Enabling distributed electronic research data collection for a rural Appalachian tobacco cessation study

Tara B Borlawsky,1 Omkar Lele,1 Daniel Jensen,1 Nancy E Hood,2 Mary Ellen Wewers2

ABSTRACT
Tobacco use is increasingly prevalent among vulnerable populations, such as people living in rural Appalachian communities. Owing to limited access to a reliable internet service in such settings, there is no widespread adoption of electronic data capture tools for conducting community-based research. By integrating the REDCap data collection application with a custom synchronization tool, the authors have enabled a workflow in which field research staff located throughout the Ohio Appalachian region can electronically collect and share research data. In addition to allowing the study data to be exchanged in near-real-time among the geographically distributed study staff and centralized study coordinator, the system architecture also ensures that the data are stored securely on encrypted laptops in the field and centrally behind the Ohio State University Medical Center enterprise firewall. The authors believe that this approach can be easily applied to other analogous study designs and settings.

INTRODUCTION
Tobacco use remains a significant public health problem and is increasingly prevalent among vulnerable groups.1 Adults living in Ohio counties federally designated as Appalachia have higher rates of tobacco use and tobacco-attributable diseases than those in other regions of the state.2 As such, researchers have been partnering with Appalachian communities to design, implement, and evaluate evidence-based interventions to address tobacco use and other health disparities.3–5 Electronic data capture (EDC) tools have the potential to increase research data quality through the use of required fields, branching logic, and active validation. By enabling real-time entry of electronic source data during study encounters, such tools also help to eliminate unnecessary duplication of data entry and reduce the opportunity for transcription errors.6–9 However, residents of many rural Appalachian communities have limited access to a reliable internet or cell phone service, which are traditionally required to take advantage of EDC tools. To support EDC for community-based studies in such settings, we have implemented a workflow that allows geographically distributed research staff to collect data offline and later synchronize it with a central database.

CASE DESCRIPTION
The EDC workflow described in this report was developed for a study that aims to evaluate the effectiveness of a community-based intervention in promoting long-term abstinence from tobacco, as well as examine the association between abstinence and selected individual, interpersonal, organizational, community, and societal factors among adult Appalachian tobacco users exposed to the intervention. Throughout the Ohio Appalachian region, trained field research staff members interact with community residents who are interested in participating in the research project. Personnel from the Ohio State University (OSU) Extension offices, which provide access to university resources and expertise to communities throughout the state of Ohio, are also involved with participant recruitment efforts. The field research staff members, known as lay health advisers (LHAs), are responsible for recruiting and screening potential participants, delivering the intervention, and recording information about participants based on a detailed research protocol. Other field research staff members, known as interviewers (INTs), conduct face-to-face baseline, 3-, 6-, and 12-month interviews with study participants.

Specifically, to support the data collection and exchange requirements of the above described study, our EDC workflow aims to:
1. Enable data to be collected offline and later synchronized with a central database
2. Support role-based data collection and synchronization
3. Eliminate manual manipulation of data files
4. Maintain audit logs for data monitoring and end-user technical support
5. Comply with enterprise clinical application data access and security policies

METHODS OF IMPLEMENTATION
Our workflow takes advantage of the readily available and widely adopted Research Electronic Data Capture (REDCap) system.10 REDCap provides a secure, web-based application with an intuitive interface for data entry. It has allowed our investigators to rapidly develop data collection instruments, which are codified using a data dictionary consisting of descriptions of all forms, fields, and validation rules. REDCap also offers audit trails, ad hoc reporting functionality, and an automated export mechanism to common statistical packages.

We designed and implemented a complimentary data synchronization tool using Java 1.6. The tool consists of a Java Swing-based client that research field staff members use to securely synchronize
data between a central REDCap database (REDCap ENTERPRISE) and the version of REDCap that is locally installed on their laptop (REDCap MOBILE). The communication channel during synchronization is established using a secure shell (SSH) connection over a virtual private network (VPN).

