2021

ConnEDCt, a mobile-first framework for clinical Electronic Data Capture

BOSTON UNIVERSITY

METROPOLITAN COLLEGE

Thesis

CONNEDCT, A MOBILE-FIRST FRAMEWORK FOR CLINICAL ELECTRONIC DATA CAPTURE

by

CALEB J. RUTH

B.S., Skidmore College, 1993

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2021

Approved by

First Reader ______________________________

Guanglan Zhang, Ph.D.
Associate Professor of Computer Science

Second Reader ______________________________

Lou Chitkushev, Ph.D.
Associate Dean for Academic Affairs
Associate Professor of Computer Science

Third Reader ______________________________

Saurabh Mehta, Sc.D.
Associate Professor of Global Heath, Epidemiology, and Nutrition
Cornell University

ACKNOWLEDGMENTS

I am grateful to several people for their assistance and support on this project. Above all for the strong support, encouragement, and patience of Professor Guanglan Zhang, my advisor in masters studies at Boston University. Secondly to Professor Saurabh Mehta of Cornell University for the opportunity to develop ConnEDCt in the first place; and to Dr. Elaine Yu and Dr. Samantha Huey for their collaboration on the initial implementations of the platform. I also gratefully acknowledge Jacqulyn Williams, Elizabeth Lance and the rest of the Writing Together writing support group sponsored by Virginia Commonwealth University – Qatar who kept me inspired and on track during the final stages of completing this manuscript during the COVID-19 lockdown of 2020. Finally, to my wife Dr. Kuei-Chiu Chen and my daughter Naomi who encouraged me to commit to the work in the first place and consistently supported me throughout the process.

CONNEDCT, A MOBILE-FIRST FRAMEWORK FOR CLINICAL ELECTRONIC DATA CAPTURE

CALEB J. RUTH

ABSTRACT

Paper-based data capture has long served as the primary means of collecting research data and continues to be the dominant means of data capture through the present day. Despite inertia with adopting information technology in clinical research, electronic methods of information capture have important benefits over traditional, paper-based methods. Electronic Data Capture (EDC) systems can provide integrated error checking, protocol enforcement, decision support, automated randomization, and quicker access to data and results. As EDC systems become more accessible and resourceful, EDC has begun to replace paper-based data capture. Meanwhile, mobile computing, utilizing smartphones and tablets, has become commonplace in business and our everyday lives. Many EDC solutions support mobile devices, yet few were conceived with a "mobile-first" design philosophy and fewer support extensive study protocol-support features. A significant amount of clinical research is conducted in geographic regions with limited or no Internet access such as impoverished and remote communities. Current EDC solutions remain challenging to use in these contexts.

While EDC is an increasingly important tool for clinical research, when EDC solutions are built on web-centric architectures, the lack of Internet coverage means that researchers often need to fall back on paper-based data capture methods or build expensive, custom EDC tools. A customizable Mobile Electronic Data Capture (mEDC) framework with an asynchronous data transport layer will better meet the needs of distributed studies in resource-limited, geographical areas. I developed ConnEDCt, a full-featured mEDC application that is customizable for longitudinal study protocols, with regulatory-compliant security, auditability and an asynchronous data transport model. ConnEDCt is adaptable to different study protocols, has extensive study protocol-support built-in, and supports on- or off-line data synchronization to a central data repository. ConnEDCt focuses on mobility and is designed to serve the needs of complex clinical research studies in regions where other EDC platforms cannot be utilized.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

API	Application Programming Interface
CDM	Central Data Management
CDR	Central Data Repository
CDS	Clinical Decision Support
CFR	Code of Federal Regulations
CRF	Case Report Form
DM	Data Manager
eCRF	Electronic Case Report Form
EDC	Electronic Data Capture
ERD	Entity Relationship Diagram
HIPAA	Health Information Portability and Accountability Act
HIT	Health Information Technology
ICT	Information and Communication Technology
IDE	Integrated Development Environment
IRB	Institutional Review Board
JSON	JavaScript Object Notation
mEDC	Mobile Electronic Data Capture
PC	Personal Computer

PDC	Paper-based Data Capture
PHI	Protected Health Information
PI	Principal Investigator
POC	Point of Capture
RA	Research Associate
RCT	Randomized Controlled Trial
RDBMS	Relational Database Management System
SOP	Standard Operating Procedure
UI	User Interface
UUID	Universally Unique Identifier

1. INTRODUCTION

As fluorescent lights flickered and flies buzzed, my eyes traced the path of the catheter as it emerged from a hole drilled in the man's bandaged head and lead to an open glass jar under his hospital cot. His swollen brain was squeezing cerebral-spinal fluid from the ventricles in his brain. The murky liquid gathered in the glass jar under his bed. This was the late 1990's in post-Soviet Eastern Europe. The man in front of me on the cot was a victim of head trauma, an everyday accident for ordinary people. He was one of several similar patients in the ICU of this teaching hospital in Ljubljana, Slovenia. The hospital was struggling to provide adequate care.

The neurosurgeons and other attending physicians gathered around the patient and reviewed his medical chart on a clipboard while a neurosurgery resident used a new IBM ThinkPad laptop to enter medical data into the Traumatic Brain Injury Survey database (TBIS). The TBIS program was conducted to support an evidence-based treatment protocol, the Guidelines for the Management of Traumatic Brain Injury. I, along with the clinical team, built the TBIS database and managed the information technology on the project.

Figure 1.1. The author demonstrating the Traumatic Brain Injury Survey (TBIS) EDC tool to a group of physicians c. 1999

Research fellows, usually neurosurgery hospital residents, entered comprehensive data relevant to brain injury for up to 12 months post-injury into TBIS. The database captured the data and provided decision support, using protocols published in the Guidelines. Based on captured data, TBIS could highlight when a medical intervention - like drilling a hole in the cranium to relieve intra-cranial pressure - was recommended. The TBIS project provided medical training, a laptop with the TBIS app, a flatbed scanner for radiology films, and a dial-up Internet connection. The project goals were to train the Guidelines, monitor compliance, and validate and refine the Guidelines in follow-up analysis. The data was regularly delivered electronically to New York City where the project was coordinated.

The TBIS program was deployed in teaching hospitals distributed roughly linearly from Estonia on the Baltic Sea to Croatia on the Adriatic Sea, a north-south slice right through the middle of Europe. TBIS operated for several years and spawned three distinct data capture tools, including the primary clinical data capture app, an expanded version, and a tele-radiology diagnostic review app. It inspired new head trauma research programs in the U.S. and Europe. We developed these tools in an age of dial-up Internet when research data was almost exclusively captured on paper, Netflix was a DVD-by-mail service, Google was still an academic research project, and Wikipedia and Facebook hadn’t yet been created. TBIS showed that, despite limited resources, a database system with support from remote specialists could make an impact in the care of patients in the present and in the future.

This is how my adventure in Health Information Technology (HIT) began and how I came to believe in its value to improve the quality of medical science. I don’t know the outcome of the man I saw in the hospital in Ljubljana more than 20 years ago, but I’m confident that the research that his participation contributed to has helped trauma surgeons successfully treat innumerable patients of head trauma since then. Research coupled with technology has improved healthcare tremendously.

Today paper-based data capture (PDC) is still common because of low up-front costs, and few technical requirements. A pen and paper will do. However, it is also error-prone, carries a high risk of data-loss, takes time to retrieve from the field, and requires transcription in order to perform data analysis. Therefore, PDC is usually more expensive than it first appears when the total cost is tallied. Electronic Data Capture (EDC) can solve many of the issues related to PDC and provide additional benefits.

We now have a society accustomed to portable computing devices, such as phones and tablets, and a rich environment of HIT. Regulations are maturing to cope with privacy and ethics in HIT. Clinical research is a well-defined field that has developed over many hundreds of years. Meanwhile, informatics implementations for clinical research date from at least the 1980s. Nonetheless, it is only relatively recently that informatics systems have advanced to combine portability with platforms that support complex study protocols. Drs. Embi and Payne pointed out that "Clinical researchers are faced with significant and increasingly complex workflow and information management challenges. …effective and efficient information access is critical to any solution to the many challenges faced by the domain."[1] A complex clinical study protocol will often have requirements like informed consent, eligibility criteria, participant randomization, and clinical visit scheduling. These are complex design challenges for the clinical researcher as well as the information scientist.

In this thesis I report on the development and utilization of the ConnEDCt electronic data capture framework that helps structure the study protocol design process while enforcing protocol compliance and providing other benefits to clinical research teams. ConnEDCt prioritizes tablet computing and portability by design. A balance of app capability, ease of use, careful protocol design, some specialized programming, and flexibility – in short thoughtfulness – are the keys to *connect* the data to the research.

1.1. Background

Electronic data capture (EDC) systems are designed to facilitate structured data input into an electronic storage system for use with clinical research. "The main advantages of EDC would be to enter, review, and analyze data in real-time and to implement online data validation checks to assure data quality." [2] However, paper-based case report forms (CRFs) are often still used in clinical research. Limiting factors such as cost, technical requirements, or perceived shortcomings can hinder EDC implementation. Many EDC solutions require a full-time Internet connection to connect to a web site or have limited functionality without real-time Internet. Laptop computers lack the portability that tablets and mobile phones provide at the point of data capture (POC). When internet access is limited investigators often will fall back on paper-based CRFs. Data captured on paper must later be re-entered from the paper-based CRFs into an EDC system.[3] Utilizing paper-based CRFs and performing secondary entry into an EDC system erases many of the benefits of EDC such as error reduction, protocol enforcement, and faster data availability.

Recently, mobile devices such as tablets and mobile phones have begun to be used for EDC. Mobile EDC (mEDC) has the potential to overcome many of the limitations of standard EDC. However, current implementations of mEDC are still limited to custom systems or systems that require full-time Internet. Custom solutions require access to skilled, possibly expensive programmers and reliable internet access is still unavailable in much of the rural and developing world. [4, 5] These limitations can prohibit the use of EDC especially in poor, rural communities where research is most needed.

In 2015 the Mehta Research Group (MRG) of Cornell University was hitting these limitations while attempting to find an EDC solution for their research projects running off the grid in rural and under-developed areas in India and South America. MRG contacted this author to develop a system that overcame the limitations of existing EDC solutions. We confirmed that a gap exists for a mobile-EDC platform designed for sophisticated, longitudinal research that can operate in regions and communities that are off the modern, information grid. We developed a design for a framework that would be re-usable for multiple clinical studies and operate on mobile devices with limited internet access. This framework, named ConnEDCt, would support accepted clinical study protocol methodologies described below.

1.1.1. The Clinical Study Protocol

"The concept of a protocol is fundamental"[6] to clinical research and EDC. Data capture and supporting the study protocol are the two essential functions of clinical EDC. A clinical study protocol is a written document that includes the research objectives, scientific background, study design, and methods. [7] A research team led by a principal investigator (PI) designs the protocol while an Institutional Review Board (IRB) reviews and approves the protocol to ensure good research and ethical practices. [8] The aspects of a clinical study protocol that are relevant to data capture and informatics include *Case Report Forms*, *Study Schema*, *Informed Consent*, *Eligibility Criteria*, and *Randomization*.

1.1.2. Case Report Forms

A Case Report Form (CRF) is the basic design element of clinical research. The CRF is the tool used to collect participant data in a clinical study. [9] A CRF consists of a collection of related variables to be entered by interviewers or study participants. A single study is likely to rely on multiple CRFs organized by clinical topic or workflow logic. Researchers determine in advance what data elements will be captured. To facilitate the interview process, as well as later data analysis, CRFs are designed when possible with numeric or coded responses that are selected from predefined lists. [6] The design of CRFs is an in-depth process involving both technical and research team members and is a key step in ensuring the quality of the collected data. Effective CRF design includes attention to clarity of language, concise coding, layout design, organization, and workflow of data collection. [7] Various types of CRFs may collect baseline data, laboratory results, clinical observations, follow-up, and outcome data.

1.1.3. The Study Schema

The study schema is the data abstraction of the clinical study protocol and will define and enforce the workflow within the context of the EDC tool. The study schema defines when and how frequently the *CRFs* are to be completed during the course of the study. For example the *Study Schema* of a simple study may define a single participant encounter with one or more *CRFs*. A *Study Schema* of a longitudinal study may define a screening visit, a baseline clinical assessment, a number of midline assessments, and an outcome assessment. [7] The initial encounter with a potential study participant will include obtaining *informed consent* and assessing *eligibility criteria*. A baseline clinical assessment may collect demographic and initial clinical data. Midline assessments will record comparison data during the course of the study. The outcome assessment will collect the final data points and markers of the study. When data capture will be performed by different research staff or for different conceptual areas during one encounter, the questions are separated into multiple *CRFs*. Administrative staff may conduct the initial interview and capture demographic information while clinical staff may perform clinical screening and assessments. In future visits nursing staff may enter some *CRFs*, while physicians and laboratory staff complete others.

Adverse events, withdrawal from the study, or other emergent events may occur at any time during the course of the study and must be recorded. [6] These unscheduled events fall out of the predefined schedule of events, but still are accounted for in the *Study Schema*.

The study schema, as implemented in an informatics system, thus becomes both a design tool for EDC implementation as well as a compliance enforcement tool supporting data capture while the study is conducted. Richesson and Andrews referred to the informatics structure that defines which CRFs are to be filled out during which time points as an “Event-CRF Cross Table” [6]. In informatics terms this schedule of when to present CRFs for data capture is abstracted as a join table between the CRF table and the Event table. This table matches CRFs to Events in a many-to-many relationship. The benefit to the research assistant (RA) or other data entry staff is that when viewing a particular encounter in the Study Protocol, only the forms for that encounter will be presented. It will not be possible for the RA to complete a form that is not scheduled for that encounter. Conversely, all the forms scheduled for an encounter will be presented to the RA. Completing the forms scheduled for a visit can therefore be enforced.

1.1.4. Informed Consent

History documents many cases of tragic effects when medical researchers exploited patients by denying them information and the opportunity to consent to participate in clinical experiments. Section 3.1 *Ethics and Governing Regulations* details several of these cases of exploitation and abuse. In response to this appalling history of exploitation of vulnerable people, informed consent has become a standard for ensuring voluntary participation in clinical research.

The United States government has developed regulations to protect the rights of participants in clinical research. These regulations are referred to as "The Common Rule" and are explored in depth in Chapter 3. Other nations have developed their own regulations or follow the U.S. rules.

The purpose of informed consent is to protect human subjects enrolled in clinical studies. [8] Government regulations in the U.S. and other countries require that researchers inform participants in a clinical study of the risks of participation and voluntarily grant consent to be included in the study. Informed consent procedures may vary based on study protocols, nation-specific legal regulations, and the age of participants. Informed consent may occur multiple times during the course of a study and include consent for different facets of the study, for example discrete consents for participation in a study and collection of biological samples. Assent by a minor will require the consent of a legal guardian. Consent given by illiterate participants may require a third-party witness. In all cases informed consent forms must include information on the study, the potential risks and benefits of participation, and capture proof – such as a signature – of a participant's consent.[10]

1.1.5. Eligibility Criteria and Enrollment

Clinical Decision Support (CDS) is another significant benefit of computerized clinical systems. With access to data, computer systems can apply algorithms to make determinations and recommendations toward various clinical goals. CDS has been used in clinical settings since as early as the 1980s and eligibility determination has been a common usage. [6] Eligibility or inclusion criteria are decided by the study authors and translated into computed algorithms. Automating the algorithms can speed the evaluation and improve the accuracy of participant eligibility. Efficient eligibility determination can improve the quality and lower the costs of participant recruitment. At the most basic level Boolean eligibility requirements are defined and participants are evaluated based on true or false determinations for each criterion. If the answers to one or more of the criteria are false, then the participant will be ineligible. If all the criteria are met, then the participant will be eligible and can be enrolled in the study.

1.1.6. Randomization and Blinding

If a study involves treatments, participants will often be randomly divided into multiple, blinded study groups. Randomization ensures a known chance that each participant will be assigned to one or another study group and that the assignment will be unpredictable. Often this will be an equal chance between a control and one or more active interventions. Other times the chance will be designed to be unequal. In all cases the probability of assignment will be known but the actual assignment will be unpredictable.[11] To maintain the integrity of a study the investigator may utilize one of two popular approaches. Single-blind design ensures the subject is unaware of the study group assignment. Double-blinding ensures that neither participants nor researchers know the study group assignment.

Double-blinding ensures that study participants and those directly involved in the outcomes of the study or direct involvement with participants do not know the assignment of the study group to a particular intervention or control. Blinding techniques minimize bias that can influence additional clinical interventions or outcome evaluation. Investigators face internal and external pressures to influence the group assignment process. Participants may pressure an investigator to be included in the active study group over the control group. Investigators themselves may consciously or unconsciously impose bias for a variety of reasons, including the honest desire to improve the health outcomes of participants. Studies have shown the real, confounding effects of unblinded research. "Blinding is as important as randomization."[12] Randomization and blinding are entwined and integral to effective clinical research.

The randomization process needs to be protected so that assignments are effectively random and blinded. PDC requires strict manual controls. The “sealed envelope technique” is one common manual randomization control that involves random assignments that are created in advance and unsealed only when an eligible participant is enrolled in the study.[12] EDC allows machines to enforce randomization and blinding by automating the generation of randomization tables and the assignment of participants to blinded study groups.

Let’s briefly explore various study designs and techniques of randomization. The simplest type of interventional study will include a control group and an active group. Block randomization would ensure that an equal number of participants would be assigned to each group after a certain number of participants have been enrolled.[11] For example in a randomized controlled trial (RCT) with two study groups, for every 20 participants enrolled (the block size is 20), 10 would be assigned to group A, and 10 would be assigned to group B. A more advanced trial might have multiple active groups and a different proportional balance. In this case it may be desirable to control the percentages assigned to each study group (ie 30%, 30%, 40%). So, with a block size of 20, group A would have 6 participants, group B would have 6, and group C would have 8 participants.

Stratification is a technique to control confounding of the study objective. It involves “separating a sample into several subgroups according to specific criteria”[13] and ensures that a predictor of outcome can be evenly distributed among study groups.[12] For example if age were known to be associated with outcome, simple block randomization could result in an imbalance of age among study groups, with most elderly participants in one treatment group and younger participants in the other group. This could have a confounding effect on the study. Stratification would control for this and balance distribution by age group. Other confounding factors could be geography, sex, or pre-existing medical conditions.[11] Effectiveness must be protected with randomization techniques. Blocks and strata must be sized appropriately to ensure effective randomization. Too many strata may leave many blocks unfilled.[11, 12]

Other randomization techniques include cluster randomization, matched pairs, unequal group allocation, adaptive randomization, and pseudo-randomization.[11, 12] Randomization techniques can also be applied within the otherwise fixed form schedule protocol. Randomized CRF serial sampling will schedule one or more CRFs at a randomized point in the study timeline. A randomized participant subset will define a randomized sub-group of participants within the overall study cohort on whom to capture more extensive data. Documentation of the randomization process is essential for auditing the integrity of the process as well as certifying which group received which treatment. A well-designed study will ensure effective randomized assignment of participants to blinded study groups.

