German Medical Data Sciences 2023 — Science. Close to People.
R. Röhrig et al. (Eds.)

doi:10.3233/SHTI230693

Evaluating REDCap as the Central Data Collection Tool for the Hamburg City Health Study

Nils SCHÖNBECK[a,1], Yassin HUSSEIN[b], Alena HAACK[b], Axel SCHMAGER[b], Ulrike HARNEY[b], Leona TRÜBE[a], Frank ÜCKERT[a], and Karl GOTTFRIED[a]
[a] *Applied Medical Informatics, University Medical Center Eppendorf, Hamburg, Germany*
[b] *Epidemiological study center, University Medical Center Eppendorf, Hamburg, Germany*

Abstract. Introduction The collection of examination data for large clinical studies is often done with proprietary systems, which are accompanied by several disadvantages such as high cost and low flexibility. With the use of open-source tools, these disadvantages can be overcome and thereby improve data collection as well as data quality. Here we exemplary use the data collection process of the Hamburg City Health Study (HCHS), carried out at the University Medical Center Hamburg-Eppendorf (UKE). We evaluated how the recording of the examination data can be converted from an established, proprietary electronic healthcare record (EHR) system to the free-to-use Research Electronic Data Capture (REDCap) software. **Methods** For this purpose, a technical conversion of the EHR system is described first. Metafiles derived from the EHR system were used for REDCap electronic case report form (eCRF) building. The REDCap system was tested by HCHS study assistants via completion of self-developed tasks mimicking their everyday study life. Usability was quantitatively evaluated via the IBM Computer System Usability Questionnaire (CSUQ) and qualitatively assessed with a semi-structured interview. **Results** With the IBM CSUQ, the study assistants rated the usage of the basic REDCap system for HCHS examination data collection with an overall score of 4.39, which represents a medium acceptance. The interview feedback was used to formulate user stories to subsequently increase the administrative sovereignty and to conceptualize a REDCap HCHS information technology (IT) infrastructure. **Conclusion** Our work aims to serve as a template for evaluating the feasibility of a conversion from a proprietary to a free-to-use data collection tool for large clinical studies such as the HCHS. REDCap has great potential, but extensions and an integration to the current IT infrastructure are required.

Keywords. Electronic data capturing, REDCap, Hamburg City Health Study, user acceptance, infrastructure

[1] Corresponding Author: Nils Schönbeck, Applied Medical Informatics, University Medical Center Eppendorf, Martinistraße 52, 20246 Hamburg, Germany, E-mail: n.schoenbeck@uke.de.

1. Introduction

1.1. Scientific Background

Data collection is a crucial aspect of any research study. The choice of an efficient electronic data capturing (EDC) tool can have a significant impact on the accuracy and efficiency, thereby enhancing data quality [1] and reducing costs [2,3]. In recent years, there has been a growing interest in open-source data collection tools as an alternative to proprietary solutions [4]. REDCap (Research Electronic Data Capture) is a free-to-use, web-based platform for managing and collecting clinical research data by customization options through built-in software development tools. A license agreement between one´s non-profit institution and Vanderbilt is required to get codebase access [5]. The Hamburg City Health Study (HCHS) is a population-based prospective cohort study that aims to learn more about important risk and prognostic factors. The study is designed to collect detailed data on a wide range of variables that may be associated with the development and progression of chronic diseases based on 45,000 participants between 45 and 74 years old in Hamburg, Germany. For each participant, a variety of methods are used to collect data, which will be repeated at a follow-up exam six years after the first visit [6]. In this context, EDC as part of an effective information technology (IT) structure is essential for successful data collection and management in such studies [7]. REDCap is a web-based, free-to-use platform designed specifically for use in scientific studies with a broad consortium [8]. REDCap has a set of features, that make data capture, management, and delivery easier and faster. Some of the functionalities of REDCap are potentially relevant for a large research infrastructure as in the case of the HCHS: (1) procurement cost reduction due to the free-to-use character of the software, (2) flexibility in creation of electronic questionnaires for data collection, (3) various data entry options, (4) data validation and cleaning, (5) built-in user management, (6) support for various privacy standards and policies, (7) built-in version control, and (8) integration of REDCap with other tools and systems, such as Excel, SPSS or R. Furthermore, REDCap provides an application programming interface (API) which allows developers to integrate or automate REDCap functions. In addition to the advantages of open-source data collection tools, there are also some potential challenges to consider, for example the user acceptance [9] or the integration process into the given clinical and/or research IT infrastructure [10]. The evaluation of REDCap as a data collection tool for the HCHS will provide valuable insights into the potential use of it in large-scale research infrastructures. REDCap was chosen for our examination because of prior experience in the use for smaller studies at the University Medical Center Eppendorf (UKE). The results of this evaluation will be useful for researchers, data managers, and IT professionals who are considering the use of REDCap in their own studies. The use of REDCap as a central component in HCHS data processing could serve as a template for other research-based infrastructures and reduce dependence on proprietary software solutions. The implementation of REDCap as a data collection tool can help to reduce costs, increase flexibility, and improve the overall efficiency and effectiveness of data collection.

1.2. Objective of the study

In this paper, we evaluate the potential use of REDCap as a data collection tool for the Hamburg City Health Study. We assess the user acceptance for study-specific data collection tasks and the necessary integration process into the existing study IT infrastructure. With this evaluation, we aim to provide an assessment of REDCap as a tool for usage in large-scale research infrastructures. The use of REDCap as a central component in HCHS data processing could serve as a template for other research-based infrastructures and reduce dependence on proprietary software solutions.

