–2696. DOI: <https://doi.org/10.1002/sim.736>
 39. **Brandt CA, Nadkarni P, Marenco L, et al**. Reengineering a database for clinical trials management: lessons for system architects. *Control Clin Trials*. 2000; 21(5): 440–461. DOI: [https://doi.org/10.1016/S0197-2456\(00\)00070-2](https://doi.org/10.1016/S0197-2456(00)00070-2)
 40. **Sippel H, Ohmann C**. A web-based data collection system for clinical studies using Java. *Med Inform (Lond)*. 1998; 23(3): 223–229. DOI: <https://doi.org/10.3109/14639239809001402>
 41. **Lallas CD, Preminger GM, Pearle MS, et al**. Internet based multi-institutional clinical research: a convenient and secure option. *J Urol*. 2004; 171(5): 1880–1885. DOI: <https://doi.org/10.1097/01.ju.0000120221.39184.3c>
 42. **Rangel SJ, Narasimhan B, Geraghty N, Moss RL**. Development of an internet-based protocol to facilitate randomized clinical trials in pediatric surgery. *J Pediatr Surg*. 2002; 37(7): 990–994; discussion 990–994. DOI: <https://doi.org/10.1053/jpsu.2002.33826>
 43. **Unutzer J, Choi Y, Cook IA, Oishi S**. A web-based data management system to improve care for depression in a multicenter clinical trial. *Psychiatr Serv*. 2002; 53(6): 671–673, 678. DOI: <https://doi.org/10.1176/ps.53.6.671>
 44. **Chadwick BJ, Nonemaker S, Bien MR**. Realize maximum value when implementing electronic data capture. *Applied Clinical Trials*. 2002(February): 36–40.
 45. **Spink C**. Electronic Data Capture (EDC) as a means for e-clinical trial success. In: *IBM Global Services*; 2002.
 46. **Sahoo U, Bhatt A**. Electronic data capture (EDC)—a new mantra for clinical trials. *Qual*

- Assur*. 2003; 10(3–4): 117–121. DOI: <https://doi.org/10.1080/10529410390892052>
47. **Litchfield J, Freeman J, Schou H, Elsley M, Fuller R, Chubb B.** Is the future for clinical trials internet-based? A cluster randomized clinical trial. *Clin Trials*. 2005; 2(1): 72–79. DOI: <https://doi.org/10.1191/1740774505cn069oa>
 48. **Green J.** Realizing the value proposition of EDC. *Innovations in Clinical Trials*; 2003.
 49. **Bunn G.** Scaling up EDC. How to move away from paper trials. *Applied Clinical Trials*. 2002; 12–14.
 50. **Editors ACT.** Supporting EDC in a clinical trial environment. *Applied Clinical Trials*. 2002; 16–19.
 51. **Mitchel JT, You J, Kim YJ,** et al. Internet-based clinical trials: practical considerations. *Pharmaceutical Development and Regulation*. 2003; 1(1): 29–39. DOI: <https://doi.org/10.1007/BF03257363>
 52. **CenterWatch.** EDC Adoption in Clinical Trials: A 2008 Analysis. *BioITWorldcom*. 2008; February 2008.
 53. **El Emam K, Jonker E, Sampson M, Krleza-Jeric K, Neisa A.** The use of electronic data capture tools in clinical trials: Web-survey of 259 Canadian trials. *J Med Internet Res*. 2009; 11(1): e8. DOI: <https://doi.org/10.2196/jmir.1120>
 54. **Ene-Iordache B, Carminati S, Antiga L,** et al. Developing regulatory-compliant electronic case report forms for clinical trials: experience with the demand trial. *J Am Med Inform Assoc*. 2009; 16(3): 404–408. DOI: <https://doi.org/10.1197/jamia.M2787>
 55. **Cramon P, Rasmussen AK, Bonnema SJ,** et al. Development and implementation of PROgmatic: A clinical trial management system for pragmatic multi-centre trials, optimised for electronic data capture and patient-reported outcomes. *Clin Trials*. 2014; 11(3): 344–354. DOI: <https://doi.org/10.1177/1740774513517778>
 56. **Arab L, Hahn H, Henry J, Chacko S, Winter A, Cambou MC.** Using the web for recruitment, screen, tracking, data management, and quality control in a dietary assessment clinical validation trial. *Contemp Clin Trials*. 2010; 31(2): 138–146. DOI: <https://doi.org/10.1016/j.cct.2009.11.005>
 57. **Pavlovic I, Miklavcic D.** Web-based electronic data collection system to support electrochemotherapy clinical trial. *IEEE Transactions on Information Technology in Biomedicine*. 2007; 11(2): 222–230. DOI: <https://doi.org/10.1109/TITB.2006.879581>
 58. **Banik N, Mochow O.** Evaluation of EDC versus paper in a multinational asthma trial. In: Zozus M, ed. *Slides from the 8th Annual European Workshop on Clinical Data Management*. Dr. Norbert Banik, presentation slides provided by email communication, May 20, 2020. ed. Berlin, Germany: Drug Information Association; 1998.
 59. **Nahm ML, Pieper CF, Cunningham MM.** Quantifying data quality for clinical trials using electronic data capture. *PLoS ONE*. 2008; 3(8): e3049. DOI: <https://doi.org/10.1371/journal.pone.0003049>
 60. **Lu Z.** Electronic data-capturing technology for clinical trials, experience with a global postmarketing study. In. *IEEE Engineering in Medicine and Biology*. Vol March/April IEEE; 2010.
 61. **Staziaki PV, Kim P, Vadvala HV, Ghoshhajra BB.** Medical Registry Data Collection Efficiency: A Crossover Study Comparing Web-Based Electronic Data Capture and a Standard Spreadsheet. *J Med Internet Res*. 2016; 18(6): e141. DOI: <https://doi.org/10.2196/jmir.5576>
 62. **Anderson NR, Lee ES, Brockenbrough JS,** et al. Issues in biomedical research data management and analysis: needs and barriers. *J Am Med Inform Assoc*. 2007; 14(4): 478–488. DOI: <https://doi.org/10.1197/jamia.M2114>
 63. **Nahm M, Zhang J.** Operationalization of the UFuRT Methodology in the Clinical Research Domain. *Journal of Biomedical Informatics (in press)*; 2009. DOI: <https://doi.org/10.1016/j.jbi.2008.10.004>
 64. **Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG.** Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics (in press)*; 2008. DOI: <https://doi.org/10.1016/j.jbi.2008.08.010>
 65. **Welker JA.** Implementation of electronic data capture systems: barriers and solutions. *Contemp Clin Trials*. 2007; 28(3): 329–336. DOI: <https://doi.org/10.1016/j.cct.2007.01.001>
 66. **Lopez-Carrero C, Arriaza E, Bolanos E,** et al. Internet in clinical research based on a pilot experience. *Contemp Clin Trials*. 2005; 26(2): 234–243. DOI: <https://doi.org/10.1016/j.cct.2004.11.017>
 67. **Kush RD.** The future for electronic data capture. *Drug Development*. 2006; 46–47.
 68. **Howells K.** e-Clinical integration strategies. *Drug Discov Today Technol*. 2006; 3(2): 167–171. DOI: <https://doi.org/10.1016/j.ddtec.2006.06.009>
 69. **Stead W, Lin H.** (eds.). *Computational technology for effective health care: immediate steps and strategic directions, pre-publication copy*. Washington, DC: National Academy Press; 2009.
 70. **Mitchel JT, You J, Kim YJ,** et al. Clinical Trial Data Integrity using internet technology to Collect Reliable Data. *Applied Clinical Trials*; 2003. 6–8.
 71. **Marks R, Bristol H, Conlon M, Pepine CJ.** Enhancing clinical trials on the internet: lessons from INVEST. *Clin Cardiol*. 2001; 24(11 Suppl): V17–23. DOI: <https://doi.org/10.1002/clc.4960241707>
 72. **Mitchel JT, Kim YJ, Choi J, Hays V, Langendorf J, Cappi S.** Impact of IBCTs on Clinical Trial Efficiency. *Applied Clinical Trials*. 2006; 2006(August): 62–68.
 73. **Paul J, Seib R, Prescott T.** The Internet and clinical trials: background, online resources, examples