1.2. Thesis Statement

A mobile electronic data capture framework with support for complex study protocols that also functions with limited internet access will better meet the needs of clinical researchers working on distributed studies in resource-constrained geographical areas than other available EDC platforms. This thesis will explore the architecture of ConnEDCt, an EDC framework built by the author, and case studies of seven clinical research investigations that have used ConnEDCt as their primary EDC tool and provided valuable input toward the development of the framework. ConnEDCt supports complex features of longitudinal study protocols, requires minimal training to use, operates with limited internet connectivity, is regulatory-compliant, and operates on mobile devices as well as personal computers.

1.3. Contribution of the Thesis

The original contributions of this research work include the development of the ConnEDCt app, a framework and platform for mobile electronic data capture to support complex clinical studies in remote areas. ConnEDCt is built for clinical investigators to improve data quality of their research in challenging environments. ConnEDCt has a client component that runs on desktop and mobile platforms and a server component to centralize data. ConnEDCt has been used by five clinical investigations including cohort studies, cross-sectional surveys, and clinical trials for over 3,000 participants with over 13,000 participant encounters and over 55,000 completed case report forms. ConnEDCt is a full-featured mEDC solution that provides complex protocol support and easy extensibility for additional clinical research requirements. It allows investigators with limited assistance from programmers to implement case report forms, data validation, automated eligibility assessment, and scheduled events. ConnEDCt fills a gap in available EDC tools with support for complex protocols, and investigations in geographical areas with limited or no Internet availability. Multiple research teams have chosen ConnEDCt for data capture on challenging projects.

1.4. Organization of the Thesis

In this thesis, we provide a brief encapsulation of the author’s prior experience in clinical research; an overview of clinical study protocols; and the thesis statement in Chapter 1. Chapter 2 discusses the history of electronic data capture and mobile computing in clinical settings. Chapter 3 explores in detail ethical concerns around clinical research and government regulations to protect human research subjects. Chapter 4 lays out the detailed architecture of the ConnEDCt software platform including database schema, data lifecycle, functional components, implementation of new studies, system deployment, and data capture workflow. Chapter 5 examines five case studies of ConnEDCt implementations for clinical studies. Chapter 6 explores planned future enhancements and lessons learned, and Chapter 7 draws conclusions about the usefulness of ConnEDCt for clinical research.

2. EXISTING LITERATURE

2.1. Electronic Data Capture

Scientific and clinical studies mostly continue to rely on paper-based data capture (PDC) with EDC gaining more and more momentum. As computers became widely used for statistical analysis, double data entry of paper CRFs into spreadsheets or databases became a routine step in data capture. EDC is being adopted at a greater rate as the availability of computers, software, and the Internet keeps growing. When implemented well, EDC has some obvious benefits over PDC. Shantala, B., K. Binny, and M.S. Latha demonstrated that electronic CRFs eliminate redundant data entry, reduce data entry errors, improve data quality, and standardize data formats.[9] Despite the evident advantages, the perceived cost and limited technical capability often remain as barriers to adoption of EDC. This is especially true for small- to medium-scale studies without institutional or enterprise financial and technical support. As we will see from other prior research, the benefits of EDC in data quality and overall financial cost are real and justifiable.

Weber, et al., in a study on the effectiveness of web-based EDC[14], found that the time for data entry was reduced by an amazing 75% compared with the process of paper-based data entry, coding, and double entry. Meanwhile data entry errors were reduced by an astounding 100%. Even though fixed costs were higher for EDC, the variable costs of data entry were a quarter of that of PDC. The conclusion drawn was that the longer a study was conducted and the more CRFs were collected, the less the overall costs for EDC. The authors concluded that the main quantifiable benefits of EDC were reduced costs over time and lower error rate. Other tangible benefits mentioned were easier regulatory compliance and better control over data.

In a trial of various EDC methods against PDC by Walther et al. in a field site in Gambia, various factors including data quality, time, cost, and clinical staff acceptability were evaluated.[2] With no data validation the error rates were found to be similar to those of PDC. EDC was quicker than PDC, especially as the staff became more familiar with the tools. Startup cost was again higher with EDC. However, as the study progressed and time to complete CRFs was reduced, EDC became more cost-effective. At a certain point the labor cost of PDC, followed by secondary data entry into a database, and finally data verification made PDC less cost-effective than EDC. Furthermore, EDC was widely accepted by field staff. Walther et al. conclude that EDC can be “a more time effective, … cost effective method”, yet stress the importance of good study design and the risks of data loss when devices are used in remote areas.

Pawellek et al. found that being forced into a thorough study design process by the nature of using an EDC system was one of the main benefits. While their planning and design process was “more time-consuming” it was also a “great advantage” due to reduced error rates and higher quality data. The benefits of higher quality data, and therefore fewer data checks after the study completed, outweighed the longer implementation time. [15]

Pavlović et al. performed a sophisticated business process analysis of the cost effectiveness of EDC versus PDC. The authors recognized the extreme inefficiency of the double entry method used in PDC, yet acknowledged technical requirements and uncertain cost of EDC. In all five separate testing scenarios there were clear cost savings with EDC and the main benefit was reduced data entry error rates. Identified disadvantages of EDC included the need to adjust to different organizational culture and regulatory compliance frameworks, as well as extended costs such as IT infrastructure that may not already be in place. [16]

Two recent studies have reported similar advantages of EDC over PDC. [17] [18] Both observed lower total cost especially with longer-running studies and fewer errors when compared to PDC. Both studies also found data entry speed to be comparable between EDC and PDC.

Staziaki, et al. compared data entry into Excel spreadsheets to EDC. Spreadsheets were seen to be potentially less expensive while eliminating double entry. EDC won on speed and data quality lending evidence that enforcing a study protocol is another major advantage of EDC.

Researchers appreciate having greater control over the data. There are perceived benefits just by having data in hand and available. "Researcher-controlled data services and secure data collection, storage and export is a universal need for any … research study." [19] "Moreover, the large geographic areas associated with many research studies as well as the need to maintain the integrity of clinical trial data make [EDC] even more appealing."[14]

There is a large body of evidence showing that EDC is more effective than other data capture methods. Yet concerns over cost and technical demands persist. Often the specialized technical requirements such as setting up EDC software or lack of computing infrastructure are still barriers. Franklin et al. recognized that EDC solutions are difficult to set up and that this may limit small- and medium-scale studies without institutional support because many EDC solutions are designed for large-scale studies. Interestingly, Franklin et al. also concluded that research teams used to PDC are willing to work around limitations in EDC software because any "EDC software is a step up." [20]

Therefore more affordable EDC solutions designed for small and medium scale studies, may be a good fit for research teams with limited resources. Richesson and Andrews note that “while experimental expert-type systems have been developed with the idea of helping clinical investigators design their own trials, their scope is too limited to address the diverse issues that human experts handle.” [6]. In other words minimum viable functionality combined with flexibility is a better goal instead of the ability to handle every requirement imaginable.

Study designers should realistically expect to need technical personnel, to some extent, to setup and manage EDC systems. This makes sense and parallels information systems everywhere. Any technical machinery needs an engineer or technician to keep it running smoothly.

The benefits of EDC have been repeatedly demonstrated. However, it is challenging to develop an EDC system that is comprehensive enough to cover all clinical research use cases and at the same time simple enough to be fully managed by non-technical research staff. Therefore, neither a fully comprehensive feature set, nor complete elimination of programmer or technical staff involvement is realistic. A sweet spot to aim for is flexibility, intuitiveness, and extensibility.

2.2. Mobile Computing in Clinical Settings

Our further aim is an EDC system that is fully functional on mobile devices, especially tablet and smartphones. Data capture using laptops and desktop PCs is usually limited to indoor stationary use. As Weber et al. pointed out, “limitations for this method of data collection have been the size of the computer (making it difficult to use in some field situations)”.[14] Tablets and phones offer a better mobile form platform.

As early as a decade ago Morak et al. recognized that for mobile EDC, “the most challenging part is the user interface … and to synchronize.” [21] Despite implementing mEDC on now-obsolete phone hardware (e.g. a Nokia clamshell with no touchscreen), Morak et al. implemented mEDC with a custom client server architecture and concluded that their platform, although very limited in functionality, was “easy-to-use, intuitive and time-saving”.

More recently Meyer et al. focused on the need for off-line functionality with later synchronization when using mEDC for conducting trials with home visits. [3] Instead of building a full system from the ground up, the team extended an existing clinical data management system to include remote client and synchronization functionality. They implemented a flexible EDC data schema and recognized that synchronization of captured data was the “most complex part” of their implementation. Meyer et al. went into fine detail on their syncing algorithms, noting the complexity of supporting syncing with data dependencies, threaded concurrency, queuing requests, and conflict resolution.

Pakhare et al. in a survey of phone-based mEDC in Africa found that mobile devices have the ability to enrich data with built in sensors such as cameras, microphones, and GPS sensors.[22] The resulting photographs, video clips, audio clips, and location data go beyond the capabilities of PDC or laptop/desktop EDC. Pakhare et al. also noted the low training requirement for use of mobile devices now that we live in an era when they are commonly used. On the other hand they noted limitations not encountered with PDC, such as maintaining a battery charge in remote locations, theft, malfunction, and reliance on network connectivity.

Patel et al. found that mEDC on smartphones was key to their distributed, international study on smoking in vehicles. [23] Their previous study methods involved recruiting geographically dispersed observers using PDC to capture a body of data while observing occupants of moving vehicles. The team implemented mEDC with the goal of improving “time-consuming and fragmented manual methods”. The lead investigators created a data and workflow specification and then opened dialogs with commercial software developers to understand the development process and costs. In the end the team chose to work with student developers. Patel et al. discussed the challenging and lengthy process of development and highlighted the tangible and intangible costs of custom development. In the end the team completed a functional smartphone app that “may provide greater efficiency that traditional methods.” Not exactly a definitive conclusion, but the study noted the advantages seen with EDC in general: real-time access to data and improved data quality. [23]

Van Heerden et al. directly compared mEDC on smartphones with PDC and replicated results of earlier EDC studies to confirm that error rates for mEDC methods comparably low to traditional EDC. Furthermore, like in earlier EDC studies, the benefits of mEDC over PDC are “magnified as the size and complexity of the study increases.” [24] This magnification of benefit is found because despite the higher startup costs of mEDC, the lower cost per participant becomes significant with larger, more complex studies.

King et al. conducted a comparison of four mEDC software packages in order to replace their PDC methods in Malawi. [25] All fieldworkers preferred EDC to PDC for a variety of reasons. Among the notable subjective reasons to prefer electronic methods were: ease of carrying an electronic device over a stack of papers and the prestige in the community the electronic device conveyed. The research team also considered the fieldworkers' use of electronic devices to be a benefit in terms of the "opportunity of capacity building within communities." Expected downsides were failures and theft of the equipment. However, these downsides were minimal and considered reasonable in the greater context. The team reiterated the emphasis we have seen before on the need for preparation and advance planning. The preparation phase took longer with EDC, yet once the studies were being conducted they ran smoothly. They conclude that mEDC is "not only viable, but desirable."

As an important counterpoint, Vélez et al. implemented mEDC in rural Ghana and encountered resistance among local healthcare workers. [26] The main complaint was that the workflow design was disruptive to the clinicians. Instead of designing a system that matched and integrated with the clinical workflow, the mEDC system disrupted the established clinical path. This resulted in resistance by the clinical staff. Two mobile phone models were used, a hardware keyboard model and a touchscreen model. The touchscreen was shown to have a lower error rate and greater user acceptance. Despite initial optimism from the clinical staff hopeful that the system would reduce an already burdensome workload, the researchers ranked several "major usability problems” and one “usability catastrophe” in the design and implementation that became ultimately an unsustainable effort. The researchers conclude, “careful and thoughtful design is essential for successful implementation, scalability, and long-term sustainability” of an EDC project and emphasize the importance of early engagement with the clinical users of the system. In another case von Niederhäusern et al. encountered issues with a design that wasn’t rigid enough in enforcing the protocol for their mEDC system for clinical home visits. “[T]he mobile application was designed with the highest user flexibility and – usability in mind in order to minimize user fatigue or dropout. This flexibility in data entry resulted in data points which were often ambiguous (e.g. inconsistencies in automatic time stamp versus time point indicated by caregiver, or typing errors for sample codes).”[27] Full reconciliation of the data would have entailed rigorous crosschecking of the data and perhaps repeated home visits.

Several other mEDC systems have documented success in remote geographical areas with some familiar results:

- A diagnostic and mEDC system “can greatly improve the delivery of quality health care in remote locations of low- and middle-income countries. Quality, complete and timely data collection by health workers in a remote setting in Kenya is feasible." [28]

- "Despite challenges including prolonged setup times, the [mEDC system for a complex, four-armed, cluster-randomized, controlled trial in rural Nepal] met multiple data collection needs for users with varying levels of literacy and experience." [29]
- Because of the lack of existing research and technological infrastructure, "it may be possible for countries in Africa to implement leapfrog approaches that exploit the advantages of digital and mobile technologies, tools which have permeated and revolutionised many aspects of work and life around the world but have not yet been widely adopted for use in clinical trials." [30]
- In a randomized clinical trial conducted across 14 hospitals in China an mEDC system helped "doctors complete a phase IV pharmaceutical clinical trial and was feasible for management of this trial. Moreover, doctors expressed their willingness to use this tool for study implementation." [31] Special emphasis was given to high user satisfaction and protocol enforcement features that benefited staff who had no clinical trial experience.

With widespread acceptance and acknowledgement of the success of mEDC Eagleson et al. moved the discussion forward by focusing on ethics and the privacy and security issues related to electronic storage and transmission of protected health information (PHI). [32] The authors applied defensive strategies to the risk of external attacks as well as internal security layers to mitigate unnecessary access to PHI by authorized users.

We conclude that mEDC is a highly successful strategy for managing clinical studies when properly designed and integrated with the clinical workflow. Areas of special concern when implementing mEDC should be protocol development, usability, and security.

3. ETHICS & GOVERNING REGULATIONS

3.1. Ethics

Systematized clinical research and drug trials began seriously in the twentieth century with the development of penicillin and other medical interventions. During this period research on human subjects was often conducted on vulnerable populations in prisons, orphanages, homes for the mentally disturbed, and other institutionally controlled groups without the explicit consent of the individual participants. The misguided, but common, justification was that these marginalized groups could be used for medical research as long as there was a benefit to the greater society.[7]

The turning point for seriously considering ethics in research was the revelation of the experimentation Nazi doctors conducted on prisoners during World War II. The types of experiments conducted on prisoners without consent were horrific, including experiments with mustard gas, poisons, freezing temperatures, burns, high altitude, starvation, malaria and other contagious diseases, and deliberately inflicted wounds.[33] During the Nuremberg Trials following the war, the military tribunal developed a code of ethics by which to judge the accused. The Nuremberg Code established the "primacy of consent" and other protections for human research subjects.

The Nuremberg Code was limited in scope to judge defendants in the Nuremberg Trials, yet it led to further development of global ethical guidelines. In 1964 the World Medical Association published the first edition of the Declaration of Helsinki[34] and international consensus document and the basis for many future national guidelines. The Declaration of Helsinki established basic principles of ethical research on human subjects and importantly differentiated ethical treatment during clinical care from non-therapeutic research.

The Declaration of Helsinki was and remains an unenforced consensus document and no legal code for human subject research had yet been established. Meanwhile questionable research practices continued. In 1966 Dr. Henry Beecher pursued the issue in a landmark paper "Ethics and Clinical Research"[35] Dr. Beecher cited many contemporary examples of published studies conducted by well-respected research institutions where consent of human subjects was not requested, known effective treatments were withheld, or risky procedures were performed on healthy individuals. These studies resulted in serious side effects including death. In analyzing the frequency of ethical breaches Beecher noted the ease with which he identified questionable ethical practices. He noted that his preliminary evaluation identified 17 examples of likely ethical violations. These seventeen “easily increased to 50” that led to “186 further likely examples.” Beecher highlighted two primary components to an ethical approach to clinical research. The first was informed consent. “The statement that consent has been obtained has little meaning unless the subject or his guardian is capable of understanding what is to be undertaken and unless all hazards are made clear.” The second important factor he added is the “safeguard provided by the presence of an intelligent, informed, conscientious, compassionate, responsible investigator.” He found both of these factors lacking in the ethical climate at the time.

Beecher’s paper brought wide attention to an ethical crisis within the United States research community. Other revelations at the time such as of the hepatitis B studies at Willowbrook and the Tuskegee syphilis studies drew additional public scrutiny.[7] This widespread, public attention to ethics in research led to the formation of the governmental National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. This commission produced the Belmont Report in 1979. The Belmont Report[36] was the first of its kind in the United States, a groundbreaking, government-sponsored guidelines document for ethical principles for research involving human subjects. The report established boundaries between medical practice and research, basic ethical principles for research, and specific guidelines for informed consent, selection of subjects, and assessment of risk.

The Nuremberg Code, Declaration of Helsinki, and The Belmont Report together became the basis for regulations developed beginning in the 1980s that govern the protection of human research subjects internationally.

3.2. United States Regulations

3.2.1. Protection of Human Subjects

The United States regulations that govern the protection of human research subjects are referred to as “The Common Rule”. The Common Rule is codified in Title 45 of the Code of Federal Regulations (CFR) Part 46 [37], referred to as 45 CFR Part 46. This regulation provides rules for Institutional Review Boards (IRBs), informed consent, and special protections for vulnerable groups.

IRBs are governing bodies formed by the institution sponsoring the research to oversee the protocols and conduct of a study involving human subjects. IRB rules detail the role of the IRB in oversight, how to select the membership of an IRB, and when expedited review procedures may and may not be used.

Informed consent is strictly required when performing research on human subjects. “No investigator may involve a human being as a subject [without obtaining] legally effective informed consent".[37] Informed consent involves explaining to the research subject the following details at a minimum:

1. Purposes of the research,
2. Duration of participation,
3. Procedures to be followed,
4. Foreseeable risks,
5. Benefits to themselves or others,
6. Alternative treatments,
7. Extent of confidentiality of their records,
8. Compensation or treatment available for research-related injury,
9. Whom to contact for more information about the research,
10. Their rights as participants, and
11. Participants' right to withdraw from the research study at any time.

Additional requirements apply under special circumstances or increased risk. The informed consent procedures must be documented in the study protocol and approved by the IRB.

The Common Rule also codifies special protections for vulnerable groups such as pregnant women, fetuses, newborns, children, and prisoners. For example research involving pregnant women, fetuses, or newborns must minimize risk and hold the prospect of direct benefit to the fetus or newborn. Both the mother and the father must grant consent – as long as both are available. When children – defined as “persons who have not attained the legal age of consent” – are to be the subject of research, assent by the child – defined as “affirmative agreement to participate” – is required in addition to consent by the parent or guardian. When prisoners are to be involved in research, the majority of the IRB is to “have no association with the prison” and at least one member of the IRB must himself or herself be a prisoner.

The Common Rule was revised in 2017 to provide greater protections for subjects while streamlining the compliance process for researchers. [38] The revised rules allow consent forms to be more concise and to contain more easily understandable language. Consent forms should now be organized for clear understanding for participants above legal protection of the researcher. The new standard is “reasonableness” instead of “comprehensiveness”. IRBs will be allowed to spend less time overseeing low-risk studies in order to focus more on high-risk studies. Researchers will be allowed to request “broad consent” in order to use PHI or identifiable bio-specimens for future, undetermined research. Finally, the updated rules allow for single-IRB oversight for multi-institutional studies. These new rules are intended to reduce the compliance burden placed upon researchers while maintaining protections for human subjects.