**Data access**

Since protected health information is collected during the course of this study, members of the OSU Department of Biomedical Informatics collaborated with the OSU Medical Center (OSUMC) information technology team to design a system and workflows that comply with their enterprise clinical application data access and security policies. To enable this EDC workflow, we housed REDCap ENTERPRISE behind the OSUMC enterprise firewall, encrypted each laptop, and provided each user with an RSA SecurID token and access to a VPN client.

Research staff members enter the data collected in the field via REDCap MOBILE, which is populated with a subset of the overall data dictionary that is specific to their project role. Each field staff has limited read-only access to data collected by other field staff (eg, participant contact information). In comparison, REDCap ENTERPRISE contains a comprehensive repository of all study data.

**Synchronization workflow**

Internet access is not required at the time data are entered. However, once a secure internet connection is established, the field staff members synchronize study data between REDCap MOBILE and REDCap ENTERPRISE using the custom data synchronization application. The business logic embedded within this application enables role-based synchronization. The upload process is encrypted and secure, meeting all applicable Health Insurance Portability and Accountability Act (HIPAA) and Institutional Review Board (IRB) requirements.

An overview of the data collection and synchronization workflow is shown in figure 1. The first four steps, as described below, are time-sensitive and occur within several days of each other. However, multiple steps are not typically required to occur within the same day. The research staff members enter data into REDCap MOBILE, and synchronize daily between REDCap MOBILE and REDCap ENTERPRISE.

1. The LHA recruits a new participant (by telephone or face-to-face) and enters basic eligibility criteria and contact information.
2. The central program director (PD) emails the new participant’s study identifier and initials to the corresponding INT. The INT uses the synchronization tool to import the participant’s contact information into REDCap MOBILE, and contacts them to schedule a baseline interview.
3. The INT conducts the baseline interview.
4. The PD notifies the LHA of baseline interview completion, and the LHA then contacts the participant to schedule a face-to-face visit. Before that visit, the LHA may use the synchronization tool to import selected information, to which they have read-only access, from the INT baseline interview.
5. The PD notifies the INT at 3, 6, and 12 months after each participant’s baseline interview. The INT conducts the appropriate follow-up interview.

Throughout the study, the LHAs and INTs can both enter and receive any updates to the participant’s contact information in REDCap MOBILE using the synchronization tool.

Any necessary modifications to the data collection instruments are made first in REDCap ENTERPRISE, followed by staged REDCap MOBILE updates that are either coordinated with a site visit between the PD and the research field staff, or completed via remote desktop sessions. The synchronization process is largely unaffected by these changes, with the exception of modifications to coded response options, because it does not check for missing data or validate fields against the master data dictionary.

**EXAMPLE AND OBSERVATIONS**

While REDCap has built-in data import and export functionality, our synchronization tool allows direct data exchange between the REDCap MOBILE and REDCap ENTERPRISE databases without intermediate steps. This decision was motivated by the following factors:

1. Owing to the limited computer experience of the research field staff, we aimed to simplify and streamline the data collection and synchronization process as much as possible.
2. Because the study is configured using longitudinal events, the standard REDCap import and export templates, which are not modifiable, differ. That is, the event name is included in the comma-separated REDCap export file. However, during the import process, the event is manually selected through the user interface. The manual manipulation necessary to convert between file formats would require the research field staff to interact with the study data outside of the REDCap
infrastructure, and thus presents a chance for data corruption or the introduction of bias.

3. The visual similarity between the REDCap MOBILE and REDCap ENTERPRISE user interfaces, despite on-screen indicators, has the potential to confuse the research field staff if they had to switch between them for submitting and retrieving data.

4. Because of the differing form and event requirements based on the research field staff roles, the simplified REDCap MOBILE configurations streamline the data collection process and enforce role-based policies.