- and issues. *J Med Internet Res*. 2005; 7(1): e5. DOI: <https://doi.org/10.2196/jmir.7.1.e5>
74. **MacGarvey A**. EDC state of the art. *Innovations in Pharmaceutical Technology*. 2005; 116–118.
 75. **Mitchel JT, Ernst C, Cappi S**, et al. Implementing internet-based clinical trials. *DIA Forum*. 2004; 40(October): 22–23.
 76. **Kush RD, Bleicher P, Kubick WR**, et al. *eClinical Trials, Planning and Implementation*. Boston, MA: Thompson Centerwatch; 2003.
 77. **Mitchel JT, Kim YJ, Choi J**, et al. The impact of electronic data capture on clinical data management perspectives from the present into the future. *MONITOR*. 2008(August): 37–41.
 78. **Mitchel JT, Kim YJ, Choi J, Park G, Suciú L, Horn M**. The final eFrontier. *Applied Clinical Trials*. 2010; May 1.
 79. **Whyte J, Vasterling J, Manley GT**. Common data elements for research on traumatic brain injury and psychological health: current status and future development. *Arch Phys Med Rehabil*. 91(11): 1692–1696. DOI: <https://doi.org/10.1016/j.apmr.2010.06.031>
 80. **Richardson A**. Planing and running the eClinical Trial. In: *Applied Clinical Trials*. Vol January 2003.
 81. **Mitchel JT, Kim YJ, Choi J**, et al. Evaluation of Data Entry Errors and Data Changes to an Electronic Data Capture Clinical Trial Database. *Drug Inf J*. 2011; 45(4): 421–430. DOI: <https://doi.org/10.1177/009286151104500404>
 82. **Cannon CP, Battler A, Brindis RG**, et al. American College of Cardiology key data elements and definitions for measuring the clinical management and outcomes of patients with acute coronary syndromes. A report of the American College of Cardiology Task Force on Clinical Data Standards (Acute Coronary Syndromes Writing Committee). *J Am Coll Cardiol*. 2001; 38(7): 2114–2130. DOI: [https://doi.org/10.1016/S0735-1097\(01\)01702-8](https://doi.org/10.1016/S0735-1097(01)01702-8)
 83. **Bart T**. Comparison of electronic data capture with paper data collection – is there really an advantage? *Pharmatech*. 2003; 1–4.
 84. **Brown EG, Holmes BJ, McAulay SE**. Clinical trials EDC endgame. In: Forrester Research, Inc.; 2004: 12.
 85. **Mitchel JT, You J, Lau A, Kim YJ**. Paper vs. web: a tale of three trials. *Applied Clinical Trials*. 2001(August): 34–36.
 86. **Meadows B**. *A comparison of paper-based data submission to remote data capture for minimizing data entry errors in cancer clinical research*. Dissertation submitted to the faculty of the Graduate School of the University of Maryland Baltimore in partial fulfillment of the requirements for the degree of Doctor of Philosophy. University of Maryland Baltimore 2006.
 87. **Pawellek I, Richardsen T, Oberle D, Grote V, Koletzko B**. Use of electronic data capture in a clinical trial on infant feeding. *European Journal of Clinical Nutrition*. 2012; 66: 1342–1343. DOI: <https://doi.org/10.1038/ejcn.2012.141>
 88. **Sung NS, Crowley WF, Jr., Genel M**, et al. Central challenges facing the national clinical research enterprise. *JAMA*. 2003; 289(10): 1278–1287. DOI: <https://doi.org/10.1001/jama.289.10.1278>
 89. **Takasaki M, Momosaki R, Wakabayashi H, Nishioka S**. Construction and Quality Evaluation of the Japanese Rehabilitation Nutrition Database. *J Nutr Sci Vitaminol (Tokyo)*. 2018; 64(4): 251–257. DOI: <https://doi.org/10.3177/jnsv.64.251>
 90. **Zozus MN, Kahn M, Wieskopf N**. Data quality in clinical research. In: Richesson RL, Andrews JE (eds.), *Clinical research informatics*. 2nd ed. Switzerland: Springer; 2019. DOI: https://doi.org/10.1007/978-3-319-98779-8_11
 91. **Desjardins J**. The rising speed of technological adoption. *Visual Capitalist Web site*. <https://www.visualcapitalist.com/rising-speed-technological-adoption/>. Published 2018. Updated February 14, 2018. Accessed December 13, 2020.
 92. **Ritchie H, Roser M**. Technology Adoption. *OurWorldInData.org*. 'https://ourworldindata.org/technology-adoption'. Published 2017. Accessed December 20, 2020.
 93. **Mitchel JT, Gittleman D, Park G**, et al. The Impact on Clinical Research Sites When Direct Data Entry Occurs at the Time of the Office Visit: A Tale of 6 Studies. In. *InSite*. Vol Second Quarter 2014.
 94. **Laky D**. The evolution of EDC into eClinical. In. *epc*. Spring ed: samedan Ltd. Pharmaceutical Publishers; 2007.
 95. **Waife RS**. Transitioning clinical data management from the 1980s to the 2010s. *Drug Information Journal*. 2001; 35(3): 713–719. DOI: <https://doi.org/10.1177/009286150103500309>
 96. **Kush R**. *The Cost of Clinical Data Interchange in Clinical Trials: A CDISC White Paper*. Austin, TX: Clinical Data Interchange Standards Consortium; 2001.
 97. **Haak D, Page CE, Reinartz S, Kruger T, Deserno TM**. DICOM for Clinical Research: PACS-Integrated Electronic Data Capture in Multi-Center Trials. *J Digit Imaging*. 2015; 28(5): 558–566. DOI: <https://doi.org/10.1007/s10278-015-9802-8>
 98. **Comulada WS, Tang W, Swendeman D, Cooper A, Wacksman J, Adolescent Medicine Trials Network CT**. Development of an Electronic Data Collection System to Support a Large-Scale HIV Behavioral Intervention Trial: Protocol for an Electronic Data Collection System. *JMIR Res Protoc*. 2018; 7(12): e10777. DOI: <https://doi.org/10.2196/10777>
 99. **Aboulelenein S, Williams T, Baldner J, Zozus MN**. Analysis of professional competencies for the clinical research data management profession. *Data Basics*. 2020; 26(1): 6–17.
 100. **Zozus MN, Lazarov A, Smith LR**, et al. Analysis of professional competencies for the clinical research

- data management profession: implications for training and professional certification. *J Am Med Inform Assoc.* 2017; 24(4): 737–745. DOI: <https://doi.org/10.1093/jamia/ocw179>
101. **Handelsman D.** Electronic data capture: when will it replace paper? Vol 2020. December 17, 2009 ed. SAS, Cary NC: SAS Inc.; 2009.
 102. **Miksad RA, Abernethy AP.** Harnessing the Power of Real-World Evidence (RWE): A Checklist to Ensure Regulatory-Grade Data Quality. *Clin Pharmacol Ther.* 2018; 103(2): 202–205. DOI: <https://doi.org/10.1002/cpt.946>
 103. **Fleurence RL, Shuren J.** Advances in the Use of Real-World Evidence for Medical Devices: An Update From the National Evaluation System for Health Technology. *Clin Pharmacol Ther.* 2019; 106(1): 30–33. DOI: <https://doi.org/10.1002/cpt.1380>
 104. **FDA.** *Framework for FDA's Real-World Evidence Program.* U.S. Department of Health and Human Services; 2018.
 105. *Examining causes of and potential solutions to clinical data management cycle time challenges.* Tufts Center for the Study of Drug Development, Tufts University; 2018.
 106. **Getz KA, Campo RA.** New Benchmarks Characterizing Growth in Protocol Design Complexity. *Ther Innov Regul Sci.* 2018; 52(1): 22–28. DOI: <https://doi.org/10.1177/2168479017713039>
 107. **Mitchel JT, Weingard K, Markowitz JMS, Gittleman D, Efros MD.** How direct data entry at the time of the patient visit is transforming clinical research: perspective from the clinical trial research site. *InSite.* 2013; 2013(Second quarter): 40–43.
 108. The eClinical Forum and PhRMA EDC/eSource Taskforce. *The future vision of electronic health records as eSource for clinical research.* September 14 2006.
 109. **Garza M, 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. American Medical Informatics Association; 2020 in press.
 110. **Garza M, Nordo A, Eisenstein EL, Hammond WE, Walden A, Zozus MN.** EHR-to-eCRF information exchange in clinical trials: a systematic review Submitted to Information Technology and Communications in Healthcare; 2018; Victoria Canada.
 111. **Arndt S, Tyrrell G, Woolson RF, Flaum M, Andreassen NC.** Effects of errors in a multicenter medical study: preventing misinterpreted data. *J Psychiatr Res.* 1994; 28(5): 447–459. DOI: [https://doi.org/10.1016/0022-3956\(94\)90003-5](https://doi.org/10.1016/0022-3956(94)90003-5)
 112. **Shortliffe E, Cimino J.** (eds.). *Biomedical Informatics: Computer Applications in Healthcare and Biomedicine.* 2 ed. New York: Springer Science Publications; 2013. DOI: <https://doi.org/10.1007/978-1-4471-4474-8>
 113. **Mitchel JT, Weingard K, Markowitz JMS, Gittleman D, Efros MD.** Three pronged approach to optimizing clinical trial monitoring. *Applied Clinical Trials.* 2014; 23(6): 37–44.
 114. **Solow R.** We'd better watch out. *New York Times Book Review.* July 12, 1987: 36.
 115. **Rogers E.** *Diffusion of Innovations.* 5th ed. New York: Simon and Schuster; 2003.
 116. **Christensen CM.** *The innovator's dilemma: when new technologies cause great firms to fail.* Boston, MA: Harvard Business Review Press; 2016.
 117. **Christensen C.** *The innovator's dilemma: When new technologies cause great firms to fail.* Boston, MA: Harvard Business School Press; 1997.
 118. **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.
 119. **Nahm M.** Data quality in clinical research. In: Richesson R, Andrews J, eds. *Clinical research informatics.* New York: Springer-Verlag; 2012. DOI: https://doi.org/10.1007/978-1-84882-448-5_10

How to cite this article: Zozus MN, Sanns B, 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>

Submitted: 24 December 2020

Accepted: 23 January 2021

Published: 12 March 2021

Copyright: © 2021 SCDM publishes JSCDM content in an open access manner under a Attribution-Non-Commercial-ShareAlike (CC BY-NC-SA) license. This license lets others remix, adapt, and build upon the work non-commercially, as long as they credit SCDM and the author and license their new creations under the identical terms. See <https://creativecommons.org/licenses/by-nc-sa/4.0/>.



Journal of the Society for Clinical Data Management is a peer-reviewed open access journal published by Society for Clinical Data Management.

OPEN ACCESS