Other parts of the Code of Federal Regulations expand rules for informed consent, additional safeguards, IRBs, new drugs, device exemptions, and research conducted outside the United States. [39-44] Table 1 details these parts of the CFR.

U.S. Regulation	Covered Topics
Title 45 CFR Part 46	IRBs, informed consent, additional protections for vulnerable groups
Title 21 CFR Part 50	Informed consent, additional safeguards
Title 21 CFR Part 56	IRBs
Title 21 CFR Part 312	Investigational new drugs
Title 21 CFR Part 812	Investigational device exemptions
Title 22 CFR Part 225	U.S. funded research conducted outside the U.S.

Table 1. U.S. regulations for protection of human research subjects

3.2.2. *Electronic Records and Electronic Signatures*

The U.S. Code of Federal Regulations Title 21 Part 11[45] known as 21 CFR Part 11, or just Part 11, states that electronic records that meet specific requirements may be used as official records in lieu of paper records. The U.S.

Department of Health and Human Services has stated that the scope of records that are covered by 21 CRF Part 11 is to be interpreted narrowly. [46] Within that narrow interpretation the two categories of information include electronic signatures, when they are used in lieu of physical signatures, and Protected Health Information (PHI). The Health Information Portability and Accountability Act (HIPAA), 21 CFR Parts 160-164, further governs the definition and management of PHI.

U.S. Regulation	Covered Topics
Title 21 CFR Part 11	Electronic records and signatures
Title 45 CFR Part 160	HIPAA General Administrative Requirements
Title 45 CFR Part 162	HIPAA Administrative Requirements
Title 45 CFR Part 164	HIPAA Security and privacy of PHI

Table 2. U.S. regulations for electronic records and healthcare data

Electronic signatures must be based on unique biometrics, or on two identification components such as a user account name and password. Electronic signatures must be linked to their respective electronic records in a way that they cannot be transferred or copied to another electronic record by "ordinary means." Controls must be in place to de-authorize lost or stolen identification tokens, as well as to detect attempts of unauthorized use of identification tokens.

Research data when combined or crossed with clinical treatment will be governed by HIPAA regulations that have privacy and security requirements. Complete coverage of HIPAA regulations is out of the scope of this work. However, 45 CFR Part 164 [47] creates specific requirements for data management, encryption, privacy, and disclosure when breaches occur.

3.3. International Regulations

Research that is conducted outside the United States that is funded by the U.S. Federal Government agencies is regulated by 22 CFR Part 225 [48] in addition to regulations in the local jurisdictions. Many countries have developed their own regulations for protection of human subjects in research. Many are based on United States regulations.

4. CONNEDCT ARCHITECTURE

4.1. Architectural Goals

The primary goals of the ConnEDCt architecture are mobility, flexible support for complex study protocols, and reuse. Some features like data security and privacy are required for any EDC system as per regulations discussed in chapter 3. Beyond these core requirements, ConnEDCt focuses on mobility and support for complex protocols as differentiators from other EDC platforms. ConnEDCt provides sophisticated protocol support features; intuitive user-interfaces with low training requirements; offline operation; and data syncing.

ConnEDCt is built with a mobile-first philosophy with a focus on tablet data-entry. Off-line operation and a touch-based, mobile user interface (UI) are fundamental to the design and architecture. Although ConnEDCt is also suitable for data capture on a desktop or laptop PC, the user interface is optimized for data capture on mobile devices. iPads are the primary devices of the intended audience. ConnEDCt is fully functional off-line with no required connection to a central data repository (CDR) except when syncing or receiving app updates. Local file storage allows for off-line operation and faster program responsiveness, since, unlike with web-based apps, the users never wait for data to be sent or retrieved during data entry.

One of the main challenges in supporting clinical studies has been creating a re-usable system that can be implemented for the most part by the research team without help from software programmers. Like Harris, “we realized early in the project that the critical factor for success would lie in creating a simple workflow methodology allowing research teams to autonomously develop study-related metadata in an efficient manner.”[19] Although a programmer is still required for implementation of a new study, a concise procedure makes programmer involvement as minimal and efficient as possible. Future development goals that will be discussed in Chapter 6 include further reduction of programmer effort in the implementation of new studies. Most likely some programmer effort will always be required to implement novel features or assist with complex algorithms.

Richesson & Andrews noted that “special problems still arise that only custom software development can solve.”[6] So far we have implemented several independent clinical studies with ConnEDCt. These will be described in detail in Chapter 5. Although developer support was required in all cases, in three out of six studies we implemented the study protocol designs with the standard feature set, without custom programming. In the other cases custom programming was needed to support additional complexities of the study protocols. In most cases novel but important features were then adopted into the platform. Many features such as dynamically generated CRFs, randomized groups, and an agenda view were added when a study required them and were implemented as platform features. Each implementation has been built on the previous experience so that most new features were integrated into the platform for re-use with future studies.

4.2. System Components

ConnEDCt consists of a distributed database built with the Claris FileMaker platform[49] plus a couple of 3rd party modules that are integrated for discrete functions. FileMaker is a commercial app-development platform by Claris International Inc., a subsidiary of Apple Inc. FileMaker combines a rapid development environment with native compatibility for personal computers (PCs) and mobile computing devices. The FileMaker platform includes a proprietary database engine, an integrated development environment (IDE), a server component, native clients for Windows and MacOS PCs, and a native mobile client for iOS. This full suite of built-in capabilities provides advantages for rapid development, and cross-platform deployment.

4.3. Database Schema

Figure 4.1 shows ConnEDCt's entity relationship diagram (ERD) and the structure of the database. The three entities FormType, VisitType, and FormSchedule establish the basic study schema. The FormType entity represents CRFs; the VisitType entity represents defined participant encounters; and the FormSchedule join entity represents the schedule of CRFs used in each encounter. Other aspects of the study protocol such as eligibility criteria and coded list values are represented by respective database entities.

For each study implementation CRF variables are hard-coded in new "CRF_" database tables that correspond to the CRFs defined in the Form table. An arguably better schema design would be to abstract the CRF variables in a Variable table so the CRF definition could be accomplished by an end-user. There were various reasons for taking the direction of the current database model, including limitations of time, resources, and capabilities of FileMaker. An alternate database schema that allows for user-defined schema is discussed in Chapter 6, Future Enhancements.

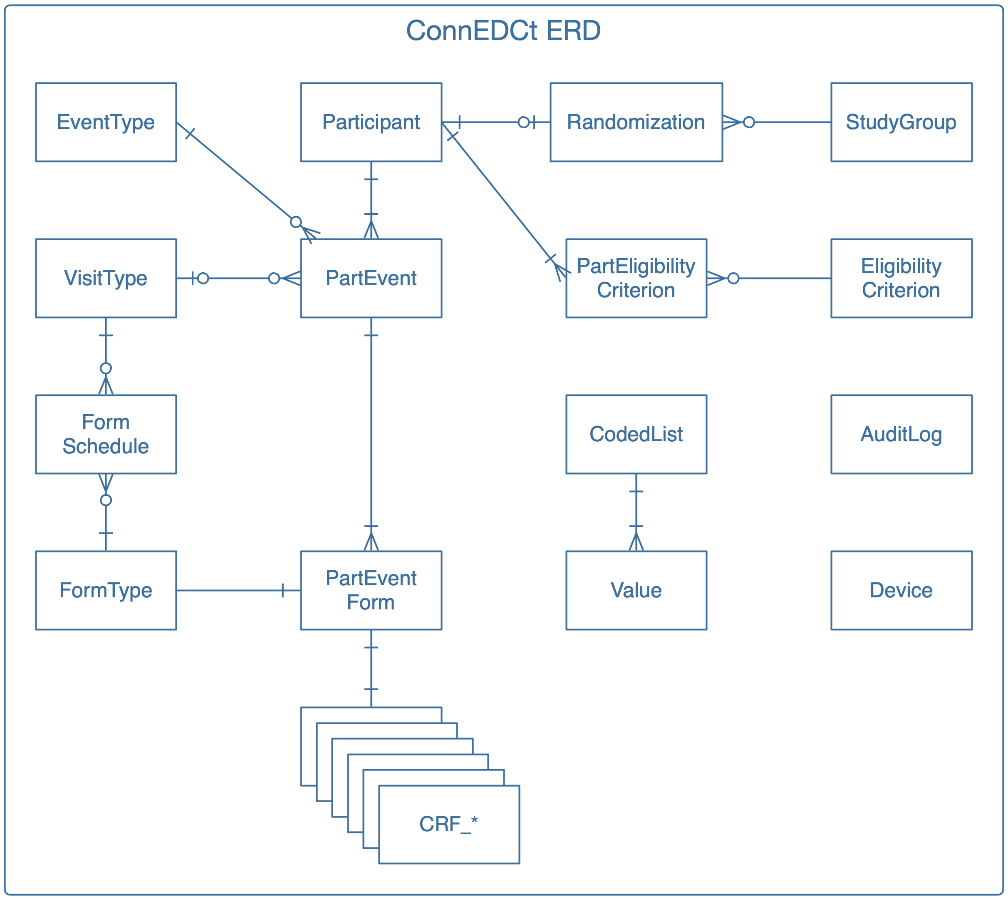


Figure 4.1. Conceptual entity relationship diagram (ERD) of ConnEDCt

Aspects of features such as device registration, eligibility, scheduling, e-signatures, and randomization are supported in the database schema with relationships and attribute values. Appendix A shows table attributes for the entities discussed below.

The Device table entity stores records for every device that was used to sign into a ConnEDCt study. Information stored for each device includes the persistent hardware ID for the device, a user-assigned device name, the study ID details discussed in section 4.5.4, as well as device attributes and hardware capabilities.

Eligibility criteria formulae are stored in the EligibilityCriterion table with relationships to the specific visit and form that the formulae references. Instances for participant eligibility status are stored in the PartEligibilityCriterion table with a relationship to the EligibilityCriterion table.

Participant encounters and assignments of CRFs to encounters are defined in the VisitType and FormSchedule tables. Visit attributes include the ordering of visits, schedule in days from start of the study, whether to automatically create the next visit when scheduling–or stop the scheduling procedure, and whether eligibility is required before creating the visit in the schedule. Randomization attributes that affect CRF scheduling – discussed in section 4.5.7 – are also stored in the FormSchedule table.

The PartEventForm table manages metadata associated with the CRF including completion and sync statuses, e-signatures, and relationships to the participant encounter and form record the CRF is associated with.

Finally, study randomization data is stored in the StudyGroup and Randomization tables. The StudyGroup table has strict access controls to enforce blinding from investigators of the description attribute. The study group code is assigned when a randomization record is assigned to a participant. The randomization record maintains the order of the list, the study group, and whether a record has been assigned. The randomization record also contains the visit number to apply the serial sampled CRFs – see section 4.5.7. Because randomization may be centralized or federated, the randomization record also stored the device ID that has authority to assign it.

The database schema manages all the persistent information on the study schema design. Whenever special requirements arise for a new study protocol feature the priority is to adopt the change in the study schema to allow for framework adoption.

4.4. Data Lifecycle and Data Quality

4.4.1. Data Security and Privacy

HIPAA regulations require electronic systems to "ensure the confidentiality, integrity, and availability" of PHI.[47] These terms have specific definitions within the code and explicitly require encryption and audit logging, while implicitly requiring business continuity and contingency plans. Encryption in ConnEDCt takes two forms: file encryption, also known as encryption at rest; and encryption in transit, in other words encrypting the online data traffic between the CDR and the client devices. ConnEDCt stores the database files on the CDR and on client devices in order to enable offline functionality. These files are encrypted with the AES-256 encryption algorithm using native FileMaker file encryption features.[50] Data in transit between the ConnEDCt CDR and client devices is encrypted using transport layer security (TLS) and third-party SSL certificates applying the SHA-256 algorithm with RSA (Rivest-Shamir-Adleman) encryption.

User account access is controlled with usernames and passwords with role-based access control. Research Associates collecting data in the field have different access requirements than statisticians analyzing data. RAs typically have access to capture a full dataset – or segmented data, if they focus only on particular subsets of study data – whereas statisticians are prohibited from access to identity data.[6] Data Managers have limited ability to make corrections to captured data.

4.4.2. Syncing

Two of the great benefits of EDC are real time data analysis and data security. A system that operates off-line must store data locally. However, local data storage leaves the data isolated and vulnerable to loss. Full data security therefore requires that data must be transmitted to a CDR. There are many strategies to accomplish this, but syncing is most efficient since it manages and tracks changes only. Syncing is a complicated programmatic problem involving comparison of changes, time zone corrections, conflict resolution, error correction among other challenges. ConnEDCt integrates a commercial syncing engine in order to achieve the best possible results. We tested several commercial solutions before settling on 360Works MirrorSync[51]. MirrorSync provides relatively simple integration with our FileMaker software platform and reliable operation over even poor network conditions. In addition 360Works has released several MirrorSync upgrades over the last few years that have improved performance.

We use a "most recent change wins" configuration and syncing is as simple as tapping a button in the interface. Sync time depends on the amount of data changes to be uploaded and downloaded.

4.4.3. Audit Logging

Audit logging is a feature that is built into many relational database management systems (RDBMS), but not FileMaker. Audit Logging is another challenging feature to be implemented from scratch with many nuances to be considered. Therefore we implemented audit logging with an add-on module, AuditLogPro 2.0[52] by 1-More-Thing. Changes are collected during data entry and saved to a separate log table with triggers. Syncing is configured to sync the log in only one direction – from devices to the CDR, so that it is not sent from the CDR to devices. Aside from maintaining data security by not storing the full log on every device, the log also becomes quite large, so syncing in one direction is a more efficient use of the network.

4.4.4. Form Validation

Form validation is performed as a pre-processing step prior to signing a CRF. Validation is accomplished in one of two ways. Specific validation checks can be programmed for each variable in each CRF, or by default every variable shown on the CRF can be checked to verify it has a non-empty value. These validation rules are a part of the custom programming step of an implementation. Validation rules should be documented in the study design stage.

4.4.5. Form Signing

Electronic signatures ensure the authenticity of the captured data.[45] An RA must verify their identity by account name and password when they sign each form after having completed it. These electronic signatures capture the identity of the RA, the timestamp, and GPS coordinates (if available on the device). In addition a signed form is locked from future data entry, although a data manager can override this. Signed CRFs are marked as complete. When all forms in a visit are signed and completed, the participant visit is marked as complete.

4.4.6. Data Export

Exported data is used for deeper statistical analysis than can be performed in a typical RDBMS. Two types of data export can be run on demand from ConnEDCt: a full, de-identified dataset, and a separate export of participant identity records. Both exports result in a collection of Microsoft Excel files. The de-identified dataset includes one Excel file for each CRF that identifies the participant only by study ID and also includes metadata for visit number, date, etc. This export also includes a data file of coded lists. After experimenting with different file formats, Excel has turned out to be the most compatible option.

4.5. Functional Components

4.5.1. Case Report Forms

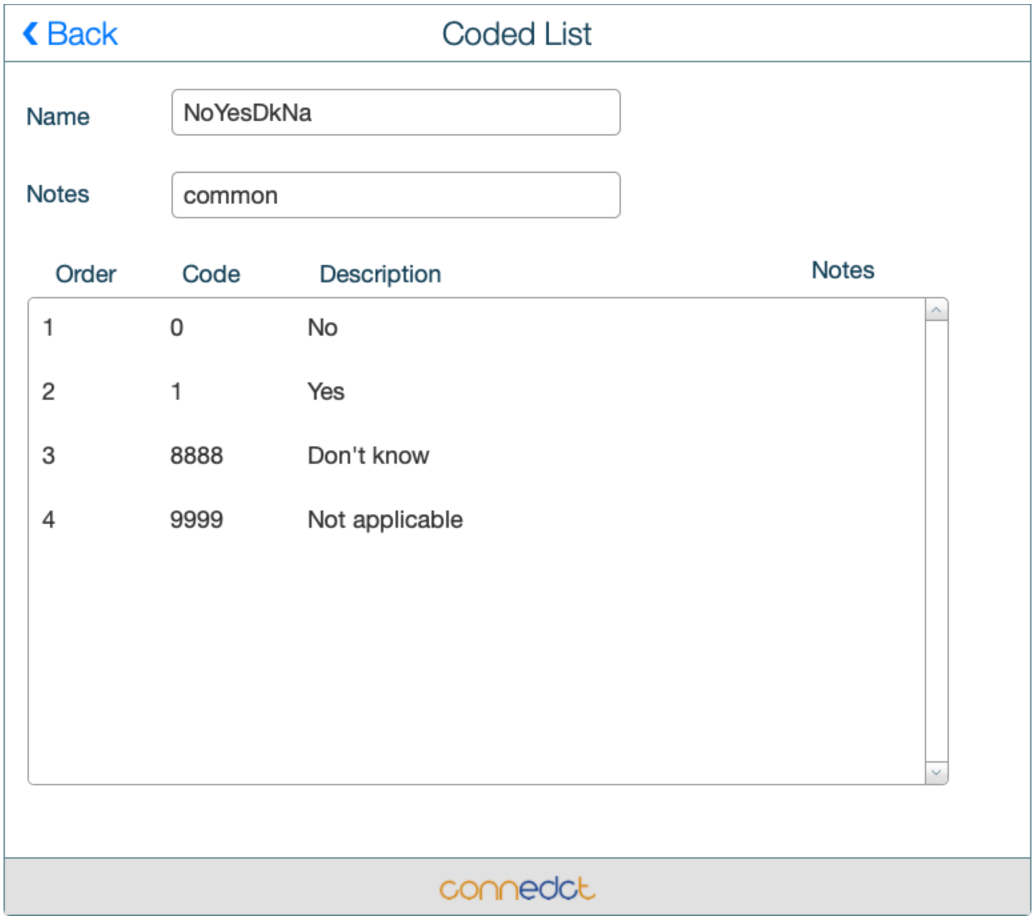
Case report forms are implemented using the FileMaker IDE. One table represents each CRF with a series of standard fields, such as database keys, and fields for data entry. It's worth mentioning that, in this intersection of disciplines, the terms field, variable, and column all represent the same concept of a structure for storing one piece of data.

FileMaker supports data types of text, number, date, time, timestamp, and container for storage of binary objects. FileMaker's "Layout Mode" environment is used for designing the data entry form. FileMaker-native field entry modalities include: free text, numeric, pop-up menu, date picker, and radio buttons. Other entry modalities can be provided with custom programming. Pop-up menus and radio buttons support coded lists so that the end user may see natural language choices, while the database stores numeric codes that enforce consistency and better data analysis.

4.5.2. Coded List Management

Coded values are essential for facilitating statistical analysis as well as optimizing data management.[19] A coded list item is represented by a text description with meaning in natural language and a stored code – usually numeric – that is more useful for data analysis. Having a code married to a natural language description also allows the description to be revised – fixing typos, adding precision language and translations – without affecting the coded value.

Coded lists can be used with pop-up menus, radio buttons, and other interface modalities that automate data entry when text description are useful to the users, yet the database should store a coded value. The end user chooses the text label value (e.g. "No" or "Yes"), whereas the database stores the coded value (e.g. 0, or 1). Radio buttons are useful for shorter lists; likewise, longer lists of coded values can be presented as pop-up menus. Multiple-choice options can be represented as checkboxes. Coded lists are exposed to data managers for customization as shown in Figure 4.2 and applied to fields in the FileMaker IDE.



Name: NoYesDkNa

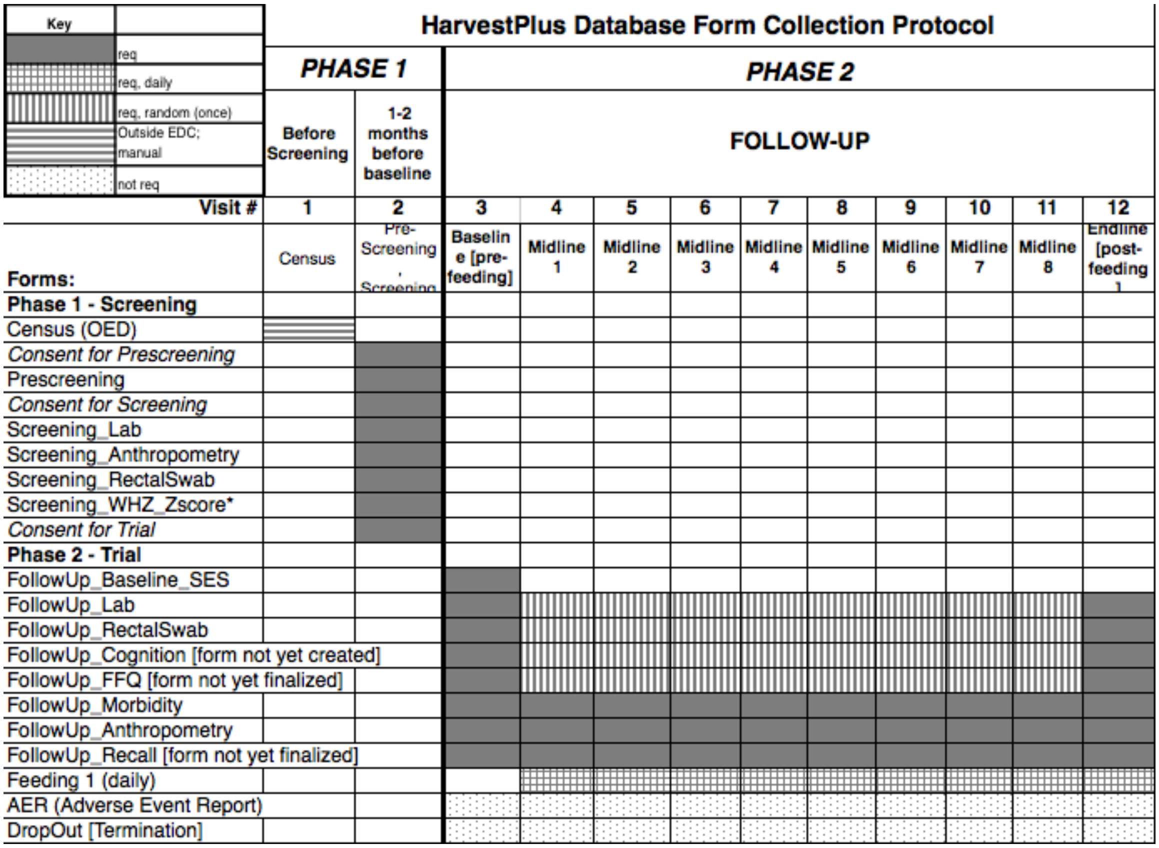
Notes: common

Order	Code	Description	Notes
1	0	No	
2	1	Yes	
3	8888	Don't know	
4	9999	Not applicable	

Figure 4.2. Defining a coded list

4.5.3. CRF Scheduling

The CRF schedule allows visits to be scheduled in advance and appropriate CRFs to be presented to RAs for data capture. It is one of the most basic elements of support for the study schema. Figure 4.3 shows the form collection protocol design document for the Biofortified Pearl Millet Trial[53], a clinical trial that employed ConnEDCt as its data capture solution. CRFs are shown down the vertical axis and participant visits are defined across the horizontal axis. The key indicates that some CRFs depend upon predicate conditions. ConnEDCt supports the conditionality of CRFs based on eligibility, randomized serial sampling, randomized subsets, and the values of specific variables in other CRFs. Based on these conditions CRFs may or may not be created for a visit.

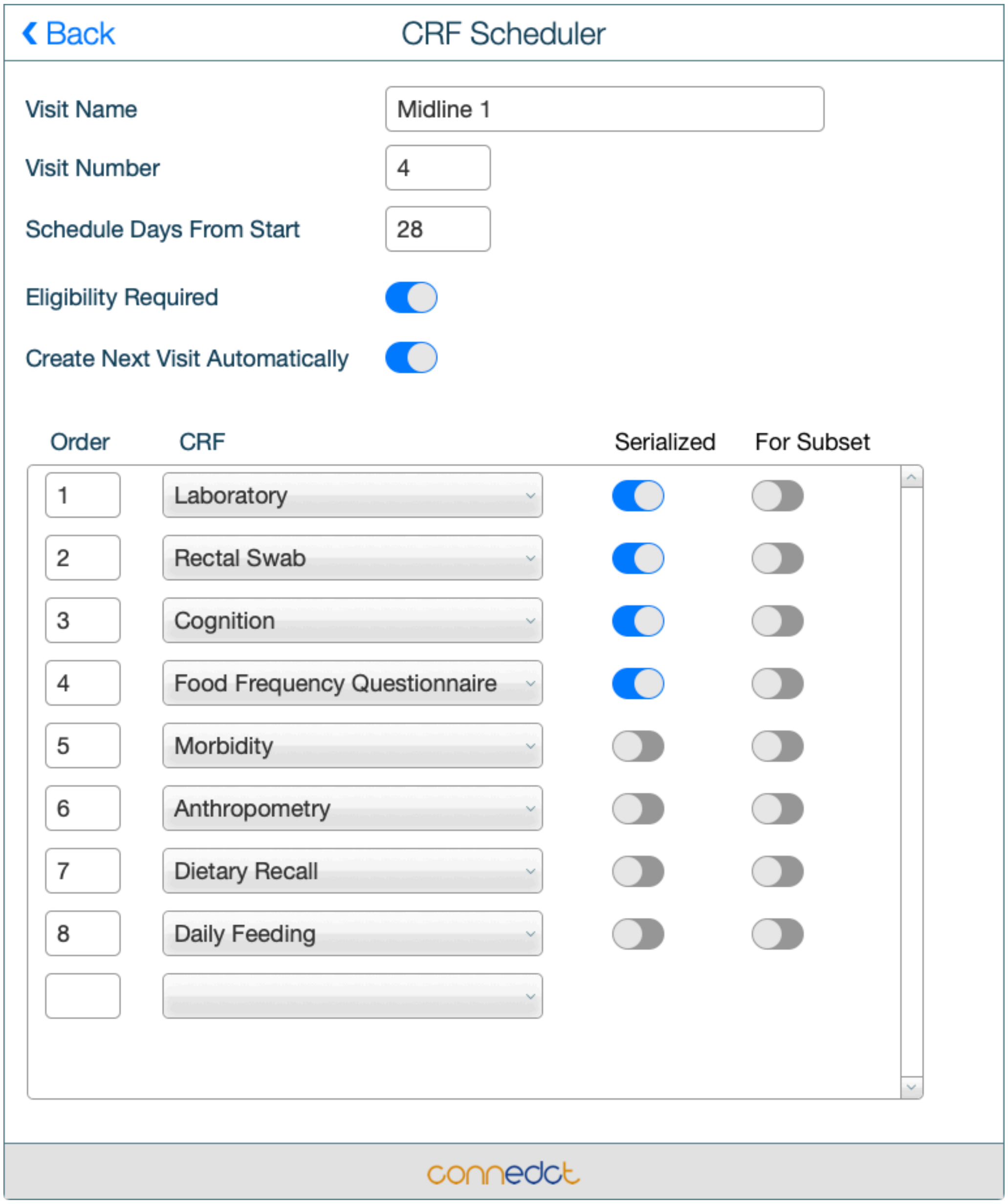


Key	
■	req
▦	req, daily
▥	req, random (once)
▤	Outside EDC; manual
░	not req

HarvestPlus Database Form Collection Protocol												
	PHASE 1		PHASE 2									
	Before Screening	1-2 months before baseline	FOLLOW-UP									
Visit #	1	2	3	4	5	6	7	8	9	10	11	12
Forms:	Census	Pre-Screening, Screening	Baseline [pre-feeding]	Midline 1	Midline 2	Midline 3	Midline 4	Midline 5	Midline 6	Midline 7	Midline 8	Endline [post-feeding]
Phase 1 - Screening												
Census (OED)	▤											
Consent for Prescreening		■										
Prescreening		■										
Consent for Screening		■										
Screening_Lab		■										
Screening_Anthropometry		■										
Screening_RectalSwab		■										
Screening_WHZ_Zscore*		■										
Consent for Trial		■										
Phase 2 - Trial												
FollowUp_Baseline_SES			■									
FollowUp_Lab			■	▥	▥	▥	▥	▥	▥	▥	▥	■
FollowUp_RectalSwab			■	▥	▥	▥	▥	▥	▥	▥	▥	■
FollowUp_Cognition [form not yet created]			■	▥	▥	▥	▥	▥	▥	▥	▥	■
FollowUp_FFQ [form not yet finalized]			■	▥	▥	▥	▥	▥	▥	▥	▥	■
FollowUp_Morbidity			■	■	■	■	■	■	■	■	■	■
FollowUp_Anthropometry			■	■	■	■	■	■	■	■	■	■
FollowUp_Recall [form not yet finalized]			■	■	■	■	■	■	■	■	■	■
Feeding 1 (daily)				▦	▦	▦	▦	▦	▦	▦	▦	▦
AER (Adverse Event Report)			░	░	░	░	░	░	░	░	░	░
DropOut [Termination]			░	░	░	░	░	░	░	░	░	░

Figure 4.3. Form collection protocol design document for the Biofortified Pearl Millet Trial[53]

ConnEDCt provides an interface for study designers to translate the study schema design into the ConnEDCt study schema. CRFs are first defined by name in the form list. Figure 4.4 shows the CRF Scheduler interface for the Midline 1 visit of the Biofortified Pearl Millet Trial. This interface includes details for scheduling as well as inclusion in the randomized CRF serial sampling and randomized participant subset protocols.



Visit Name	Midline 1
Visit Number	4
Schedule Days From Start	28
Eligibility Required	●
Create Next Visit Automatically	●

Order	CRF	Serialized	For Subset
1	Laboratory	●	○
2	Rectal Swab	●	○
3	Cognition	●	○
4	Food Frequency Questionnaire	●	○
5	Morbidity	○	○
6	Anthropometry	○	○
7	Dietary Recall	○	○
8	Daily Feeding	○	○

Figure 4.4. CRF scheduling interface

4.5.4. Study IDs

Working offline with multiple devices poses a challenge for serializing study IDs. It is given that every participant must be assigned a unique study ID for study integrity, as well as working with de-identified data. These IDs must be assigned upon the first interview with a participant and must be unique. Since participant records are created on devices that may not be connected to the CDR, the devices must be able to assign unique IDs despite not having this centralized coordination. Universally unique identifiers (UUIDs), while suitable for database primary keys, are too cumbersome to be practical as study IDs. Therefore a federated model of ID assignment must be used so that each device can assign a unique ID independently. The study ID is constructed using an identifier for the device followed by a participant number from a list maintained independently for each device.

To begin, every individual data capture device used for the study is registered with a unique, serialized identifier (e.g. 1, 2, 3, etc.). This involves capturing a persistent device ID, a unique hardware code that identifies the device, and associating it with a record in the Device table. Each record in the Device table matches a device registered with the study. The Device table stores a serialized NextParticipantNumber field for each device. When a new participant is created, the study ID is constructed from the serialized device number and the next participant number on that device. Then the value in NextParticipantNumber is incremented. For the first participant created on the first registered device, the study ID may be “1.1001”. The number of digits in the participant number segment is determined by the estimated enrollment in the study. Other segments may be added to differentiate multiple studies by a PI or institution and to indicate the study group or trial arm. An example of a full study ID would be “PM.01.1001-B” using the formula “[study code].[device number].[participant number]-[study group]”.

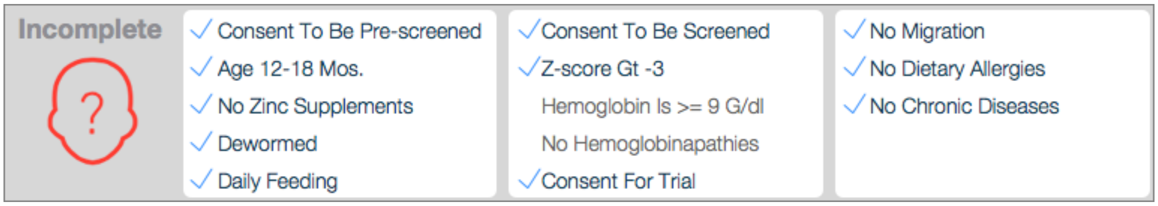
4.5.5. Consent Forms

Consent forms are implemented like any CRF plus including details for recording the method of consent plus optional capture of hand-written signatures. However, because participants are not authenticated users of ConnEDCt, their electronic signatures are not considered valid.[45] Therefore the consent record is recorded separately. Chapter 5 explores the various methods used for maintaining records of informed consent, including video capture of the informed consent interview, and paper-based capture of signatures or thumbprints. In all cases the status of consent is recorded for evaluating participation and eligibility.

4.5.6. Eligibility Controller

Automating eligibility as another protocol support feature lifts the burden of managing eligibility from the RAs and makes the interview and recruitment process easier. With PDC eligibility requirements were often consolidated on a single eligibility CRF. However, consolidation of unrelated questions could create an awkward workflow, for example mixing physician exam details with anthropometry, or lab results. This information is captured by different clinical staff or, in the case of lab results, take time to process. Automating the eligibility process, on the other hand, allows eligibility questions to be placed on relevant CRFs where they belong in the workflow.

The eligibility algorithm is defined in the EligibilityCriterion table as a series of formulae referencing one or more variables in a CRF. Any CRF, and the study as a whole, can have unlimited eligibility formulae. When a participant is created, records are instantiated in the ParticipantEligibilityCriterion join table for each EligibilityCriterion record. The eligibility controller runs each time an RA signs a CRF. At that moment any eligibility criteria referencing that CRF are evaluated and ConnEDCt updates the participant's eligibility status. Figures 4.4-4.6 show an eligibility algorithm and the three statuses: incomplete, ineligible, and eligible. Eligibility status begins and remains incomplete until either one criterion is negative (i.e. ineligible), or all criteria are positive (i.e. eligible).



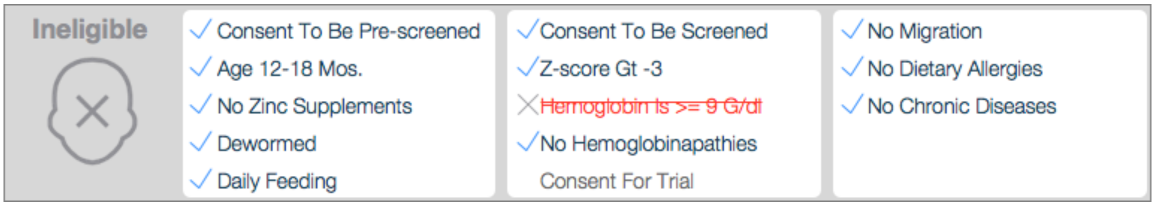
Incomplete

✓ Consent To Be Pre-screened
✓ Age 12-18 Mos.
✓ No Zinc Supplements
✓ Dewormed
✓ Daily Feeding

✓ Consent To Be Screened
✓ Z-score Gt -3
Hemoglobin Is >= 9 G/dl
No Hemoglobinapathies
✓ Consent For Trial

✓ No Migration
✓ No Dietary Allergies
✓ No Chronic Diseases

Figure 4.4. Incomplete eligibility status



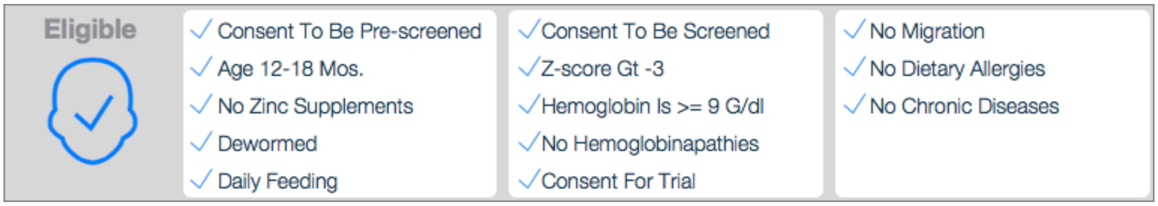
Ineligible

✓ Consent To Be Pre-screened
✓ Age 12-18 Mos.
✓ No Zinc Supplements
✓ Dewormed
✓ Daily Feeding

✓ Consent To Be Screened
✓ Z-score Gt -3
✗ ~~Hemoglobin Is >= 9 G/dl~~
✓ No Hemoglobinapathies
Consent For Trial

✓ No Migration
✓ No Dietary Allergies
✓ No Chronic Diseases

Figure 4.5. Ineligible status



Eligible

✓ Consent To Be Pre-screened
✓ Age 12-18 Mos.
✓ No Zinc Supplements
✓ Dewormed
✓ Daily Feeding

✓ Consent To Be Screened
✓ Z-score Gt -3
✓ Hemoglobin Is >= 9 G/dl
✓ No Hemoglobinapathies
✓ Consent For Trial

✓ No Migration
✓ No Dietary Allergies
✓ No Chronic Diseases

Figure 4.6. Eligible status

4.5.7. Randomization Controller

The randomization controller handles three randomization functions: assignment to a study group, CRF serial sampling, and assignment to a participant subset. Randomization can be centralized or federated depending on protocol requirements and workflow considerations. ConnEDCt can import a randomization table that includes a study group code, a serialized CRF value, and a CRF subset flag. The StudyGroup table also stores the meaning behind the study group codes. The StudyGroup table is confidential and access is controlled in order to maintain blinding.

After eligibility is evaluated and the participant is enrolled in the study, they are assigned a randomization key from the randomization table and the randomized study group code is appended to the study ID. The key to the randomization record is added to the participant record for referencing other randomized parameters. The randomization record is marked as used and cannot be assigned again.

In addition to the study group ConnEDCt supports randomized CRF serial sampling and randomized participant subsets. The form collection protocol in Figure 4.3 shows the use of randomized CRF serial sampling. The key shows a pattern indicating “req. random (once)”. This key is assigned to four CRFs over eight visits. The result is that each of these CRFs will be assigned on one of the eight visits depending on the randomized value in the SerialSampleNumber attribute of the Randomization record assigned to that participant. The biological samples will be collected only once during the midline of the study, determined by the randomized serial number value.

The workflow may demand that the study group is known immediately upon participant eligibility, if, for example, a treatment is administered at this visit. In this case a subset of the complete randomization table – maintaining block cohesion – can be loaded onto each device and randomization can be performed with decentralization of the system. An estimate must be made of the maximum recruitment numbers and the randomization series is divided among the devices that will be used for recruitment. Since each device will not recruit an equal number of participants a margin of error of extra randomization records must be added to each device.

4.6. Study Protocol Implementation

The implementation in ConnEDCt of each new study protocol is a collaborative effort between the research teams and ConnEDCt developers. Each implementation has provided the opportunity for more features to be added to the platform and for existing features to be re-used. Developer effort decreases with each implementation.

Two design documents are key to define the study protocol and standardize it for implementation. The form collection protocol design document (figure 4.3) identifies broad scope of a study including the CRFs and the Visits.

The data dictionary document provides the granular details of each CRF including variable names, data types, variable labels, descriptive text and coded lists for radio buttons, checkboxes, and drop-down menus. A well-defined data dictionary facilitates efficient communication among team members and reduces ambiguity when implementing data structures. Together these two documents cover most of the specifications. However, implementation is always an iterative process that involves revisions to these documents as well to ConnEDCt before participant recruitment begins. Invariably some refinements continue even after recruitment begins.

The Form Collection Protocol and Data Dictionary combined with an integration procedure allow fast implementation of the study schema in ConnEDCt. The implementation procedure follows these steps:

1. Create a new blank instance of the ConnEDCt system.
2. Create records for CRFs and participant encounters.
3. Create database tables for each CRF.
4. Create coded lists or import from the data dictionary.
5. Write validation scripts.
6. Add the eligibility criteria.
7. Create the file encryption key and user accounts.

Once the implementation steps are complete, the system is ready for testing by the research team to verify workflow and all other details. Notes are taken and revisions made until the system is approved for deployment.

4.7. Deployment and Updates

Deployment happens in two stages. First the ConnEDCt files are loaded on a server and then they are deployed to client devices. The server is running the FileMaker Server and MirrorSync services. This can be on any server with reliable high-performing internet service including on premise servers or cloud servers. Several commercial hosting providers offer this specific FileMaker and MirrorSync configuration or it can be built to spec.

Once the files are hosted the sync configuration is implemented in the MirrorSync console. Once sync is implemented the hosted files and client files are keyed to one another. MirrorSync provides a download link for client-side deployment. Since ConnEDCt runs within the FileMaker Go client application on iOS and FileMaker Pro on Mac and Windows, client devices must have one of these apps installed. Then deployment consists of installing an empty audit log file and then clicking the MirrorSync download link for the main ConnEDCt file.

Many revisions to the study protocol – such as changes to protocol schedule and eligibility criteria – can simply be made on the hosted version of the file. They will be delivered to the clients upon the next sync. Changes to database schema-related features such as variables and validation will require redeploying a new ConnEDCt file. This is a two-step procedure of syncing data to the server and then downloading a new ConnEDCt file with the MirrorSync download link that will replace the old file. The audit log file should never change during the course of a study so should only ever need to be deployed once.

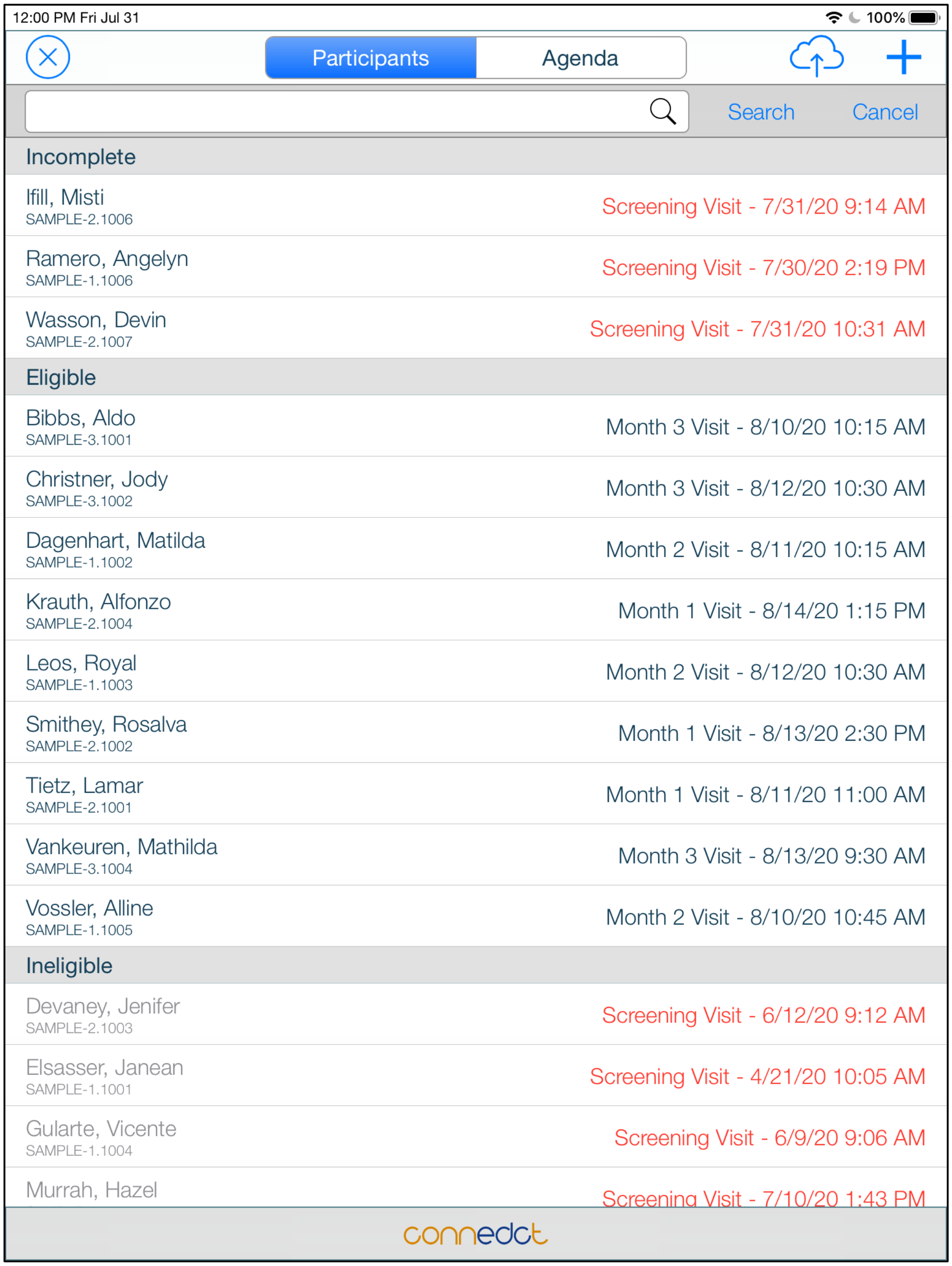
4.8. Data Capture Workflow

ConnEDCt's data capture workflow follows some common clinical interview patterns including informed consent interview, eligibility assessment, and one or more clinical encounters. We have made assumptions that many patterns are common to the majority of clinical studies. Elements including informed consent, CRFs, participant encounters, and eligibility criteria exist in most, if not all, clinical studies. Some patterns such as randomized study groups are specific only to particular study designs. The data capture workflow is designed around these assumptions.

The typical interview session begins with the RA opening the ConnEDCt app on an iPad. After receiving the encryption key and user authentication, ConnEDCt opens to a list of study participants as shown in Figure 4.7. The

Participant List interface features navigation buttons to switch between the participant list and the agenda view; buttons for ‘new participant,’ ‘sync to cloud,’ and ‘exit’; and a field to search for participants by name or study ID. The list of participants is grouped by eligibility status.

When creating a new participant, minimal information is captured, usually just the participant’s name, and, if a minor, the guardian’s name. ConnEDCt then generates a study ID; the initial visit or visits, depending on the study schema; and the CRFs associated with those visits. The new participant workflow is shown in Figure 4.8.



Incomplete

Ifill, Misti SAMPLE-2.1006	Screening Visit - 7/31/20 9:14 AM
Ramero, Angelyn SAMPLE-1.1006	Screening Visit - 7/30/20 2:19 PM
Wasson, Devin SAMPLE-2.1007	Screening Visit - 7/31/20 10:31 AM

Eligible

Bibbs, Aldo SAMPLE-3.1001	Month 3 Visit - 8/10/20 10:15 AM
Christner, Jody SAMPLE-3.1002	Month 3 Visit - 8/12/20 10:30 AM
Dagenhart, Matilda SAMPLE-1.1002	Month 2 Visit - 8/11/20 10:15 AM
Krauth, Alfonzo SAMPLE-2.1004	Month 1 Visit - 8/14/20 1:15 PM
Leos, Royal SAMPLE-1.1003	Month 2 Visit - 8/12/20 10:30 AM
Smithey, Rosalva SAMPLE-2.1002	Month 1 Visit - 8/13/20 2:30 PM
Tietz, Lamar SAMPLE-2.1001	Month 1 Visit - 8/11/20 11:00 AM
Vankeuren, Mathilda SAMPLE-3.1004	Month 3 Visit - 8/13/20 9:30 AM
Vossler, Alline SAMPLE-1.1005	Month 2 Visit - 8/10/20 10:45 AM

Ineligible

Devaney, Jenifer SAMPLE-2.1003	Screening Visit - 6/12/20 9:12 AM
Elsasser, Janean SAMPLE-1.1001	Screening Visit - 4/21/20 10:05 AM
Gularte, Vicente SAMPLE-1.1004	Screening Visit - 6/9/20 9:06 AM
Murrah, Hazel	Screening Visit - 7/10/20 1:43 PM

Figure 4.7. Participant list interface with sample participants

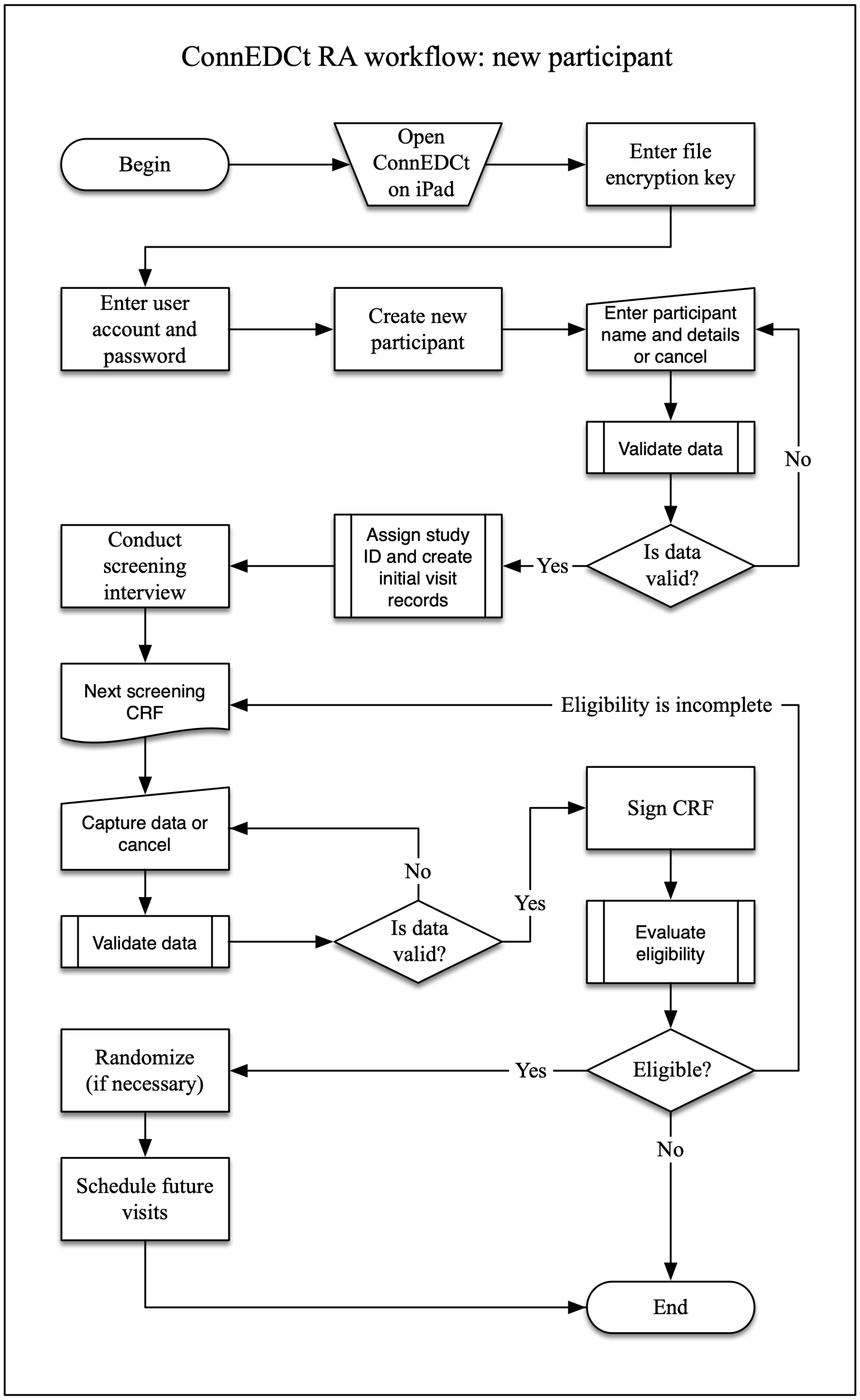
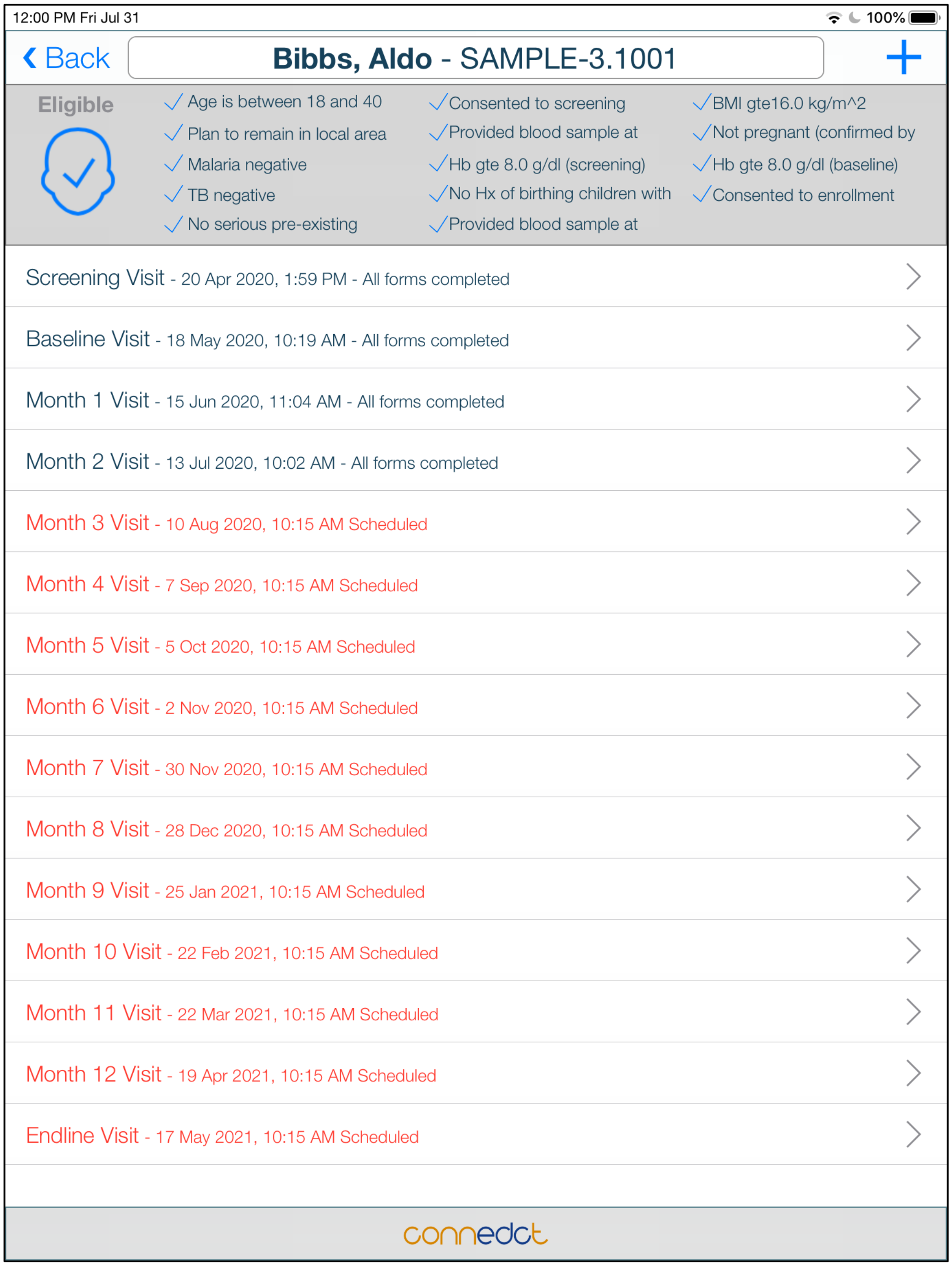


Figure 4.8. New participant workflow

Once a participant is enrolled – eligibility questions and consent are complete and randomization is applied, if necessary – the RA will schedule the remaining visits for the study. ConnEDCt then generates the visit and form records following the study schema. CRFs dependent upon predicate criteria will be created as needed when signing a form containing the predicate variables. Figures 4.9 and 4.10 show the visit list and form list interfaces.

RAs keep track of follow-up appointments with the agenda view, shown in Figure 4.11. The agenda view allows the RA to see all visits scheduled for a specific day or an entire week, as well as export the list if needed to hand off to a receptionist. The RA is able to step forward and backward by date to view the past, present, and future appointments.

An RA will create forms to capture unscheduled events such as dropouts, adverse events, and special notes from the event list interface with the 'new event' button. These ad hoc, special forms are essentially single form events and are listed within the events list. Using these form generation and data management features a study protocol can be completed.



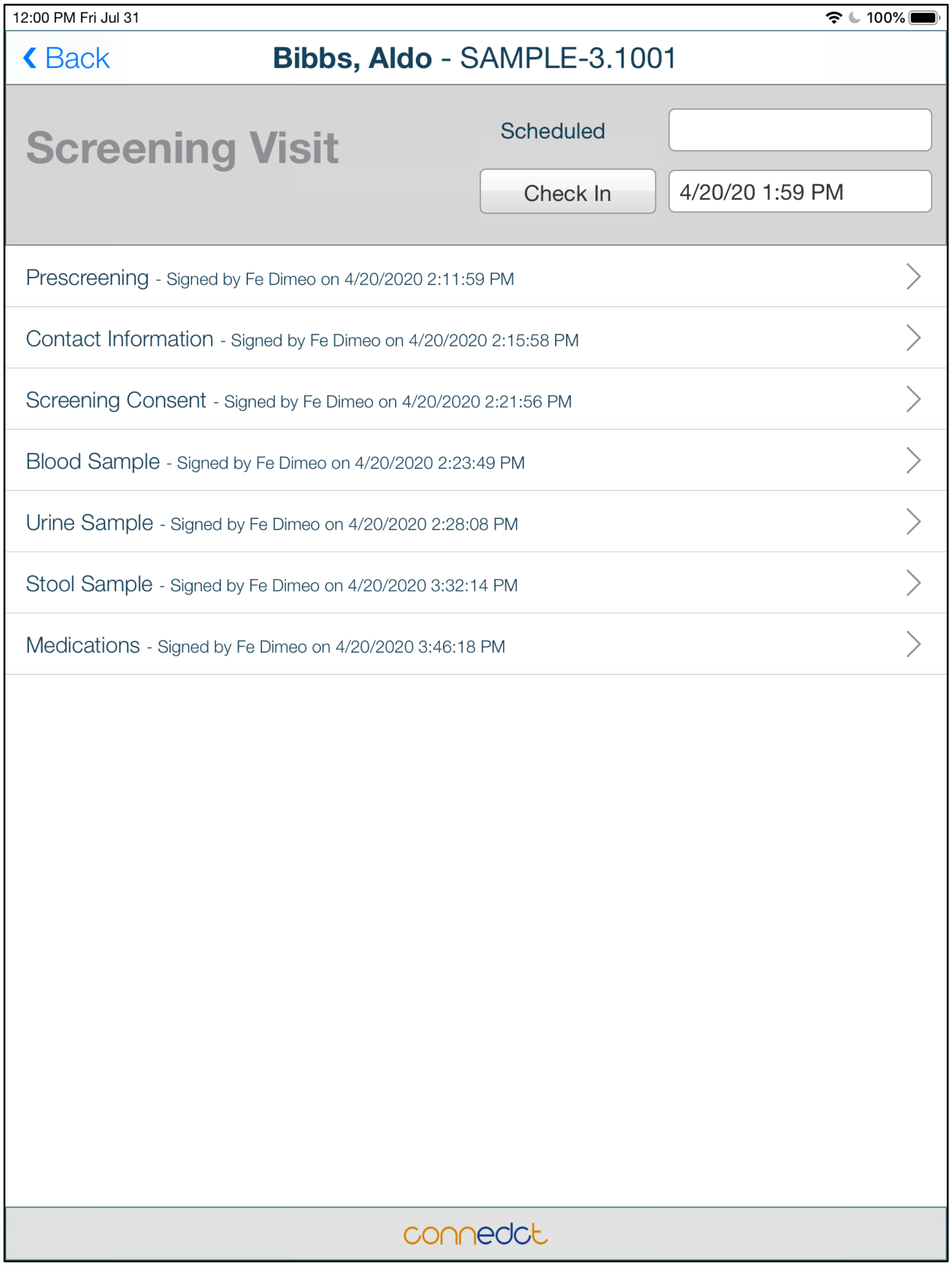
Bibbs, Aldo - SAMPLE-3.1001

Eligible

✓ Age is between 18 and 40	✓ Consented to screening	✓ BMI gte16.0 kg/m^2
✓ Plan to remain in local area	✓ Provided blood sample at	✓ Not pregnant (confirmed by
✓ Malaria negative	✓ Hb gte 8.0 g/dl (screening)	✓ Hb gte 8.0 g/dl (baseline)
✓ TB negative	✓ No Hx of birthing children with	✓ Consented to enrollment
✓ No serious pre-existing	✓ Provided blood sample at	

- Screening Visit - 20 Apr 2020, 1:59 PM - All forms completed
- Baseline Visit - 18 May 2020, 10:19 AM - All forms completed
- Month 1 Visit - 15 Jun 2020, 11:04 AM - All forms completed
- Month 2 Visit - 13 Jul 2020, 10:02 AM - All forms completed
- Month 3 Visit - 10 Aug 2020, 10:15 AM Scheduled
- Month 4 Visit - 7 Sep 2020, 10:15 AM Scheduled
- Month 5 Visit - 5 Oct 2020, 10:15 AM Scheduled
- Month 6 Visit - 2 Nov 2020, 10:15 AM Scheduled
- Month 7 Visit - 30 Nov 2020, 10:15 AM Scheduled
- Month 8 Visit - 28 Dec 2020, 10:15 AM Scheduled
- Month 9 Visit - 25 Jan 2021, 10:15 AM Scheduled
- Month 10 Visit - 22 Feb 2021, 10:15 AM Scheduled
- Month 11 Visit - 22 Mar 2021, 10:15 AM Scheduled
- Month 12 Visit - 19 Apr 2021, 10:15 AM Scheduled
- Endline Visit - 17 May 2021, 10:15 AM Scheduled

Figure 4.9. Visit list interface



Bibbs, Aldo - SAMPLE-3.1001

Screening Visit

Scheduled:

Check In: 4/20/20 1:59 PM

Prescreening - Signed by Fe Dimeo on 4/20/2020 2:11:59 PM

Contact Information - Signed by Fe Dimeo on 4/20/2020 2:15:58 PM

Screening Consent - Signed by Fe Dimeo on 4/20/2020 2:21:56 PM

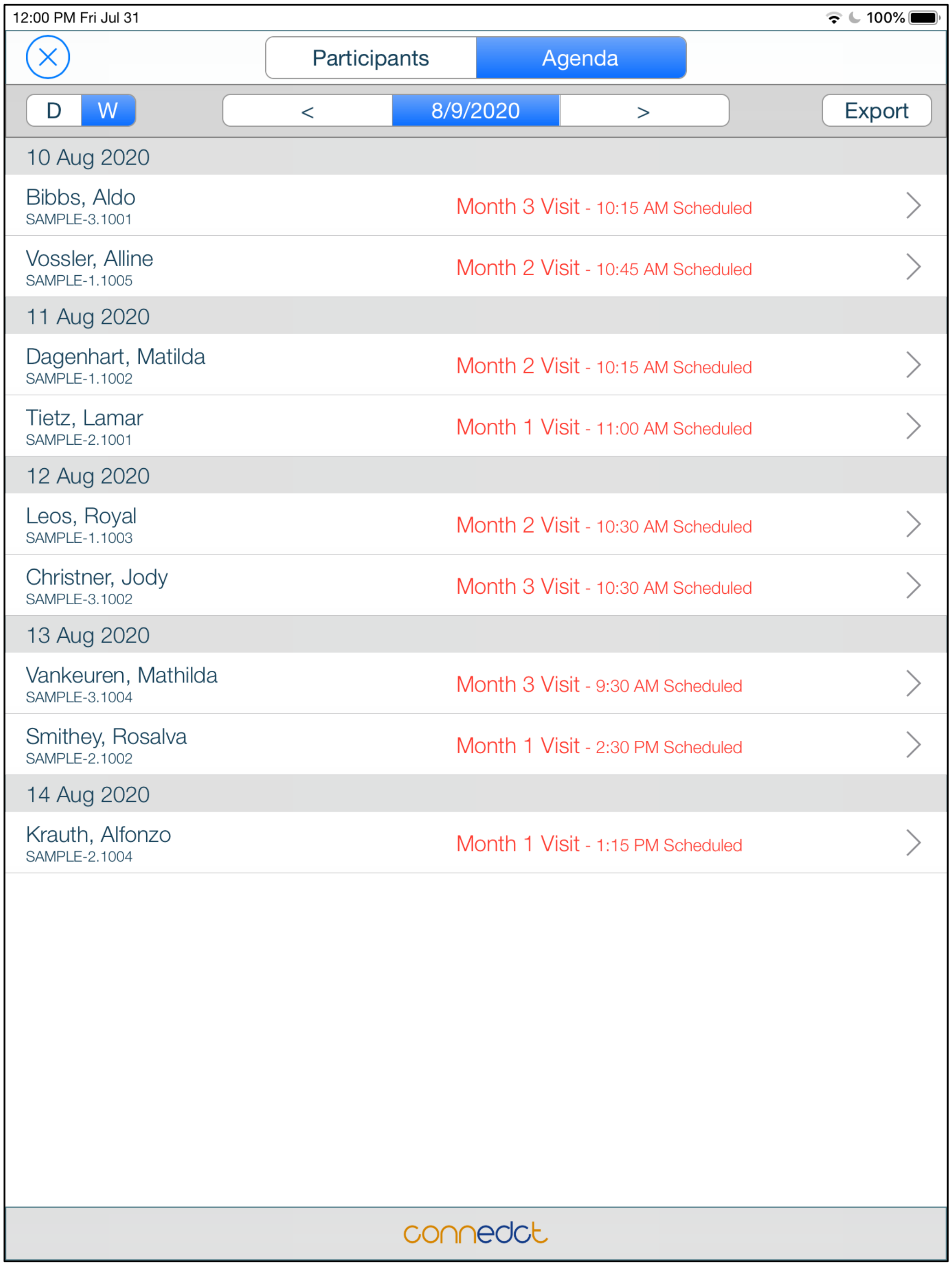
Blood Sample - Signed by Fe Dimeo on 4/20/2020 2:23:49 PM

Urine Sample - Signed by Fe Dimeo on 4/20/2020 2:28:08 PM

Stool Sample - Signed by Fe Dimeo on 4/20/2020 3:32:14 PM

Medications - Signed by Fe Dimeo on 4/20/2020 3:46:18 PM

Figure 4.10. CRF list interface



10 Aug 2020

Bibbs, Aldo
SAMPLE-3.1001
Month 3 Visit - 10:15 AM Scheduled

Vossler, Alline
SAMPLE-1.1005
Month 2 Visit - 10:45 AM Scheduled

11 Aug 2020

Dagenhart, Matilda
SAMPLE-1.1002
Month 2 Visit - 10:15 AM Scheduled

Tietz, Lamar
SAMPLE-2.1001
Month 1 Visit - 11:00 AM Scheduled

12 Aug 2020

Leos, Royal
SAMPLE-1.1003
Month 2 Visit - 10:30 AM Scheduled

Christner, Jody
SAMPLE-3.1002
Month 3 Visit - 10:30 AM Scheduled

13 Aug 2020

Vankeuren, Mathilda
SAMPLE-3.1004
Month 3 Visit - 9:30 AM Scheduled

Smithey, Rosalva
SAMPLE-2.1002
Month 1 Visit - 2:30 PM Scheduled

14 Aug 2020

Krauth, Alfonzo
SAMPLE-2.1004
Month 1 Visit - 1:15 PM Scheduled

Figure 4.11. Agenda view

5. CASE STUDIES

Six study protocols have been implemented with ConnEDCt. Five of those studies have moved ahead with enrollment and are continuing to use ConnEDCt in the field for primary data capture or have completed the investigatory phase of the study. These five active or completed studies are: Pearl Millet Biofortification (PMB), FeverPhone, Periconceptional Surveillance in India (PSI), Multi-biofortified Food Crops (MBFC), and Environmental Determinants of KSHV in Uganda (EDKU). Figure 5.1 shows the geographic distribution and relative participant volumes of the active and completed studies.

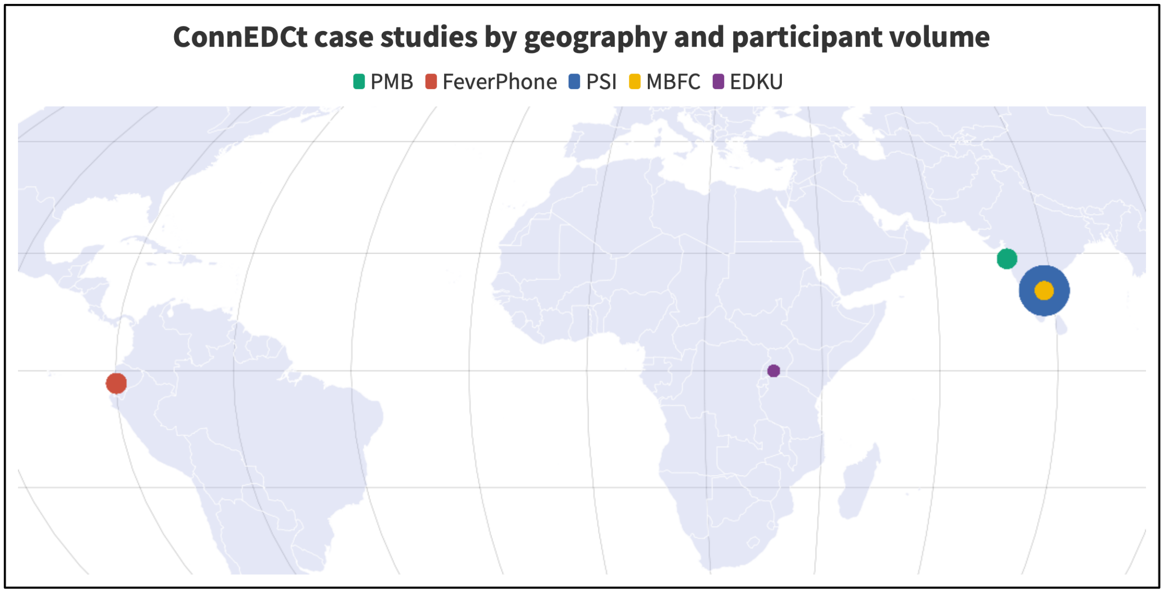


Figure 5.1 The five studies that have used or are using ConnEDCt for data capture by geography and participant volume. The radii of the dots are proportional to the participant volumes.

These studies have widely varying protocols and levels of complexity in their schemas, and the five studies that have gone ahead with data capture have widely varying volumes of data. This chapter will address specific complexities for each case study. Figure 5.2 shows the relative complexity of the schemas. FeverPhone shows a lower level of complexity with only three visits and a total of 18 instances of scheduled CRFs, whereas MBFC has 11 visits and almost ten times as many instances of scheduled CRFs.

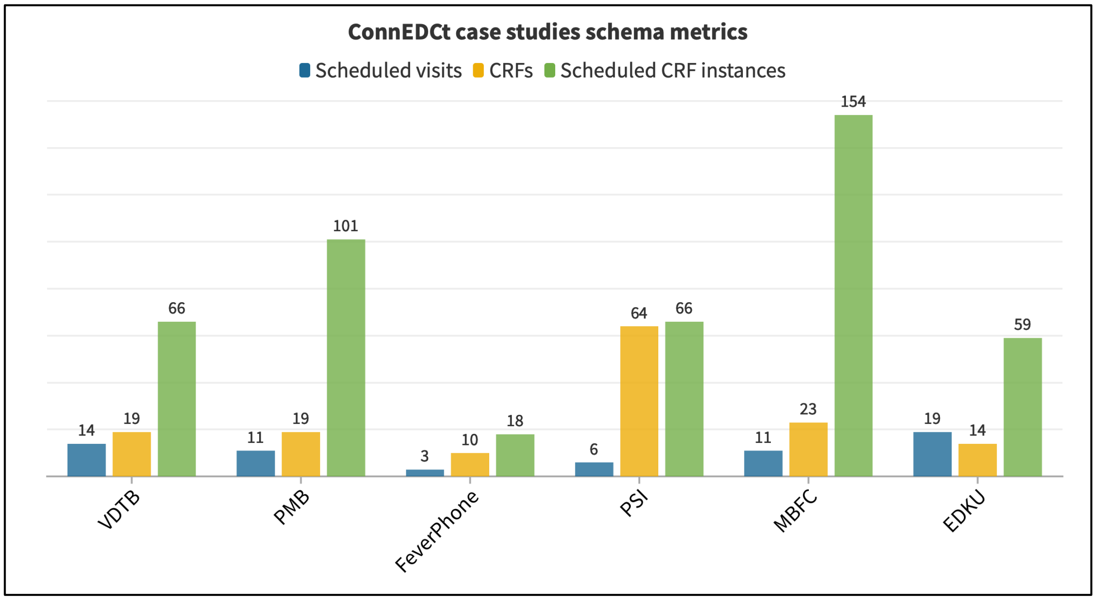


Figure 5.2 The schema metrics of the six case studies.

The static schema of scheduled visits for a study does not demonstrate the full complexity of the data capture protocol. Unscheduled CRFs can represent a significant volume of data to be captured and is the case with the PSI study. Even though the numbers of CRFs and scheduled CRF instances are roughly the same for the PSI study (the yellow and green bars in Figure 5.2), Figure 5.3, data capture metrics, shows the total number of signed CRFs (the green bars) to be highly differentiated from the actual participant encounters (the yellow bars). This is due to a high number of dynamically generated CRFs based on predicate variables. Additional CRFs were dynamically scheduled based on participants' responses. Dynamically scheduled CRFs are a significant part of this protocol. We will look at other specifics of the case studies in the following sections. Detailed schema and data capture metrics are shown in the appendixes.

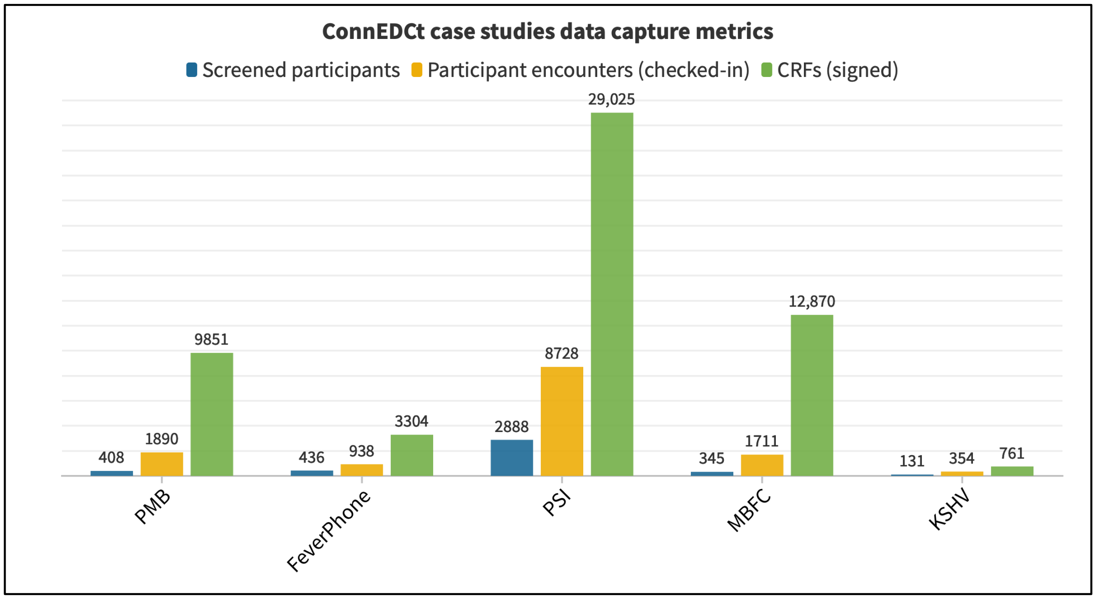


Figure 5.3 ConnEDCt case studies data capture metrics

5.1. Vitamin D Supplementation Among TB Patients in South India

The genesis of ConnEDCt was the need to have an EDC tool for an RCT that would operate in rural settings with only occasional, slow access to the Internet. The VDTB trial was designed "to assess how vitamin D supplementation affects immunity"[54] in a participant pool of 200 tuberculosis-positive adults in southern India. The trial was expected to require a combination of home visits and clinic visits with a staff of trained RAs conducting the data capture on iPads. The requirement for home visits in rural India necessitated disconnected data capture operations.

This trial was our pilot project and the first implementation of ConnEDCt and was the initial basis of the framework. The research team developed the initial CRF designs within the FileMaker IDE. Meanwhile, framework development proceeded, including interface designs, navigation, data integrity assurance, electronic signatures, syncing, consent, eligibility, coded lists, and scheduling methodologies. A great deal of adjustment to the organization of variables and CRFs was made during the design process. This work included consolidating CRFs, moving variables from one CRF to another, and adding or eliminating variables. These adjustments were made to expand or limit the scope of captured data, improve clinical workflow, and reduce the time needed to capture data. Because these adjustments were made after the CRFs had been implemented within ConnEDCt, they required a large amount of developer effort and iterative review by the research team. So early on we learned of the need for greater autonomy by the research team in the design process.

This trial was put on hold and never proceeded to recruitment. However, the development process revealed to us several areas in EDC, such as eligibility determination and complex scheduling, that would be beneficial to adopt in the framework.

5.2. Pearl Millet Biofortification Trial in Mumbai

The PMB trial was the first full implementation of ConnEDCt. This RCT in Mumbai, India investigated "the effect of the consumption of foods prepared with iron- and zinc-biofortified pearl millet (FeZn-PM) by children on biomarkers of iron and zinc status, growth, and immune function."[53, 55] We added a number of enhancements to the ConnEDCt framework to support this trial. Feature development focused on offline operation and randomization. Again the CRF design process was highly iterative with extensive collaboration between the study design team and the engineering team.

Newly developed features to support offline operation included syncing, offline device registration, visit scheduling, and the federated study ID algorithm. To minimize the creation of irrelevant records and optimize syncing efficiency, RAs manually scheduled future visits after eligibility is satisfied. New features to support randomization included randomization of blinded study groups and randomized CRF serialization. Randomization began with a federated model and adjusted to a centralized model. Both modes are still supported in the framework.

At the end of the implementation process, the trial began with a three-month pilot project that allowed the team to refine the clinical workflow and extensively test the syncing functionality. The initial syncing methodology proved to be unreliable in rural India where Wi-Fi and mobile internet was slow, with high latency, and often-dropped connections. Early on we changed to the MirrorSync module that performed faster and recovered from dropped connections more reliably. Even after switching to MirrorSync, syncing was challenging for several months until more reliable internet connectivity was established. Frequently dropped connections resulted in incomplete syncing and created syncing conflicts that required manual resolution. Manual resolution of syncing conflicts required both a deep understanding of the syncing concept and a firm grasp of the data model. Frequently consultation between an engineer and the DM was required to resolve syncing conflicts.

Participant recruitment began in March of 2017 and data capture was completed in August 2018. We encountered a few issues during the year and half of data capture. Because RAs scheduled visits manually, it was possible for duplicate visits to be created on different iPads by different RAs. After syncing, the duplicate visits appeared on all devices. It required careful attention by the DM to mark and delete duplicates and repeat the sync to remove the duplicates from all devices. We mitigated this issue in future studies by revising the framework to automate scheduling.

The study design team extensively planned and tested the CRFs and the study protocol throughout implementation and the pilot. The DM and local RA team tested many iterations of CRF designs and the newly added platform features. Despite this extensive testing, additional changes were required to CRFs after recruitment had begun. In some cases the order of questions on CRFs was modified to better match the clinical workflow. In other cases a few variables were added to CRFs. One change was made to correct a flaw in the eligibility algorithm. The requirements for changes despite the heavy testing pre-launch further affirmed the need for investigator-defined study schema and a data model that facilitates deployment of schema changes.

5.3. FeverPhone

The ongoing FeverPhone study in Ecuador aims to support the research and development of a point of care diagnostic device for febrile illnesses. [56] The implementation of the FeverPhone study required very few programmatic additions to ConnEDCt. It was, therefore, a good model for developing an efficient implementation procedure.

Some modifications were made specifically to support the FeverPhone study protocol. The interface was localized in the Spanish language to support the local research staff in Ecuador. This localization was hard-coded. Since ConnEDCt is used internationally, language localization is a candidate for future development. The protocol also called for separate consent or assent forms based on the age of the participant.[57] Adult participants receive a consent form, whereas minors require a parental consent form and a participant assent form. FeverPhone required very little modification or support post-launch. Participant recruitment began in June of 2017 and is continuing indefinitely. The timing of data capture focuses on the seasonality of febrile illnesses in the region.

5.4. Periconceptional Surveillance in India

Several severe, neurological birth defects, such as spina bifida, are linked to vitamin B-12 and folate status during pregnancy.[58, 59] The PSI study provided pre-intervention biomarker data to aid the development of a future, randomized, controlled efficacy trial.

PSI has been the most challenging implementation of ConnEDCt due to the large number of CRFs and the complex protocol that went through extensive revisions and refinements during implementation. Although ConnEDCt's customizable framework model was well suited for the study, the design process further reinforced the need for more of the protocol implementation to be independent of developer involvement. ConnEDCt became a useful tool for the research team to experiment and clarify issues with the study protocol. Previous studies with simpler protocols required far less programmer effort and fewer revisions. It is understandable that a complex protocol is more difficult to design and may require customization and new features added. Ultimately, ConnEDCt became the study protocol-modeling tool for the research team.

PSI was a complex study with a very detailed, rules-based schema and lengthy CRFs. To ensure data quality a great deal of effort was put into applying precise validation rules to ensure values fell into controlled ranges.[57] Many of these validation rules also referenced other variables, creating complex dependencies. Implementing the rules required a great deal of testing and debugging to refine the validation rules and eliminate issues such as circular references and other design or coding issues. The effectiveness of these rules, as well as their complexity, demonstrated the value of having technical resources available during the study design process.[57]

New feature development for this study focused on dynamically generated CRFs and integration with external systems. In this case most of these features were executed as custom programming. While some new concepts were identified as being important to the framework, because of the limited timeframe little of the custom programming was incorporated into the framework during the implementation of the study. The dynamic CRF features were later implemented as framework elements and used in the MBFC study. These dynamic CRF creation events are triggered either upon completion of a CRF or triggered during the scheduling procedure to evaluate a condition when additional CRFs should be generated.

The PSI study protocol designates two sets of CRFs, defined by whether the participant had a history of pregnancy. The number of past pregnancies is captured in the health history CRF. When an RA signes the CRF, the event triggers the creation of a number of new pregnancy CRFs, based on the total number of pregnancies. During scheduling a controller evaluates the prior pregnancy status to determine whether the set of pregnancy or non-pregnancy

CRFs should be created. This pre- and post-processing CRF creation model was later adopted as a framework element.

PSI also required integration with an external system that stored census data of the cohort of potential participants. RAs used the census tool to create participants in ConnEDCt. The census tool created a JSON data object payload with participant identity and demographic data and delivered it to an external hook in ConnEDCt. Upon receiving the JSON payload, ConnEDCt parses it, and creates the participant record and the study ID. RAs conducted informed consent and eligibility screening in ConnEDCt to complete participant enrollment.

The PSI study began recruiting in June of 2018 and was completed in July of 2019. Because of the thorough design phase, no adjustments to the data schema were made during the data capture phase of the study and little post-deployment support was needed.

5.5. Effect Of Multiple Biofortified Food Crops On Micronutrient Status, Immune, And Cognitive Function Among Indian Infants

"Iron, zinc, and vitamin A deficiency remain a major worldwide public health problem especially in developing countries."[60] MBFC is an RCT to compare young children consuming meals prepared with multiple biofortified food crops with children receiving a typical diet "to measure growth, cognitive changes, and immune function."[60]

MBFC followed closely the protocols established with the PMB trial with a few exceptions. The close adherence to a previously implemented protocol allowed us to focus efforts on streamlining the implementation procedure, implementing style sheets for the CRFs, improving workflows such as scheduling and device registration, and building user interfaces to manage some data such as coded lists, that previously had been managed on the back-end.

Significant changes were made to the CRF scheduling procedure. In the previous PMB study RAs had manually scheduled each visit. For MBFC, ConnEDCt automated the scheduling of the entire protocol after randomization. The use of randomized midline serial-sampled CRFs required scheduling to occur after the randomization of eligible participants. The DM randomized participants centrally by manually triggering the randomization of eligible participants in batches. The randomization process would then automatically trigger the scheduling process.

Another improvement was to automate the device registration process. Because ConnEDCt works off-line with a decentralized study ID numbering scheme, each device that runs ConnEDCt has a serialized ID of its own that is assigned by registering the device with the centralized data store prior to use. In other words each device is assigned a unique serialized number, like 1, 2, 3, etc. The registration process was improved so that devices could self-assign their device ID and prompt a data manager to validate the registration status.

After being tested in the clinical environment, MBFC underwent extensive revisions to CRFs, including adding variables, removing variables, moving variables to other CRFs, removing CRFs, adding new CRFs, changing the visits that CRFs are scheduled in, and changing eligibility formulae.

MBFC had more user errors than other studies, including data being entered in the wrong scheduled visit several times – RAs used the correct CRF, but in the wrong visit. The error was corrected by manually moving the CRF record to the correct scheduled visit or by clearing all data in the record and the RA re-recording the CRF. Such errors rarely happened in prior studies, yet it indicated that users need more help identifying and navigating to the correct visit. The MBFC trial began recruiting in March 2019 and is ongoing.

5.6. Environmental Determinants of KSHV in Uganda

This longitudinal cohort study looks at Kaposi's sarcoma herpes virus, the causative agent of Kaposi's sarcoma, and its association with malaria infection, with the goal of identifying strategies for the prevention of this type of cancer.[61] EDKU provided an opportunity to standardize and adopt into the framework several features that had been conceived of previously, but had not been implemented or had been implemented only as custom code. These standardized features included dynamically generated CRFs, better scheduling automation, and an agenda view of participant appointments.

Because the EDKU study schema also required CRFs being generated dynamically based on captured data in real time, dynamic CRF generation has been incorporated into the pre-processing controller in the CRF generation process and the post-processing controller of the CRF signing process. These are now framework features that are easy to implement when a study schema requires.

To make the scheduling process more user and participant-friendly ConnEDCt now requests a preferred day of the week and a time of day and then schedules visits adaptively. Scheduling is triggered manually for each participant after eligibility is confirmed so that the RAs can confirm the best appointment days and times with the participant.

The agenda view (figure 4.11) provides a daily or weekly view of appointments to help RAs avoid selecting a visit from the wrong date. In addition the visit closest to the current date in the visit list interface (figure 4.9) is highlighted to further help identify the relevant visit.

The EDKU study reaffirmed the need for a more flexible data model for participants. The study tracks mother/child dyads, whereas ConnEDCt had been designed to track only individual participants, so we adapted ConnEDCt to store mother and child information. However, adopting the well-established party data model[62] would provide a more effective way of linking family members and other participant associations. The EDKU study began recruiting in February 2020 and is ongoing.

6. FUTURE ENHANCEMENTS

Implementing ConnEDCt for three clinical surveys and three RCTs provided broad experience with varying protocols and plenty of inspiration for improving the framework. Some features were placed on the development roadmap from early in the planning process, and others have been uncovered along the way.

6.1. Investigator-defined Study Schema

From the beginning the need to reduce developer involvement in implementing a study schema became apparent in order to scale the use of ConnEDCt. The more studies that can be implemented in a shorter time, the lower cost of each implementation, the less reliance on finite resources, and the more flexibility for researchers. CRF development has been the most time-intensive aspect of every ConnEDCt implementation. The back-and-forth between study designers and software developers takes many rounds before a design can be finalized. Iteration is simply a part of the study design process. Even finalized designs are frequently revised once study recruitment has begun. Revisions of the CRF design should be within the control of study designers.

The present data dictionary design document is too abstract for study designers to envision the final questionnaires. The better approach will be to build a design tool to construct the data dictionary that translates automatically into usable CRFs. This ConnEDCt design tool will allow study designers to define variables and data validation to directly generate the end result forms. Custom development will still be required from time to time to implement new ideas for innovative protocols. However, providing a tool that allows designers to construct the meta-data associated with a basic study schema would provide two advantages. First, it will reduce or eliminate developer involvement in the study protocol design phase, and therefore lower costs. Second, it will allow the study designers to test and iterate their designs rapidly and thus save time.

6.2. Statistics and Reporting

Better statistics availability will help RAs and data managers keep track of study progress during the data capture phase of studies. ConnEDCt currently shows numbers of eligible, incomplete, and ineligible participants. To get other stats DMs must export datasets and run their own analysis to get more in-depth information. A dashboard view on opening can provide a useful overview of the study. Helpful stats will be the number of enrolled participants in each study group, the number of dropouts, and the number of appointment no-shows with links to the participants to help RAs follow-up. Breaking out informed consent as an independent number from eligibility can help differentiate issues with consent versus other eligibility factors. Viewing other data over time, such as enrollment trends or missed appointments, can help manage the performance of the study staff or communication issues with participants. A dashboard will provide simple real-time statistics and eliminate the need for data exporting and analysis to get key performance indicators on the study progress.

The CONSORT Statement[63] is a formalized, respected reporting format for RCTs developed by a group of scientists and editors that includes a checklist and a flowchart to avoid systemic error in the assessment of study results. CONSORT has been supported by hundreds of academic journals "to provide guidance to authors about how to improve the reporting of their trials."[64] Because the CONSORT Statement is structured information consisting of a table and a flow chart (figure 6.1), the format can be readily constructed in template form. Some information can be directly derived from data that ConnEDCt captures or generates. The rest can be presented in a template form for the investigators to complete. The full CONSORT Statement can then be exported or printed as a reference for investigators producing manuscripts for publication.

CONSORT
TRANSPARENT REPORTING of TRIALS

CONSORT 2010 Flow Diagram

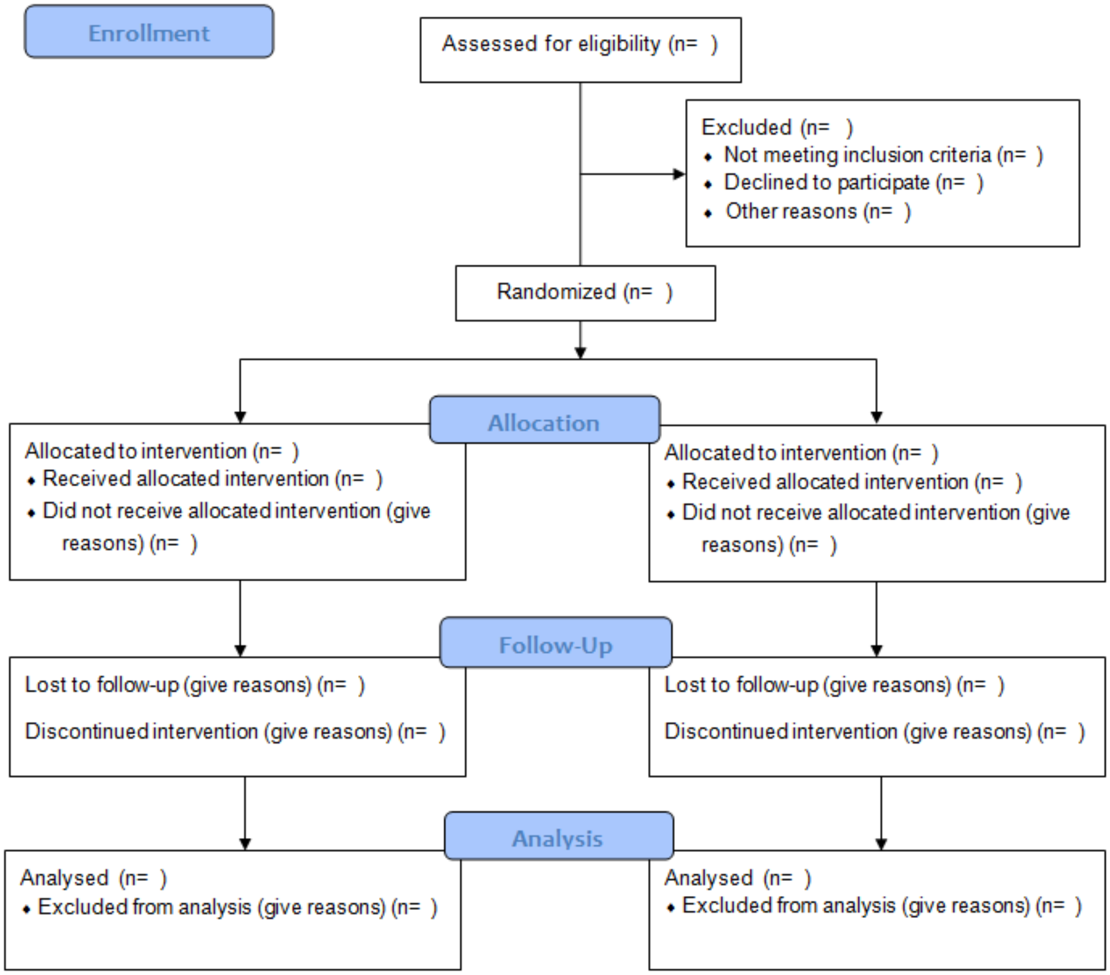

Enrollment
Assessed for eligibility (n=)
Excluded (n=)
• Not meeting inclusion criteria (n=)
• Declined to participate (n=)
• Other reasons (n=)
Randomized (n=)
Allocation
Allocated to intervention (n=)
• Received allocated intervention (n=)
• Did not receive allocated intervention (give reasons) (n=)
Allocated to intervention (n=)
• Received allocated intervention (n=)
• Did not receive allocated intervention (give reasons) (n=)
Follow-Up
Lost to follow-up (give reasons) (n=)
Discontinued intervention (give reasons) (n=)
Lost to follow-up (give reasons) (n=)
Discontinued intervention (give reasons) (n=)
Analysis
Analysed (n=)
• Excluded from analysis (give reasons) (n=)
Analysed (n=)
• Excluded from analysis (give reasons) (n=)

Figure 6.1. CONSORT flow diagram[65]

6.3. Electronic Signatures for Informed Consent

ConnEDCt supports electronic signatures for users of the system, including DMs and RAs, but not participants who are not direct users of ConnEDCt. This lack of support for participants' e-signatures is driven primarily by the challenge of determining the authenticated identity of participants as defined for electronic systems.

Providing support for participants' electronic signatures would make the informed consent process completely electronic, totally eliminating the need for paper records. Username and password combinations authenticate study investigators. Username and password authentication validates electronic signatures under current regulations. However, since participants are not assigned user accounts in ConnEDCt, another method must validate identities within the electronic system. This is especially challenging in an offline context. Some options we will consider for creating validated electronic signatures for participants are photo or video capture of physical identity verification, such as an identity card, along with witness attestation and biometrics validated with machine learning models.

6.4. Improved Data Transport and Integrations

Data transport to external statistics packages has been challenging. While the Microsoft Excel format seems to be universally accepted, some issues such as column headers and null values are not fully resolved. Other options will be explored, including translation to native file formats.

We are actively developing an integration solution to acquire machine-generated lab results and link them with participants. Few laboratory devices support direct integration, so we are prioritizing an import methodology for the data files generated by lab devices. DMs will be able to select the data file, expunge irrelevant headers, customize column mapping, and associate sample IDs with participant CRFs.

6.5. Data Model Flexibility

With FeverPhone and EDKU, we adapted ConnEDCt to support parent/child dyads in the data single-table entity that stores participant data. This functioned well enough but was not ideal. In two opportunities for census enumeration where households and family groups were to be enumerated together, ConnEDCt was deemed unsuitable for data capture due to the rigidity of the single-table entity for tracking participants.

The party data model[62] is a data model that can support the relationships between participants such as parent/child dyads, family groups, households, and other relationships. Adopting the party data model will expand the scope of ConnEDCt's data capture utility.

All of the studies that employed ConnEDCt required some extent of localization of language. UI elements were presented in a local language only, English with modifications for dialect, or both English and a local language side by side. This applied to the interview questions, coded list options, and UI navigational elements. Localization features built into the investigator-defined data model will allow study designers to have full control over these elements as well as provide users with a choice of localization option. Therefore, Spanish-speaking users, for example, can view the interface in Spanish, while English-speaking users can view the same interface in English.

6.6. Automated Updates

Even with the investigator defined study schema and flexible data model, experience has shown that there will be customizations made post-deployment that will require updates. The current update process requires users to replace files on their devices manually. An automated update process would involve detecting the existence of an update, verifying the sync status, optionally triggering data sync, downloading an updated database file in the background and replacing the old database file. This process could be completed with minimal user-interaction with a user prompt to begin and a confirmation notice at the end.

6.7. Documentation

Finally, we have developed several standard operating procedure (SOP) documents for ConnEDCt. These SOPs will be collected and organized into a volume of comprehensive documentation. While ConnEDCt is intuitive to use and requires little training, the need for documentation is evident from the repetition of some of the same questions new users ask. In addition the U.S. Department of Health and Human Services has published non-binding recommendations regarding documentation of electronic records systems.[66] Some of the current SOPs, such as those for deploying updates, will become obsolete when features described in this chapter are completed, yet some documentation will still be essential.

7. CONCLUSION

In Chapter 1, I introduced the elements of clinical data capture protocols and ConnEDCt, the software framework I built in cooperation with clinical investigators to support complex EDC protocols in resource-constrained areas. But what makes ConnEDCt a framework? A software framework should provide context-specific yet generic features that can be further adapted to additional requirements. The framing elements for clinical electronic data capture include data security and integrity, consent forms, eligibility criteria, visit scheduling, CRFs, and randomization with study groups. ConnEDCt provides an implementation framework that supports these features on a mobile platform and it has been expanded with customized features during almost every implementation process.

After implementing six successful and highly differentiated study protocols, ConnEDCt has proven to be an effective EDC framework. It has provided standard study protocol features and compliance with government regulations, while the framework architecture has supported feature expansion. With nearly every new implementation, new opportunities have been uncovered to add greater sophistication to ConnEDCt while maintaining its underlying code. The framework architecture has supported this ongoing expansion of features like randomized serial sampling and contingent CRFs. The six successful implementations for research in resource-constrained locations, have proven ConnEDCt's effectiveness and flexibility[53-61].

ConnEDCt has demonstrated value for investigators who wish to perform research in remote areas where internet is unreliable and for those who need advanced protocol support. While ConnEDCt will benefit from improvements in user-defined schema definition, real-time stats, and some other areas, we are actively improving it and hope to raise the visibility of ConnEDCt as a commercial option for clinical EDC.

ConnEDCt is far more advanced than the tool I created for the TBI survey in the late 1990s and includes features for protocol enforcement, offline function, and flexibility that other popular EDC tools, including the highly popular REDCap[19], lack. Although information technology has evolved tremendously since then, there are still poor, resource-constrained geographical areas that suffer not just from lack of pervasive internet but also from what many of us would consider basic healthcare backed by medical science. ConnEDCt is a flexible platform that prioritizes mobility, is adaptable to different study protocols, has extensive study protocol support built-in, and functions on- or offline with data synchronization to a central data repository. ConnEDCt serves clinical research needs with a focus on mobility and flexibility. I intend to continue with the development of ConnEDCt and the engagement with clinical researchers, thereby supporting clinical research, especially those studies conducted in rural or resource-constrained areas, and improving clinical investigators' access to quality EDC tools.

APPENDIX A: Selected Schema Entities' Attributes

Note that FileMaker data types do not conform to standard SQL data types.

Number and text are valid FileMaker data type definitions.

Device table

Field Name	Data Type
ID	Text
PersistentID	Text
Type	Text
Name	Text
Number	Number
NextStudyID	Number
ScreenHeight	Number
ScreenWidth	Number
Sensors	Text
SystemLanguage	Text
SystemPlatform	Text
IsRegistered	Boolean

Appendix A: Table 1. Device table attributes

EligibilityCriterion table

Field Name	Data Type
ID	Text
ID_FormType	Text
ID_VisitType	Text
Description	Text
Formula	Text

Appendix A: Table 2. EligibilityCriterion table attributes

PartEligibilityCriterion table

Field Name	Data Type
ID	Text
ID_EligibilityCriterion	Text
ID_Participant	Text
Status	Number

Appendix A: Table 3. PartEligibilityCriterion table attributes

VisitType table

Field Name	Data Type
ID	Text
Name	Text
Order	Number
ScheduleDaysFromStart	Number
CreateNextVisit	Boolean
IsEligibilityRequired	Boolean

Appendix A: Table 4. VisitType table attributes

FormSchedule table

Field Name	Data Type
ID	Text
ID_FormType	Text
ID_VisitType	Text
Order	Number
IsSerialized	Boolean
IsSubset	Boolean

Appendix A: Table 5. FormSchedule table attributes

PartEventForm table

Field Name	Data Type
ID	Text
ID_FormType	Text
ID_PartEvent	Text
isComplete	Boolean
IsSynced	Boolean
SignatureLocation	Text
SignatureTS	Timestamp
SignatureName	Text
SignatureString	Text

Appendix A: Table 6. PartEventForm table attributes

StudyGroup table

Field Name	Data Type
ID	Text
Code	Text
Description	Text

Appendix A: Table 7. StudyGroup table attributes

Randomization table

Field Name	Data Type
ID	Text
ID_StudyGroup	Text
ID_Device	Text
isAssigned	Boolean
Order	Number
SerialSampleNumber	Number

Appendix A: Table 8. Randomization table attributes

APPENDIX B: Case Studies Tabular Data

Case studies schema metrics	VDTB	PMBT	FeverPhone	PSI	MBFC	KSHV
CRFs	19	19	10	64	23	14
Variables	881	753	386	3239	808	354
Consent forms	2	3	3	1	1	1
Scheduled visits	14	11	3	6	11	19
Scheduled CRF instances	66	101	18	66	154	59
Eligibility criteria	9	13	3	5	15	4
coded lists	93	73	30	112	104	26
Randomized study groups	4	2	0	0	4	0

Appendix B: Table 1. Case studies schema metrics

Case studies data capture metrics	PMBT	FeverPhone	PSI	MBFC	KSHV	VDTB
status	completed	ongoing	completed	completed	ongoing	not started
data compiled on	23-Aug-20	23-Aug-20	23-Aug-20	23-Aug-20	23-Aug-20	NA
Data capture devices (iPads)	4	5	17	12	5	-
Screened participants	408	436	2888	345	131	-
Participant encounters (checked-in)	1890	938	8728	1711	354	-
Participant encounters (scheduled)	2612	938	17305	3323	1453	-
Signed CRFs	9851	3304	29025	12870	761	-
Scheduled CRFs	19660	5108	41683	26310	4369	-

Appendix B: Table 2. Case studies data capture metrics

BIBLIOGRAPHY

1. Embi, P.J. and P.R.O. Payne, *Clinical Research Informatics: Challenges, Opportunities and Definition for an Emerging Domain.* Journal of the American Medical Informatics Association, 2009. **16**(3): p. 316-327.

2. Walther, B., et al., *Comparison of electronic data capture (EDC) with the standard data capture method for clinical trial data.* PLoS One, 2011. **6**(9): p. e25348.

3. Meyer, J., et al., *A mobile and asynchronous electronic data capture system for epidemiologic studies.* Computer Methods and Programs in Biomedicine, 2013. **110**(3): p. 369-79.

4. *Measuring the Information Society Report*. 2017, International Telecommunication Union: Geneva Switzerland.

5. Wu, H. *900 million Indians can't get online. Here's why.* CNN Tech 2016 March 9, 2016 [cited 2018 March 19, 2018]; Available from: http://money.cnn.com/2016/03/09/technology/india-internet-access/index.html.

6. SpringerLink, R.L. Richesson, and J.E. Andrews, *Clinical research informatics*, in *Health Informatics,*. 2012, Springer,: London ; New York. p. 1 online resource (ix, 419 p.

7. *Principles and Practice of Clinical Research*. 3rd ed, ed. J.I. Gallin and F.P. Ognibene. 2012, Amsterdam; Boston: Elsevier Science & Technology. 797.

8. Brody, T., *Clinical Trials: Study Design, Endpoints and Biomarkers, Drug Safety, and FDA and ICH Guidelines*. 2011: Academic Press.

9. Shantala, B., K. Binny, and M.S. Latha, *Basics of case report form designing in clinical research.* Perspectives in Clinical Research, 2014. **5**(4): p. 159-166.

10. Grady, C., MD, *Ethical Principles in Clinical Research*, in *Principles and Practice of Clinical Research*, J.I. Gallin, F.P. Ognibene, and L.L. Johnson, Editors. 2018, Academic Press: Cambridge, MA.

11. Shaw, P.A., L.L. Johnson, and C.B. Borkowf, *Issues in Randomization*, in *Principles and Practice of Clinical Research*, J.I. Gallin, F.P. Ognibene, and L.L. Johnson, Editors. 2018, Academic Press: Cambridge, MA.

12. Cummings, S.R., MD, D.G. Grady, MD, MPH, and S.B. Hulley, MD, MPH, *Designing a Randomized Blinded Trial*, in *Designing Clinical Research*, S.B. Hulley, et al., Editors. 2013, Lippincott Williams & Wilkins: Philadelphia, PA.

13. Porta, M.S., *A dictionary of epidemiology*. Sixth edition / edited by Miquel Porta. [ed. Epidemiology. 2014, New York]: New York : Oxford University Press.

14. Weber, B.A., et al., *A comparison study: paper-based versus web-based data collection and management.* Applied Nursing Research, 2005. **18**(3): p. 182-185.

15. Pawellek, I., et al., *Use of electronic data capture in a clinical trial on infant feeding.* European Journal of Clinical Nutrition, 2012. **66**(12): p. 1342-1343.

16. Pavlović, I., T. Kern, and D. Miklavčič, *Comparison of paper-based and electronic data collection process in clinical trials: Costs simulation study.* Contemporary Clinical Trials, 2009. **30**(4): p. 300-316.

17. Dillon, D.G., et al., *Open-source electronic data capture system offered increased accuracy and cost-effectiveness compared with paper methods in Africa.* Journal of Clinical Epidemiology, 2014. **67**(12): p. 1358-1363.

18. McLean, E., et al., *Implementing electronic data capture at a well-established health and demographic surveillance site in rural northern Malawi.* Global Health Action, 2017. **10**(1): p. 1367162.

19. Harris, P.A., et al., *Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support.* Journal of Biomedical Informatics, 2009. **42**(2): p. 377-81.

20. Franklin, J.D., A. Guidry, and J.F. Brinkley, *A partnership approach for Electronic Data Capture in small-scale clinical trials.* Journal of Biomedical Informatics, 2011. **44**: p. S103-S108.

21. Morak, J., et al. *Electronic data capture platform for clinical research based on mobile phones and Near Field Communication technology*. in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2008.

22. Pakhare, A., S. Bali, and G. Kalra, *Use of mobile phones as research instruments for data collection.* Indian Journal of Community Health, 2013(2): p. 95-98%V 25.

23. Patel, V., et al., *Developing a smartphone 'app' for public health research: the example of measuring observed smoking in vehicles.* Journal of Epidemiology and Community Health, 2013. **67**(5): p. 446-452.

24. Heerden, A.C.v., et al., *Collecting Health Research Data: Comparing Mobile Phone-assisted Personal Interviewing to Paper-and-pen Data Collection.* Field Methods, 2014. **26**(4): p. 307-321.

25. King, C., et al., *Electronic data capture in a rural African setting: evaluating experiences with different systems in Malawi.* Global Health Action, 2014. **7**: p. 10.3402/gha.v7.25878.

26. Vélez, O., et al., *A Usability Study of a Mobile Health Application for Rural Ghanaian Midwives.* Journal of Midwifery & Women's Health, 2014. **59**(2): p. 184-191.

27. von Niederhäusern, B., et al., *Validity of mobile electronic data capture in clinical studies: a pilot study in a pediatric population.* BMC Medical Research Methodology, 2017. **17**(1): p. 163.

28. Soti, D.O., et al., *Feasibility of an innovative electronic mobile system to assist health workers to collect accurate, complete and timely data in a malaria control programme in a remote setting in Kenya.* Malaria Journal, 2015. **14**: p. 430.

29. Style, S., et al., *Experiences in running a complex electronic data capture system using mobile phones in a large-scale population trial in southern Nepal.* Global Health Action, 2017. **10**(1): p. 1330858.

30. van Dam, J., et al., *Open-source mobile digital platform for clinical trial data collection in low-resource settings.* BMJ Innovations, 2017. **3**(1): p. 26-31.

31. Zhang, J., et al., *Mobile Device–Based Electronic Data Capture System Used in a Clinical Randomized Controlled Trial: Advantages and Challenges.* Journal of Medical Internet Research, 2017. **19**(3): p. e66.

32. Eagleson, R., et al., *Implementation of clinical research trials using web-based and mobile devices: challenges and solutions.* BMC Medical Research Methodology, 2017. **17**: p. 43.

33. *Trials of war criminals before the Nuernberg military tribunals under control council law No. 10*. 1949, Washington: U. S. Govt. Printing Office: Nuernberg.

34. *DECLARATION OF HELSINKI*. 1964, 18th World Medical Assembly: Helsinki, Finland.

35. Beecher, H.K., *Ethics and Clinical Research.* New England Journal of Medicine, 1966. **274**(24): p. 1354-1360.

36. Ryan, K.J., M.D., et al., *The Belmont report: ethical principles and guidelines for the protection of human subjects of research*, in *DHEW publication ; no. (OS) 78-0012- 78-0014*. 1979, The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research: [Bethesda, Md.].

37. *PART 46—PROTECTION OF HUMAN SUBJECTS*, in *45 CFR Part 46*. 2020, United States Office of the Federal Register.

38. Menikoff, J., J. Kaneshiro, and I. Pritchard, *The Common Rule, Updated.* New England Journal of Medicine, 2017. **376**(7): p. 613-615.

39. *PART 50—PROTECTION OF HUMAN SUBJECTS*, in *21 CFR Part 50*. 2020, United States Office of the Federal Register.

40. *PART 56—INSTITUTIONAL REVIEW BOARDS*, in *21 CFR Part 56*. 2020, United States Office of the Federal Register.

41. *PART 312—INVESTIGATIONAL NEW DRUG APPLICATION*, in *21 CFR Part 312*. 2020, United States Office of the Federal Register.

42. *PART 812—INVESTIGATIONAL DEVICE EXEMPTIONS*, in *21 CFR Part 812*. 2020, United States Office of the Federal Register.

43. *PART 225—PROTECTION OF HUMAN SUBJECTS*, in *22 CFR Part 225*. 2020, United States Office of the Federal Register.

44. *§312.120 Foreign clinical studies not conducted under an IND.*, in *21 CFR Part 312.120*. 2020, United States Office of the Federal Register.

45. *PART 11—ELECTRONIC RECORDS; ELECTRONIC SIGNATURES*, in *21 CFR Part 11*. 2020, United States Office of the Federal Register.

46. *Guidance for Industry: Part 11, Electronic Records; Electronic Signatures — Scope and Application.* U.S. Department of Health and Human Services, 2003.

47. *PART 164—SECURITY AND PRIVACY*, in *45 CFR Part 164*. 2020, United States Office of the Federal Register.

48. *Foreign Relations - Protection of Human Subjects*, in *22 CFR Part 225*. 2018, Code of Federal Regulations.

49. *Claris FileMaker — Tackle any task.* [cited 2020 July 12, 2020]; Available from: https://www.claris.com/filemaker/.

50. *Claris FileMaker 19 Technical Specifications*. [cited 2020 June 28, 2020]; Available from: https://support.claris.com/s/article/FileMaker-Server-19-System-Requirements?language=en_US.

51. *360Works MirrorSync: FileMaker Sync for FileMaker Server, FileMaker Pro, and FileMaker Go on iPhone and iPad, FileMaker offline sync*. [cited 2020 June 30, 2020]; Available from: https://www.360works.com/filemaker-sync.html.

52. *FM AuditLog Pro 2.0 - 1-more-thing*. [cited 2020 June 30, 2020]; Available from: https://www.1-more-thing.com/en/products/fm-auditlog-pro-2-0/.

53. Mehta, S., et al., *Effect of iron and zinc-biofortified pearl millet consumption on growth and immune competence in children aged 12–18 months in India: study protocol for a randomised controlled trial.* BMJ Open, 2017. **7**(11).

54. Yu, E., W. Bonam, and S. Mehta, *Vitamin D Supplementation and TB*, in *U.S. National Library of Medicine*. 2013, National Institutes of Health.

55. Mehta, S., et al., *Effect of Iron/Zinc-Biofortified Pearl Millet on Growth and Immunity in Children Aged 12-18 Months in India*, in *ClinicalTrials.gov*. 2014, U.S. National Library of Medicine.

56. Mehta, S. and D. Erickson. *FeverPhone*. 2018 [cited 2018 May 1, 2018]; Available from: http://insight.cornell.edu/feverphone.

57. Ruth, C.J., et al., *An Electronic Data Capture Framework (ConnEDCt) for Global and Public Health Research: Design and Implementation.* Journal of Medical Internet Research, 2020. **22**(8).

58. Finkelstein, J., et al., *Periconceptional Surveillance for Prevention of Anemia and Birth Defects in Southern India*. 2017, World Health Organization: International Congress of Nutrition, Buenos Aires, Argentina.

59. Finkelstein, J.L., *Periconceptional Surveillance in India*, in *ClinicalTrials.gov*. 2019, U.S. National Library of Medicine.

60. Mehta, S., et al., *Effect Of Multiple Biofortified Food Crops On Micronutrient Status, Immune, And Cognitive Function Among Indian Infants*, in *ClinicalTrials.gov*. 2016, U.S. National Library of Medicine.

61. Rochford, R. and R. Newton, *ENVIRONMENTAL DETERMINANTS OF KSHV TRANSMISSION IN RURAL UGANDA*. 2020, National Institutes of Health: Uganda.

62. Silverston, L., *Health Care*, in *The Data Model Resource Book*, R.M. Elliott, Editor. 2001, John Wiley & Sons, Inc.: New York.

63. Schulz, K.F., D.G. Altman, and D. Moher, *CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials.* BMJ: British Medical Journal, 2010. **340**: p. c332.

64. Moher, D., et al., *CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials.* BMJ: British Medical Journal, 2010. **340**: p. c869.

65. *The CONSORT Flow Diagram*. 2010 [cited 2020 August 12, 2020]; Available from: http://www.consort-statement.org/consort-statement/flow-diagram.

66. *Use of Electronic Records and Electronic Signatures in Clinical Investigations Under 21 CFR Part 11 – Questions and Answers*, 2017, U.S. Department of Health and Human Services. Available from https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-electronic-records-and-electronic-signatures-clinical-investigations-under-21-cfr-part-11

VITA

[redacted]