Additionally, the following criteria were taken into consideration when developing the data synchronization tool. During the synchronization process, it is important to avoid potentially overwriting information. For example, the PD may have revised contact information for Participant A in REDCap ENTERPRISE, and the INT has entered new interview data for Participant B in REDCap MOBILE. When the INT transfers data for Participant B to REDCap ENTERPRISE, the old contact information for Participant A residing on REDCap MOBILE should not be automatically transferred as well. To limit the data elements transferred between REDCap MOBILE and REDCap ENTERPRISE at a given time, the synchronization tool was developed such that the research field staff must select a participant from a list that is auto-generated from REDCap MOBILE, and an event from a list that is populated with only those events that have associated completed forms for the selected participant (figure 2). It is important for the synchronization tool to limit access to participants and forms based on the role and geographical location of the research field staff. To meet this criterion, the synchronization tool was developed such that the research field staff must select an existing participant from a list that is auto-generated from REDCap MOBILE (figure 3). The logic embedded within the synchronization tool uses the REDCap MOBILE data dictionary to determine the role of the field staff and the appropriate forms from which to import data.

LESSONS LEARNED
During the course of deploying REDCap and the custom data synchronization tool in support of the previously described geographically dispersed community-based tobacco cessation study, we encountered and addressed the following challenges.

Socio-technical factors
The multidisciplinary nature of implementing this study introduced challenges such as:

- Acculturation of information technology professional staff, who almost exclusively interact with enterprise clinical applications, to community-based research tools and workflows, including applicability and practicality of strong authentication policies for remote access to research data and circumstances under which approval by an IRB is required.
- Resolving differences in terminologies, including definitions of virtual machines, strong authentication and secure communication channels, study role differentiations, and blinding methodologies.
- Competing priorities between the various stakeholders, including balancing the timelines and resource constraints associated with the deployment of REDCap, development of the custom synchronization tool, ongoing implementation of an enterprise-wide electronic medical record, and provision of secure data access. For example, when a new research field staff was hired through the College of Public Health, it took several weeks to obtain the necessary OSUMC credentials and hardware (ie, RSA SecurID token). While this process has since been streamlined through bidirectional education and workflow modifications, the investigators must still plan for a delay between employee hiring and training.
Computer literacy
Many of the field staff had limited previous experience with computers. We generated comprehensive user guides providing step-by-step instructions for accessing the system, protocols for data entry and exchange, and use of the data synchronization tool. Team members with expertise in the design and implementation of community-based research in Appalachia (NH, MEW) conducted in-person hands-on training sessions. Subsequently, the field staff members were given several weeks to acclimate to the tools, enter pilot-test data, conduct mock interviews, and exchange data. Additionally, we have implemented a multidisciplinary method of technical support by which the informatics collaborators (TB, OL, DJ) have trained the public health researchers (NH, MEW) to troubleshoot and resolve common issues.

Internet connectivity
Given the remote locations of the field staff members, in-home access to the internet is often either unreliable or too slow to support the data synchronization workflow, or provide remote support and training. As such, the field staff members are often required to use nearby public internet access points.

DISCUSSION
We have demonstrated that the combination of off-the-shelf EDC tools and a custom data synchronization application can be used to facilitate the central coordination of distributed research studies conducted in communities with limited internet access, as well as provide near-real-time exchange among field project staff members and the study coordinator. Although the current custom data synchronization application is specific to the described tobacco cessation study and REDCap application, future work includes modifications to allow the associated data collection workflow to be easily adopted by other community-based studies and adapted for use with other EDC tools. Additionally, the case study described in this paper has been provided to the REDCap Consortium to help guide their development of a built-in module for asynchronous data collection.

Acknowledgments We acknowledge the contributions made by the local field research staff members who have implemented the study.

Funding This work was supported by UL1 RR025755 and R01 CA129771.

Competing interests None.

Ethics approval This study was conducted with the approval of the Ohio State University Biomedical Sciences Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES