Research Electronic Data Capture (REDCap) for population-based data collection in Low- and Middle-Income Countries: Opportunities, Challenges, and Solutions

Ha Thanh Le, Dung Viet Tien Vu, Anh Ngoc Thi Nguyen, Hang Thi Tran, Tan Viet Nguyen, Thao Phuong Tran, Aria Kekalih, Samita Rijal, Dewi Friska, Raph L Hamers, Abhilasha Karkey, Mary Chambers, Jennifer Ilo Van Nuil, SPEAR and CoAct team, Sonia Lewycka

Submitted to: Journal of Medical Internet Research
on: August 15, 2024

Table of Contents

Research Electronic Data Capture (REDCap) for population-based data collection in Low- and Middle-Income Countries: Opportunities, Challenges, and Solutions

Ha Thanh Le[1] MPH; Dung Viet Tien Vu[1*] PhD; Anh Ngoc Thi Nguyen[1]; Hang Thi Tran[1]; Tan Viet Nguyen[1]; Thao Phuong Tran[1]; Aria Kekalih[2]; Samita Rijal[3]; Dewi Friska[2]; Raph L Hamers[2, 4, 5]; Abhilasha Karkey[3, 4]; Mary Chambers[4, 6]; Jennifer Ilo Van Nuil[4, 6]; SPEAR and CoAct team[7*]; Sonia Lewycka[1, 8]

[1]Oxford University Clinical Research Unit, Hanoi, Vietnam Hanoi VN
[2]Department of Community Medicine Faculty of Medicine Universitas Indonesia Jakarta ID
[3]Oxford University Clinical Research Unit, Kathmandu, Nepal Kathmandu NP
[4]Nuffield Department of Medicine Tropical Medicine and Global Health University of Oxford Oxford GB
[5]Eijkman Oxford Clinical Research Unit, Jakarta, Indonesia Jakarta ID
[6]Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam Ho Chi Minh VN
[7]Oxford University Clinical Research Unit Hanoi VN
[8]Nuffield Department of Medicine Tropical Medicine and Global Health University of Oxford Hanoi VN
[*]these authors contributed equally

Corresponding Author:
Sonia Lewycka
Nuffield Department of Medicine
Tropical Medicine and Global Health
University of Oxford
78 Giai Phong, Dong Da
Hanoi
VN

Abstract

Background: Health research requires high-quality data, and population-based health research comes with specific opportunities and challenges. Electronic data collection can mitigate some of the challenges of working with large populations in multiple, sometimes difficult to reach, locations.

Objective: To discuss the opportunities, challenges, and solutions when using Research Electronic Data Capture (REDCap) for designing, collecting, and managing data.

Methods: We implemented two mixed methods studies combining surveys, in-depth interviews, and social media surveillance in Vietnam, Nepal, and Indonesia to understand lived experiences of the COVID-19 pandemic across three countries, and to understand knowledge and behaviours related to antibiotic use in Vietnam. In this paper, we discuss how we used REDCap to gather and manage data, and the benefits and drawbacks throughout the process.

Results: Electronic data capture using REDCap made it possible to collect data from large populations in different settings. Challenges related to working in multiple languages, unstable internet connections, and complex questionnaires with nested forms. Some data collectors lacked digital skills to comfortably use REDCap. We solved these problems through regular team meetings, training, supervision and automated error checking procedures. The main types of errors that remained were incomplete and duplicate records due to disruption during data collection. However, with immediate access to data, we were able to identify and troubleshoot these problems quickly, while data collection was still in progress. Lessons learned will be beneficial to any research team working with electronic data capture for population-based data.

Conclusions: REDCap is cost-effective, easily accessible, and has comprehensive functionality that allows for confidential, secure interactions with participants and robust data management.

(JMIR Preprints 15/08/2024:65377)
DOI: https://doi.org/10.2196/preprints.65377

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Research Electronic Data Capture (REDCap) for population-based data collection in Low- and Middle-Income Countries: Opportunities, Challenges, and Solutions

Authors
Ha Thanh Le[1], Dung Viet Tien Vu[1], Anh Ngoc Thi Nguyen[1], Hang Thị Tran [1], Tan Viet Nguyen[1], Thao Phuong Tran[1], Aria Kekalih[4], Samita Rijal[6], Dewi Friska[4], Raph L Hamers[3,4,5], Abhilasha Karkey[3,6], Mary Chambers[2,3], Jennifer Ilo Van Nuil[2,3], SPEAR and CoAct team, Sonia Lewycka[1,3]
Affiliations
1 Oxford University Clinical Research Unit, Hanoi, Vietnam
2 Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam
3 Nuffield Department of Medicine: Tropical Medicine and Global Health, University of Oxford, Oxford, UK
4 Department of Community Medicine, Faculty of Medicine, Universitas Indonesia, Jakarta, Indonesia
5 Eijkman Oxford Clinical Research Unit, Jakarta, Indonesia
6 Oxford University Clinical Research Unit, Kathmandu, Nepal
Corresponding Author:
Sonia Lewycka

Abstract

Background: Health research requires high-quality data, and population-based health research comes with specific opportunities and challenges. Electronic data collection can mitigate some of the challenges of working with large populations in multiple, sometimes difficult to reach, locations.

Objective: To discuss the opportunities, challenges, and solutions when using Research Electronic Data Capture (REDCap) for designing, collecting, and managing data.

Methods: We implemented two mixed methods studies combining surveys, in-depth interviews, and social media surveillance in Vietnam, Nepal, and Indonesia to understand lived experiences of the COVID-19 pandemic across three countries, and to understand knowledge and behaviours related to antibiotic use in Vietnam. In this paper, we discuss how we used REDCap to gather and manage data, and the benefits and drawbacks throughout the process.

Results: Electronic data capture using REDCap made it possible to collect data from large populations in different settings. Challenges related to working in multiple languages, unstable internet connections, and complex questionnaires with nested forms. Some data collectors lacked digital skills to comfortably use REDCap. We solved these problems through regular team meetings, training, supervision and automated error checking procedures. The main types of errors that remained were incomplete and duplicate records due to disruption during data collection. However, with immediate access to data, we were able to identify and troubleshoot these problems quickly, while data collection was still in progress. Lessons learned will be beneficial to any research team working with electronic data capture for population-based data.

Discussion: REDCap is cost-effective, easily accessible, and has comprehensive functionality that allows for confidential, secure interactions with participants and robust data management.

Keywords: REDCap, population-based research, strength, limitation.

Introduction

Research quality is highly dependent on appropriate design, methodology, and data collection (1). An important component of public health research quality is data accuracy and reliability (2). Growing international research, increased complexity of study designs, and stricter legislation of sharing health data including ethics, privacy, and data infrastructure make systems for high-quality data collection, management, and storage more crucial, but also more challenging than ever (3), (4). Electronic data collection methods can reduce common data entry errors and enhance data collection, storage, and analysis, compared to traditional data collection tools. (5), (6). However, when planning research, investigators need to understand the trade-offs and potential advantages and disadvantages of using different data collection methods.

Population-based data are essential in epidemiology for identifying, tracking, and addressing a wide range of health problems. Population-based data are used internationally to understand population health and social, economic, behavioural, and environmental determinants of health, and to provide valuable insights for public health decision-making and evidence-based policy-making. Unlike clinical research, population-based research often relies on surveys involving large numbers of participants sampled from multiple, sometimes difficult to reach, locations. While clinical research can make use of data on demographics, diagnoses, and treatments recorded in patient records at specific institutions, population-based research requires the collection of detailed information from representative samples of people in the community. Rather than enrolling patients as they arrive at healthcare institutions, population-based researchers must go out, find participants, and interview them.

Opportunities offered by electronic data capture

Poor and incomplete data collection reduces the usefulness of population-based data and the validity of analyses. Data quality can be affected at every stage of the research process, including data collection, management, and storage (7). Paper-based questionnaires are still widely used in population-based research, but the development of electronic data capture systems provides opportunities to reach more people, better manage data collection problems in the field in real-time, and better integrate data collection, management, and security processes.

Data collection

For both paper-based and electronic data capture, data collection tools must be well-designed and data collectors well-trained in order to minimize incorrect, biased, or missing data. Errors can happen at different times during data collection and in almost all research, even in a well-designed and controlled study (8), (9). A study developing the Data Error Criteria for retrospective studies analyzed 16 publications, and identified 2515 general errors, of which 1920 (18.9%) were blank cells, 556 (5.5%) were caused by spreadsheet mismanagement, 32 (0.3%) by transcribing errors, and 7 (0.1%) were coder-related (10). In light of this, some studies have concentrated on how to handle missing data, issues caused by missing data, and the methods to prevent or minimize missing data in medical research (11), (12). Rigorous training for everyone involved in the study before participant enrollment begins is an important step to minimize errors, and should include everything from enrolling participants to detailed instructions and practice for data collection and data entry processes (3). Electronic data capture can also reduce errors at the point of data collection and entry.

Paper-based data collection has the advantage that no technical or IT skills are needed to develop forms or enter data, no hardware or software is needed, and no internet connection to upload data to a central server. This makes tools accessible in remote settings. However, costs and resources required for printing and distributing questionnaires may be considerable, and it is easy for data collectors to make mistakes with branching logic or record unrealistic values. In contrast, electronic data capture can save time with data entry. Field restrictions and branching logic may reduce errors at the point of entry, allowing fields to be concealed or shown dependent on characteristics of survey respondents. Data collection instruments can be shorter and much more logical if related questions are grouped and only presented to respondents when relevant. Furthermore, self-completed surveys can be shared *en masse* using survey links, providing a cost-effective solution for large-scale survey deployment across countries and regions (13). However, a basic level of IT skills is required by data collectors or participants, which may limit access for some populations (e.g. low IT literacy) (6), (14).

Data management

Large-scale studies have distinct advantages in the field of etiology research because of large sample sizes and increased statistical power, but they are challenging in terms of data management and quality control. Population-based studies may use a range of data collection tools, administered by different data collectors at different times. A range of data management platforms are also available, each with different features and utilities. Research design must take into account the needs of the research, the skills of the research team, and the logistics of implementation (15). Recognizing these challenges, funders and research institutions are increasingly requesting researchers to develop data management plans that include quality control measures before, during, and after data has been collected. Data management plans must outline how each step in data gathering will be monitored so that it is possible to keep track of what happens to the data (16), (17).The advantage of paper-based data collection for data management is that a physical record of the interview is retained and can be used to verify data in the system and correct possible data entry errors. The main disadvantage is the time taken to enter and check large volumes of data. Electronic data capture allows data to be stored immediately, making it available for real-time checking and call-backs to participants to correct errors. It provides researchers and data managers with an overview of the dataset, even for a single variable, at any time, allowing them to keep control over data quality. Electronic data capture also enables simultaneous data collection and management in multiple languages using a single tool and database, which is an advantage for large multi-country surveys. Furthermore, many data capture platforms have integrated functionality that allows automatic reports to be generated in real-time, as data collection is ongoing.

Data storage and sharing

There are many different rules and guidelines regarding how long data and records should be maintained, since longer-term and permanent retention of research outputs is necessary as long as possible after publication (18), (19). When working with large amounts of data, data storage capacity may be a challenge. Paper-based records may be bulky, especially for large surveys, and require adequate physical space to retain them. They must be protected from damage by water, insects, and fire. Well-organised and maintained filing systems are required to easily retrieve specific documents. Electronic data capture allows data to be immediately uploaded to a central server, where regular back-ups can be made to avoid data loss. However, sufficient data storage capacity may be needed for large surveys. In order to strengthen scientific transparency, and increase value-for-money of data collected, it is increasingly recommended by journals and research funding institutions to share data and project documents as widely as possible. Fortunately, there are several features of electronic data collection systems that can as automatically create data dictionaries, codebooks, and forms, which facilitates the understanding and interpretability of the data and the research project in general (16).

Security

Research ethics incorporate the ideals of beneficence, justice, protection of individual's identity, and dignity (20), (21). Data must be processed in a manner that ensures the proper confidentiality of personal and identifiable information, including protection against unintentional loss, destruction, or damage, using appropriate technical or organizational measures (22), (23). When it is absolutely necessary to be kept in a data file, personal data should be encrypted (24). Research needs secure data management tools/systems that can tackle this issue. Hard copies of data should be stored in locked rooms or cupboards that are only accessible by the study team. Access to these is easily restricted, and it is unlikely that copies of physical documents will be inadvertently made and shared. Electronic data likewise, should only be accessible by authorized members of the study team. Institutions should utilize a user authentication system and an advanced encryption algorithm to ensure that data remains secure. It may be difficult to ensure data security for data collected with traditional tools such as paper or spreadsheets (25). In the building and developing process, electronic data collection systems are designed to be consistent with different information security standards worldwide, which have some features such as user right, data access group, record lock, to improve the procedures for setting up secure data collection (26).

The REDCap platform is a web-based application for building and managing online databases and surveys that supports several types of research, such as cross-sectional studies, clinical trials, retrospective studies, cohort studies, and many more (26), (27). The utility of REDCap goes beyond data collection because it includes a number of tools that are advantageous in the research environment, such as interview scheduling and customized data analysis. It is also freely available, making it a good option for research centers to use while conducting epidemiological investigations (6). REDCap is compliant with FISMA (Federal Information Security Management Act), GDPR (General Data Protection Regulation), HIPAA (Health Insurance Portability and Accountability Act), and 21 CFR Part 11 (Part 11 of Title 21 of the Code of Federal Regulations) (13). REDCap's consortium includes 7231 partners spread across 156 countries, and it has been cited in over 39000 articles (information updated in May 2024) (28). As a result, the academic community commonly uses this tool for ongoing research initiatives, ensuring a safer environment for research outcomes.

This article provides an overview of experiences using REDCap in two multi-site population-based studies: Social Science and Public Engagement Action Research (SPEAR) and Collective Action against antimicrobial resistance in Vietnam (CoAct). ***The aim of this article is to discuss the opportunities, challenges, and solutions when using REDCap for designing, collecting, and managing data.*** The lessons learned will be beneficial to research teams conducting population-based health research, as well as software development teams aiming to improve effective data management in the field.

Methods

Project settings

SPEAR project

The SPEAR study (Social Science, Public Engagement and Action Research study) was a mixed methods social science study, combining surveys, in-depth interviews, and social media surveillance to gain insights into lived experiences of the COVID-19 pandemic in Vietnam, Indonesia, and Nepal (all sites of Wellcome Africa Asia Program's Oxford University Clinical Research Unit (OUCRU)) (29). Three surveys, two in Phase 1 and one in Phase 2, were conducted in 13 districts in a range or urban, rural, and remote settings across the three countries.

Phase 1: Aimed to explore the experiences and impact of COVID-19 on healthcare workers and health-related staffs, and vulnerable communities.

Phase 2: Aimed to explore themes around the acceptance and accessibility of vaccines.

In Phase 1 healthcare workers (HCWs) and health-related staff in clinical and community settings were targeted, including physicians, nurses, pharmacists, laboratory scientists, community health-workers, ambulance drivers, cleaners, administration staff, and other health professionals. In the Phase 1 community survey and Phase 2 vaccine survey, general populations and vulnerable communities were targeted. These included specific populations, such as people in quarantine areas, recovered patients, elderly, new mothers, Tuberculosis and Hepatitis C patients, and ethnic minority groups.

Throughout the project, communications between the teams at all three sites (Vietnam, Nepal, and Indonesia) occurred weekly via Zoom. This involved discussions around the initial project setup, development of data collection forms and processes, and data quality checks.

CoAct project

The Collective Action against AMR (CoAct) project is also a mixed methods study to evaluate the impact of educational and participatory learning and action interventions in communities on antibiotic prescribing, antibiotic use, and antibiotic resistance. A baseline household survey was conducted in 64 rural communes from three districts in Nam Dinh province, northern Vietnam – Nghia Hung, Xuan Truong, and Hai Hau. The target population was a representative sample of the normal residents of 64 communes who were willing and able to give informed consent for participation.

Tool/survey development

For both studies we implemented our research data collection using REDCap to provide security, privacy, and confidentiality to all participants. We set up REDCap for the two projects on a dedicated REDCap server in Hanoi, Vietnam, thus, all data was received in a central location. For each study we established a working group to design the REDCap data collection forms, test, pilot, update, and review changes.

Our standard questionnaires for the three SPEAR surveys were designed with country-specific adaptations and translation into local languages. We used the multilingual hook package in REDCap for translating every variable into four languages (English, Nepali, Vietnamese, and Bahasa) without the need to create a new survey for each country. For the CoAct study, the entire survey was conducted in Vietnamese. The data collection tool consisted of eight instruments covering general and specific information (demographic, socioeconomic, social capital, knowledge of antibiotics and antibiotic resistance, birth history, immunization, illness and treatment). Instruments two (household member), six (child nutrition and vaccination history), and seven (household member illnesses) were repeat instruments for entering multiple members in a household.

The surveys were tested and piloted at all sites and changes were implemented before initiation of data collection. Data access groups (DAGs) were set up to organize the data for participating sites for SPEAR project. In that way, local researchers could only work on and access the project from their site, which facilitated subsequent data management. For CoAct, we did not have DAGs because we data was collected in one location, though a variable was included to identify districts. After revisions and updates were finished, surveys were moved to “production mode” and the data collection period started.

Data collection/deployment

For both SPEAR and CoAct, data collection was mainly performed using the web version rather than the REDCap mobile app. For SPEAR surveys, we were able to obtain consent electronically. Information about the study was provided on the first page, and participants were asked to click a link stating that they agree to participate, and this allowed access to the survey questions. For the CoAct survey, we first created a supplementary screening survey to obtain the consent form from every selected household before implementing the main survey.

SPEAR: Data collection for Phase 1 was between June 2020 and December 2021. Data collection for Phase 2 was between September 2021 and August 2022. Some surveys were self-completed, where participants had access to an electronic device and internet. However, some interviews were conducted in person or over the phone, where local internet access was unreliable, and depending on the public health restrictions in place in the area at the time.

Phase 1: Links to online surveys and QR codes were posted on institutional websites and shared through professional networks and social media channels, including Facebook and institutional websites, in Nepal and Vietnam, to reach as wide an audience as possible. We sought permission to contact discharged COVID-19 patients and people within or discharged from quarantine centers, in the hospital sites managing patients and quarantine centers. These participants were contacted by phone. We used our extensive networks within the healthcare systems and communities in each country to identify participants without internet access to take part in telephone interviews. For targeted healthcare workers interviews, we randomly sampled participants from staff lists at selected hospitals and health centers linked to partner institutions, using computer-generated random numbers and stratified by department and sex. For targeted community interviews, we used random sampled participants from household listing obtained from our local community-level partners where possible, as well as purposive sampling.

Phase 2: We first targeted participants from the community survey in Phase 1 for whom we had contact phone number or email address. We then randomly selected additional participants from household lists using computer-generated random numbers, in order to reach our minimum sample size for each setting (400 per country per survey). We also posted links to the survey online to be shared through social media, including Facebook and institutional websites, and professional networks.

CoAct: The survey was a cross-sectional household survey forming the baseline of a cluster randomised trial. Data collection was completed district by district: Nghia Hung in April 2022, Hai Hau in April 2023, and Xuan Truong in May 2023. A list of communes with active Community Health Centers (CHCs) was compiled by province-level officials. Sixty-four communes meeting eligibility criteria were chosen for the study. In each community, we randomly sampled approximately 195 people for each survey. Assuming 10% refusal, we had 175 per cluster and 11,200 in total. One adult household member who was identified as the main care-giver was invited to answer for all other household members. We trained 64 local interviewers, who were mainly community health workers. They used the web-based version of REDCap on internet tablets, to interview and support participants throughout the interviewing process. Some data collectors used paper questionnaires as they were unfamiliar with digital technology. We originally decided to use the REDCap mobile app for collecting data, however, there were some disadvantages with the app: (1) although the app could display survey questions in local languages, some buttons, options, and system messages could only be displayed in English; (2) we sometimes made changes to the questionnaire, which could be updated instantly on the web version but which data collectors would have to update manually from the app on their device; and (3) data was not uploaded immediately and automatically onto the server but data collectors would have to upload manually from the app on their device. Given that all study areas had mobile network coverage, and many data collectors lacked digital literacy skills, we decided to use the web-based version instead of the mobile app. Unlike the SPEAR survey, for some participants, the CoAct survey was not completed at one time. For households with children under five, data collectors needed to complete the first part of the questionnaire on the first visit, then finish the questionnaire when they came back to collect children's samples. The level of digital literacy among data collectors was not sufficient for them to easily navigate and find records in REDCap, and it would be complicated to manage permissions for 64 user accounts. Instead, re-entering the survey later for households with children under five or following internet disconnection, was done by supervisors sending survey links and return codes to the data collector's email. This allowed them to save their progress and resume the survey where they left off.

Data storage and management

SPEAR: For participants who filled in the survey online, their responses were routed directly into a REDCap server hosted by OUCRU. For participants who completed a phone or interview administered survey, responses were entered into REDCap directly (if the interviewers had access to internet) or documented on paper and then entered into REDCap from the paper form. All laptops and devices used for data collection and data storage were encrypted and password protected. Email addresses and phone numbers for future contact and linkage were stored separately from the survey responses.

CoAct: Data were collected using internet tablets, and data were entered using the REDCap web platform. Data were uploaded immediately to a secure REDCap server hosted by OUCRU. All devices used for data collection and storage were encrypted and password protected. Data monitoring and quality control measures were implemented in real-time using a collaborative approach. Every week, our data manager at OUCRU Hanoi ran an automated check (type of variables, length of the information, range of values, duplication records, coherence between information, etc.) by using Data Exports, Reports and Stats feature to make the survey reports for quality control purposes. When an issue was identified, the data manager raised a query and notified the site coordinator. Data quality reports were updated whenever new issues for monitoring and checking were identified.

On the other hand, the site coordinator regularly used automated reports and arranged a meeting or contacted the data collection team to resolve queries and provide missing or corrected data. Queries were later reviewed and closed by the data manager. In case problems were experienced by interviewers during data collection, these were reported to site coordinators and data managers to resolve. Weekly data backups were performed to maintain up-to-date datasets.

Ethical approvals

Both studies received Institutional Review Board (IRB) approvals from the Oxford Tropical Research Ethics Committee (OxTREC), as well as the National Hospital for Tropical Diseases Ethics Committee (Hanoi, Vietnam), Hospital for Tropical Diseases Ethics Committee (Ho Chi Minh City, Vietnam), Ethics Committee of Nepal Health Research Council (Kathmandu, Nepal), Patan Hospital Ethics Committee (Kathmandu, Nepal), and Ethics Committee of the Faculty of Medicine, University of Indonesia (Jakarta, Indonesia). In addition, we obtained local government permission as required by the local regulations in each project.

Results

Data collection

As of December 2022, the SPEAR study had completed data collection across 13 sites in three countries. Three surveys were completed, and data was obtained. In phase 1, after data cleaning, the community survey collected 1825 records overall, including 809 records from Indonesia, 504 records from Nepal, and 512 records from Vietnam. Meanwhile, the healthcare worker survey collected 2321 in total, with 487 Indonesia records, 537 Nepal records, and 1297 Vietnam records. In phase 2, a total of 1915 records were collected, of which 891 were from Indonesia, 405 from Nepal, and 619 from Vietnam. With CoAct study, our REDCap system was use to collect data from 3237 households in Hai Hau, Nghia Hung, and Xuan Truong districts in Nam Dinh province.

Strengths

REDCap is a user-friendly software without requirements of knowing programing to set up a database or project. Implementation of REDCap data capture allowed our researchers to gather data easily. We were able to quickly obtain large, diverse datasets without having to travel across nations, which was a great advantage, especially during the COVID-19 pandemic. Our labour force requirements and financial expenditures for the data collection procedures were greatly lowered by this solution. Surveys could be developed in multiple languages, including error messages. We were able to collaborate across multiple countries during remote working periods, and the flexibility of REDCap allowed us to make updates and corrections to problems identified in data collection tools even once data collection was in progress, without having to re-print many paper questionnaires.

The screening and consent processes could be integrated with data collection, allowing for collection and storage of meta-data including refusals. Real-time monitoring, data checking, and correction was possible. Data collectors could check data immediately after collection, which minimized errors going unnoticed. It also provided easy exports so users were in control of their data at any time. Additionally, REDCap automatically generated a codebook and data dictionary, which provided good documentation on the meaning of the variables and made collaborative work analysing data easier. Fields in REDCap can be marked as identifiable and the data manager had the option to de-identify the data during export. This function is extremely valuable for datasets including personal information such as name, address, health insurance number, allowing us to share those datasets without being afraid of disclosing personally identifiable information. Table 1 summarizes the advantages and disadvantages of electronic data capture compared to paper-based data collection that were identified during implementation of the two projects.

Table 1: Advantages and disadvantages of electronic data capture in comparison with paper-based data capture

Paper-based data capture	**Electronic data capture**
Data collection	
Advantages	
- No IT skills needed to develop forms. - No IT skills needed to use data collection tools. - Data can be collected in a variety of formats (typed, handwritten, etc.). - Data collection tools are accessible without the requirement for hardware or software. - No need for electricity or internet. - Data can be collected without reliable technology infrastructure.	- There are many options for free data capture software, particularly for educational institutions. - Data entry at the point of data collection saves time - Field restrictions and branching logic can avoid common problems at point of data entry. - Having mandatory data fields reduces the possibility of missing important questions or sections. - Surveys can be self-completed, thus saving data collection costs. - Survey links can easily be shared *en masse* for large-scale survey deployment for multiple sites and institutions across countries and regions. - Mobile apps allow for flexible data entry particularly in remote areas without internet connection.
Disadvantages	
- Cost and resources required for printing and distributing questionnaires. - It is easy for data collectors to make mistakes following branching logic or recording values outside realistic ranges. - Risk of missing data due to illegible writing, misplaced forms, or incomplete responses.	- Cost of devices (such as tablet or smartphone). - Needs trained and experienced IT personnel to develop forms, and maintain electronic devices. - Internet connectivity is required to transmit data to central server. - Requires basic level of IT skills to use by participant/interviewer. - Need for electricity and internet connection may limit use in some

	areas.
Data management	
Advantages	
- Data entry errors can be verified by checking the original paper record.	- Data can be collected and entered in real-time, reducing the risk of errors and missing data. - Data are stored immediately, minimizing data loss, and available for real-time verification. - Ensures timely tracking of activities. - Easy to edit and delete data if errors are made or updates are required. - Multi-language management is available where a default language for the project can be chosen, and then new languages added later. - Many data capture software platforms have functions for easily viewing dataset reports using tables and graphs. - Advanced analytics and visualization tools for faster and more efficient data processing. - Easier to share platforms and data and collaborate with others remotely.
Disadvantages	
- Time-consuming to enter and check large amounts of data.	- Data entry errors cannot be verified with an original record. - Electronic devices may be damaged or malfunction, leading to data loss or corruption.
Data storage	
Advantages	
- No technical skills required to access and check paper-based records.	- Data is immediately stored and can be backed up to avoid data loss. - It is easy to search digital records.
Disadvantages	
- Storage of paper questionnaires can be bulky, especially for large surveys - Well-maintained filing systems are needed to retrieve specific documents. - Paper can easily be stolen, damaged, or destroyed with incorrect storage.	- Sufficient data storage capacity may be needed for large datasets. - Risk of data breaches or hacking can compromise data security.
Security	
Advantages	
- Physical access to raw data storage units can easily be monitored.	- Access to data can be restricted from the point of data collection.

- Unlikely that copies of physical documents will be inadvertently made and shared.	- Better security measures, including encryption and back-up systems.
Disadvantages	
- It is difficult to ensure data security of paper questionnaires in transit from collection to storage facilities - Once physical data is entered into a database management system it may still be inadvertently copied and shared	- Data may be inadvertently copied and shared.

Challenges and solutions

We encountered several challenges using REDCap for data collection. Some of the issues were shared between the two surveys, others were unique to each study. Problems and proposed solutions are summarised in table 2.

Table 2: Challenges and solutions for REDCap electronic data capture during SPEAR and CoAct survey implementation

Challenges	**Solutions**
Tool development	
For SPEAR:	
- Setting the multi-language tool with different answer options for some country-specific questions, led to branching logic conflicts.	- Participants had to choose the language when starting the survey to be presented with the correct options. Questions with different country-specific responses were recorded in different variables, which were merged during data processing. - Careful piloting in each language/site before uploading to REDCap to minimized the need for later revisions to the REDCap tool.
- It was time-consuming to develop the tools and make changes due to the complexity of the questionnaire and the separate time zones for collaborating across three sites.	- We held weekly Zoom meetings across three countries to share feedback from survey development, testing and piloting.
For CoAct:	
- A lengthy questionnaire with complex logical connections between instruments required back-and-forth testing between survey creators and community data collectors.	- Careful piloting, training, close supervision, and regular, routine data checks to ensure completeness and logical responses.
- REDCap does not allow data to be collected with relational data structure for nested forms – e.g. members within households. Unique	- We created separate REDCap instruments relating to nested forms (households, household members, children, illness episodes). These

primary identifiers (keys) are automatically created as record IDs, but difficult and meaningless for data collectors to work with.	could be identified by Record ID, instrument ID, and instance ID. So that data collectors could select a specific household member in nested forms, we created a separate field and used an SQL command to display member household names in a dropdown list wich they could choose from. This meant that data collectors did not need to reenter the member’s name to link the records, thus avoiding typing errors.
- To reduce survey time, we wanted to randomly present only two questions out of a set of seven questions to each household.	- We created a random combination of questions for each household by group outside of REDCap, then we used branching logic to show only the questions that were pre-assigned for that household.
Data collection	
For SPEAR:	
- Some real data was mistakenly entered before production mode and this made data cleaning time-consuming.	- We downloaded the data and reimported it to REDCap once production mode had started. - We regularly scheduled communications between the teams at all three sites to discuss the initial project setup, data quality checks, and verification of data collection. - Future efforts should be clear in communication with team members when making the transition to production mode.
- Incomplete and duplicate records due to internet disruption during data collection. – i.e. respondents started the survey, got cut-off, and started again with a new, duplicate record.	- We added a time stamp and field to indicate when the survey was actually completed, rather than cut-off midway.
For CoAct:	
- Some data collectors were not comfortable using electronic devices and opted to use paper surveys which were entered into REDCap later.	- Carefully planned REDCap training to provide basic working knowledge hands-on experience using the software, ensuring enough training time user-friendly materials. - The option for dual modes of data collection was made possible. Supervisors checked and entered paper surveys into REDCap.
- Incomplete and duplicate records due	- If data collectors were cut-off during

to internet disruption during data collection. – i.e. respondents started the survey, got cut-off, and started again with a new, duplicate record.	the interview, they had to ask the supervisor to send a link to re-enter the survey. - We created automated REDCap reports to display all records entered for each commune, so that supervisors could use these to monitor data collection progress and check for errors. - Ongoing technical support and real-time data checking to identify and remove incomplete and duplicate records. - We also added a time stamp and field to indicate when the survey was actually completed, rather than cut-off midway.
- Data collectors lacked sufficient digital literacy to navigate the REDCap system or mobile app. To simplify the process, they entered REDCap through the webform, rather than logging in with an account. This meant that they could not directly re-enter or check records. Following internet disruption or to complete a later section, they had to request a link email to re-enter. This was confusing and complicated because it had to be done manually case-by-case.	- We developed a standardized data collection SOP to provide guidance for managing data and guaranteeing that all data was gathered, checked, and properly evaluated. - Supervisors were trained to use the REDCap system to check data, identify records, and send survey links to data collectors.
- Some error messages could not be displayed in languages apart from English.	- We added footnotes or pop-ups for more information in the local language and to explain questions.
Data management	
For SPEAR:	
- Because many surveys were self-completed, we had difficulty distinguishing between incomplete surveys and surveys with a lot of skipped questions.	- We used the variables for time-stamp and completed survey status to identify incomplete surveys. - During data cleaning we removed any remaining records with less than 50% of the total questions answered.
For CoAct:	
- Managing incomplete and duplicate records for some households.	- We used the time stamp and field to identify completed surveys. - Careful checking by supervisors and cleaning by the data team to remove incomplete and duplicate records after data collection.

- Missing and incorrect data were identified during automated quality control checks.	- Phone numbers were collected during data screening, and call-backs were made to 20% of households to check data quality. - Phone calls were also made to check any missing or incorrect data found through error reports. - We collected GPS data from the tablets for location tracking and used this to identify data collectors who needed extra supervision to ensure interviews were done.
- Because a record ID rather than household number was used as the primary key, duplicate records for households were possible. It was then difficult to identify duplicate records for the same household, often created when multiple attempts were made to start the survey.	- We added an extra field for supervisors to indicate the correct record when multiple entry attempts were made for the same household. These were reported weekly for checking and correction.
- REDCap has many additional modules, functions, and solutions that we were not familiar with. Some of these required IT skills and time to solve.	- We used the REDCap community to seek advice and ideas to solve problems. - For future research, a dedicated person with IT and data management skills would improve capacity to modify tools and reports efficiently.

Table 3: Suggestions for future REDCap features

1. The mobile app has modules to translate the interface into multiple languages, however, some buttons, options, and system messages can only be displayed in English. These need also to be translatable into local languages. 2. Management of nested forms is complicated. More efficient management and storage of relational data structures would be beneficial for population-based surveys or longitudinal studies with repeated visits. 3. REDCap automatically creates a record ID as the primary key. A secondary unique field can be added as a constraint to make sure that a record does not already exist. However, this must be a text field. To avoid typing errors, we wanted to use a dropdown list of preselected households from our sampling frame to check if a record for this household already existed in the database. But it was not possible to set this non-text field as a secondary key. 4. A codebook can automatically be created and downloaded, however, this is not currently in table format. All values for a question are contained within one cell, rather than as separate rows.

Discussion

Advantages of using REDCap

Utilization of REDCap has significantly increased in epidemiology, particularly in studies undertaken in low-and middle-income countries (30). The key advantage of using REDCap software is reduction of research costs and improved efficiency. There is no charge for REDCap for use by educational institutions, which is an advantage compared to some other software. REDCap provides their tools freely and has a lot of effective functions, which make global research collaboration efficient. Examples of efficiency have been reported in several cohort studies and large international collaborations illustrating that there were reductions in time and costs associated with interviewing, travelling, printing documents, and communicating between sites (6), (31), (32). As compared to paper-based research methods, these savings have led to a lower cost per interview and overall expenditure. Financial advantages increase with the size of the research population (33).

In our two studies, REDCap enabled us to collect data from large populations across a range of urban, rural, and remote settings. We were able to control data entry errors and increase data quality by using a range of functions in REDCap, including field validation, question requirements, and branching logic. Data inconsistencies in research tend to occur quite commonly for several reasons, mostly due to errors in questionnaire or interviewer's inaccuracies (34). However, with the help of REDCap's function for field restrictions, it was possible to minimize data recording errors when conducting interviews. Additionally, branching logic could be applied so that given a participant response to one question, data collectors would automatically be directed to the correct follow-up question, or skip subsequent questions that were not relevant. Checking this branching logic took time during set-up, but it was an important step to have a well-prepared survey, and saved time later for error checking. Some errors or inconsistencies still arose, particularly with managing duplicate or incomplete records. The only option to resolve these issues was contacting the interviewee to verify the information given. With immediate access to electronic data, it was possible to automate error and consistency checking, and identify problems. Research supervisors were thus able to identify errors during the data collection period and fix them soon after the interview. This contrasts with paper-based questionnaires, where it can take months to achieve a closed database since questionnaires must be gathered, checked manually, and entered.

A vital element of REDCap tool development and implementation is effective and close collaboration between sites implementing projects. During the development period of SPEAR and CoAct studies, weekly meetings were held to develop the tools, transfer to REDCap, test the system, pilot, and revise. Weekly meetings continued during implementation periods to identify and troubleshoot common problems and collect status updates. One frequently discussed issue was the occurrence of multi-language and branching logic errors, which appeared as notifications at the beginning of the web survey. Continuous feedback allowed data collectors to identify and report inaccuracies, enabling data managers to resolve them promptly. This iterative process of feedback and correction facilitated the enhancement of the tools and the overall approach.

Time saving was another benefit of the use of electronic data capture. REDCap features such as automated survey invitations and reminders streamlined the process of reaching participants without the need for in-person contact. Additionally, its built-in data validation and real-time error checking reduced the need for extensive data cleaning, thereby speeding up the analysis phase. Our findings strongly support the idea that REDCap is a good choice for developing and deploying population-based data collection, especially during the COVID-19 period when conducting face-to-face interviews was difficult due to social-distancing measures in place (35). Overall, REDCaps capabilities greatly enhanced the efficiency and effectiveness of our research efforts during project implementation period.

Disadvantages

We chose REDCap because of the relatively simple interface, meaning that it was easy to get started without specific programming skills. However, preparing the REDCap server for data collection and developing survey tools with multiple instruments and complex branching logic was not such a straightforward process. It required a professional with sufficient competence to install and configure the necessary tools and tailor the REDCap environment using some SQL programming. Thus, the project team needs to have an IT expert to develop and test the tools, as well as sufficient technical expertise among the supervisory team to troubleshoot, communicate issues to the IT expert, and manage data checking, editing, and correction within REDCap. This would speed up the data collection and verification processes. In populations with low digital literacy, thorough training, piloting, easy to follow guidance materials, and close supervision is needed.

Conclusions

For both SPEAR and CoAct studies, our research involved collaboration with multiple partners and data collection in different locations. The entire data management process, which included gathering, storing, and monitoring the data, was made easier with REDCap. Some features were not straightforward for staffs and participants with basic computer literacy skills. These problems could be rectified with carefully designed training, supervision, and dedicated IT support. More intensive piloting and supervision was needed to familiarize data collectors with the REDCap tool and troubleshoot problems than might have been needed with a paper-based survey, especially in the early phases of data collection. But the advantages of having branching logic to navigate complex surveys, and electronic data immediately available for monitoring and checking were huge, and costs and time were saved from data entry and processing.

Author Contributions: SL, MC, JIVN designed the studies and directed the projects. TVN, DVTV, HTL designed the data collection tools. ANTN, THT, TPT led the data collection process. AK, SR, DF, RLH, AK provided insightful information to the studies. The manuscript was first drafted by HTL and thoroughly revised by SL, TPT, DVTV, MC, and JIVN. All authours approved the final version of the manuscript and are accountable for all aspects of the work in ensuring its accuracy and integrity.

Funding: This work was supported by WellcomeTrust, UK. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We would like to acknowledge all the participants who gave their time to take part in SPEAR and CoAct study. We also would like to acknowledge all of our collaborators across the study sites.

Conflicts of Interest

The authors declare that there are no conflicts of interest in this article.

Abbreviations

REDCap: Research Electronic Data Capture
FISMA: Federal Information Security Management Act
GDPR: General Data Protection Regulation
HIPAA: Health Insurance Portability and Accountability Act
SPEAR: Social Science, Public Engagement and Action Research study
CoAct: Collective Action against antimicrobial resistance in Vietnam
OUCRU: Oxford University Clinical Research Unit

References

1. Haug A, Zachariassen F, Liempd D van. The costs of poor data quality. Journal of Industrial Engineering and Management. 2011 Jul 4;4(2):168–93.

2. Choi BCK. The Past, Present, and Future of Public Health Surveillance. Scientifica (Cairo). 2012;2012:875253.

3. de Kok JWTM, de la Hoz MÁA, de Jong Y, Brokke V, Elbers PWG, Thoral P, et al. A guide to sharing open healthcare data under the General Data Protection Regulation. Sci Data. 2023 Jun 24;10:404.

4. Van Bulck L, Wampers M, Moons P. Research Electronic Data Capture (REDCap): tackling data collection, management, storage, and privacy challenges. Eur J Cardiovasc Nurs. 2022 Jan 11;21(1):85–91.

5. Jibb LA, Khan JS, Seth P, Lalloo C, Mulrooney L, Nicholson K, et al. Electronic Data Capture Versus Conventional Data Collection Methods in Clinical Pain Studies: Systematic Review and Meta-Analysis. J Med Internet Res. 2020 Jun 16;22(6):e16480.

6. Barros AJD. Electronic data collection in epidemiological research. The use of REDCap in the Pelotas birth cohorts. Appl Clin Inform. 2016 Jul 13;7(3):672–81.

7. Bernardi FA, Alves D, Crepaldi N, Yamada DB, Lima VC, Rijo R. Data Quality in Health Research: Integrative Literature Review. J Med Internet Res. 2023 Oct 31;25:e41446.

8. Kang H. The prevention and handling of the missing data. Korean J Anesthesiol. 2013 May;64(5):402–6.

9. Dhudasia MB, Grundmeier RW, Mukhopadhyay S. Essentials of Data Management: An Overview. Pediatr Res. 2023 Jan;93(1):2–3.

10. Buczek L, Azar F, Bauzon J, Batra K, Murphy C, Wahi-Gururaj S. The Data Error Criteria (DEC) for retrospective studies: development and preliminary application. J Investig Med. 2023 Apr;71(4):448–54.

11. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. N Engl J Med. 2012 Oct 4;367(14):1355–60.

12. O'Neill RT, Temple R. The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. Clin Pharmacol Ther. 2012 Mar;91(3):550–4.

13. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009 Apr;42(2):377–81.

14. Bangdiwala AS, Boulware DR. Technical procedures and REDCap tools for internet-based clinical trials. Contemp Clin Trials. 2022 Mar;114:106660.

15. Cené CW, Haymore LB, Enga Z, Sallah SH, Ritchwood T, Wynn M, et al. Data Collection and Management in Community Engaged Research: Lessons Learned From Two Community-Based Participatory Research Partnerships. Prog Community Health Partnersh. 2015;9(3):413–22.

16. Meyer J, Ostrzinski S, Fredrich D, Havemann C, Krafczyk J, Hoffmann W. Efficient data management in a large-scale epidemiology research project. Comput Methods Programs Biomed. 2012 Sep;107(3):425–35.

17. Ramsey I, Corsini N, Hutchinson A, Marker J, Eckert M. Challenges and opportunities for using population health data to investigate cancer survivors' quality of life in Australia. Qual Life Res. 2022 Oct 1;31(10):2977–83.

18. Management of Data and Information in Research: A guide supporting the Australian Code for the Responsible Conduct of Research.

19. Ducato R. Data protection, scientific research, and the role of information. Computer Law & Security Review. 2020 Jul 1;37:105412.

20. Ke Z, Yongzhen L. Research on Internet data security and privacy protection. J Phys: Conf Ser. 2021 Aug 1;2005(1):012004.

21. Wang R. Research on Data Security Technology Based on Cloud Storage. Procedia Engineering. 2017 Jan 1;174:1340–55.

22. Petrova E, Dewing J, Camilleri M. Confidentiality in participatory research: Challenges from one study. Nurs Ethics. 2016 Jun;23(4):442–54.

23. Bertino E. Data Security – Challenges and Research Opportunities. In: Jonker W, Petković M, editors. Secure Data Management [Internet]. Cham: Springer International Publishing; 2014 [cited 2024 Jun 3]. p. 9–13. (Lecture Notes in Computer Science; vol. 8425). Available from: https://link.springer.com/10.1007/978-3-319-06811-4_2

24. Rumbold JMM, Pierscionek B. The Effect of the General Data Protection Regulation on Medical Research. J Med Internet Res. 2017 Feb 24;19(2):e47.

25. The importance and challenges of data sharing. Nat Nanotechnol. 2020 Feb;15(2):83–83.

26. Patridge EF, Bardyn TP. Research Electronic Data Capture (REDCap). J Med Libr Assoc. 2018 Jan;106(1):142–4.

27. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform. 2019 Jul;95:103208.

28. Consortium – REDCap [Internet]. [cited 2022 Dec 13]. Available from: https://projectredcap.org/about/consortium/

29. COVID-19 Social Science and Public ... | Wellcome Open Research [Internet]. [cited 2024 Apr 2]. Available from: https://wellcomeopenresearch.org/articles/6-352

30. Odukoya O, Nenrot D, Adelabu H, Katam N, Christian E, Holl J, et al. Application of the research electronic data capture (REDCap) system in a low- and middle income country- experiences, lessons, and challenges. Health Technol (Berl). 2021 Nov;11(6):1297–304.

31. Gesell SB, Halladay JR, Mettam LH, Sissine ME, Staplefoote-Boynton BL, Duncan PW. Using REDCap to track stakeholder engagement: A time-saving tool for PCORI-funded studies. J Clin Transl Sci. 2020 Apr;4(2):108–14.

32. Khan JS, Jibb LA, Busse JW, Gilron I, Choi S, Paul JE, et al. Electronic Versus Traditional Data Collection: A Multicenter Randomized Controlled Perioperative Pain Trial. Canadian Journal of Pain. 2019;3(2):16–25.

33. Pavlović I, Kern T, Miklavcic D. Comparison of paper-based and electronic data collection process in clinical trials: costs simulation study. Contemp Clin Trials. 2009 Jul;30(4):300–16.

34. Kyte D, Ives J, Draper H, Keeley T, Calvert M. Inconsistencies in quality of life data collection in clinical trials: a potential source of bias? Interviews with research nurses and trialists. PLoS One. 2013;8(10):e76625.

35. Nuil JIV, Friska D, Kekalih A, Bhandari AR, Bogh C, Brindle H, et al. COVID-19 Social Science and Public Engagement Action Research in Vietnam, Indonesia and Nepal (SPEAR): Protocol for a mixed methods study exploring the experiences and impacts of COVID-19 for healthcare workers and vulnerable communities [Internet]. Wellcome Open Research; 2021 [cited 2022 Dec 5]. Available from: https://wellcomeopenresearch.org/articles/6-352