2. Methods

2.1. HCHS- data source, data collection process and IT infrastructure

The HCHS is collecting data from four main sources: (1) Data from 18 examinations [6]. The data are collected using the proprietary hospital information system (HIS) Soarian® (Siemens), including an electronic healthcare record (EHR) system. This source includes laboratory and pathological test results. (2) Data from medical devices like Magnetic Resonance Imaging (MRI). (3) Data from 36 self-reported questionnaires [6]. (4) Secondary used variables calculated from source (1) to (3). Within the current HCHS IT infrastructure, data are collected in the hospital or external networks by multiple devices: The HIS with the laboratory information and management system (LIMS) and the radiology information system (RIS), holding the laboratory devices or the MRI device, respectively. The proprietary EHR system for data collection is also integrated in the HIS. The HCHS department network also has its own research devices for bodyplethysmography or for echocardiography (ECG) examinations. External networks are used for questionnaire data entry devices. The hospital and external networks are required to be connected to fulfill the study's needs. Finally, all data are stored in the HCHS database. An overview of the current HCHS IT infrastructure is shown in figure 2.

2.2. REDCap eCRF building

REDCap electronic case report forms (eCRFs) were created based on four HCHS metadata files which were provided by the HCHS data management team. These metadata files were semi-automatically scanned and filtered to the main required eCRF variables (variables from source (1) and (2)). This set of variables was then extracted and adapted to the REDCap structure. Subsequently, it was manually mapped to a so-called REDCap "data dictionary", which is a csv-file formatted for an upload to REDCap. An iterative adaptation step was then performed mainly to improve the visual appearance of the eCRFs, for instance through text enrichment in the descriptions of the variables, and to enhance the workflow within the created eCRFs. Over 6000 study variables are recorded per HCHS subject [6]. For this purpose, examination devices such as ultrasound devices are used, but also questionnaires. In fact, 1897 variables in 47 examination instruments were generated through direct data entry by the study assistant. The additional more than 4000 data points originate from the remaining sources ((2), (3), and (4)) and are intended to be integrated into the study database independent of the REDCap

data collection, as currently done (see infrastructure in figure 2). From all 47 instruments, three instruments showed the most significant REDCap functions for daily task completion and were therefore used for the evaluation study.

2.3. Study design and evaluation

To test the acceptance and requirements of REDCap as EDC system, the study assistant team as main users for data input were included in a two-step evaluation study. For the recruitment process, ten study assistants volunteered to perform the evaluation study. Inclusion criteria was the experience with the old system of at least 12 months. The evaluation study was carried out at the epidemiological study center of the University Medical Center Eppendorf (UKE) from September to December 2022. The study assistants independently performed the evaluation during their working hours. They were able to use their own workspace with a laptop provided for this purpose. The required instructions for conducting the evaluation were the same for all participants and provided via predefined structured worksheets. As a first evaluation step, the study assistants performed two tasks. Task one consisted of watching a self-produced, three-minute introduction video for REDCap´s basic functions. This included log in to REDCap, select a project, choose an existing or add a new record, open an instrument, save entered data and download data entry as PDF file. The study assistants had the opportunity to watch the video multiple times. The frequency of watching was documented for each assistant. As a second task, three case studies were designed to simulate the REDCap usage as realistic as possible. For each case study, a wide range of REDCap functions were considered while keeping the focus on comparability to the daily routine of a study assistant. Case study one involved a blood sampling examination with data entry and PDF file creation. Case study two involved an ECG examination with data entry and correction of false entered data. Case study three involved a dental status examination with data entry and error message handling. Each study participant received his or her own random user account to support the verification of successful task completion. For further support during the evaluation, a contact person for telephone consultation was made available. The processing time for each task was documented by the participants. To obtain a quantitative, standardized user experience feedback, the IBM Computer System Usability Questionnaire (CSUQ) [13] was used. The CSUQ has been shown to be valid and reliable for software usability evaluation [11,12]. As a second evaluation step, a semi-structured, ten-question telephone interview was offered to qualitatively specify the exact requirements for REDCap. The questions and answers were predefined by the technical experts to get feedback as standardized as possible. The study assistants were read aloud the interview questions and all answers were documented. The interview answers were manually collected, sorted, counted, and finally merged in case of semantically similar answers. Based on the qualitative user experience feedback, the study assistants' requirements for REDCap were extracted and listed in the form of user stories. For this purpose, two members from the technical team first screened the answers to the standardized questions separately and sorted them regarding to a set of possible categories. The following categories were defined and differentiated: "User interface", "Functionality", "Error handling", and "Others". The classifications of both experts were compared, restructured in an evaluation process, and formulated as user stories. In a final assessment, the list of formulated user stories was verified and prioritized by the HCHS study assistants to derive their respective importance. Each user story was given a prioritization score from 1 (highest priority) to 10 (lowest priority) by the study assistants.

For data analytics, Python 3 was used for the analysis of the HCHS eCRF, the REDCap metadata and semantic as well as for the upload of the REDCap data dictionary. Microsoft Excel and the Python package pandas were used for data analysis. The Python package seaborn was used for data visualization and draw.io and Microsoft PowerPoint for IT infrastructure design.

3. Results

3.1. REDCap test system and task completion

From all volunteering study assistants (n=10), five completed all tasks, answered every IBM CSUQ question and every semi-structured interview question. Six study assistants watched the introduction video (task one) once, two watched it twice. The mean processing time of case study one was five minutes recorded by seven users, case study two seven minutes recorded by eight users and case study three four minutes recorded by seven users (task two of evaluation study). The duration of employment at the HCHS was more than three years for five and less than twelve, but more than six months for one study assistant. Six study assistants stated to have no prior REDCap experience.

3.2. Evaluation of the REDCap efficiency using quantitative user experience feedback

For quantitative evaluation of the user satisfaction of the REDCap performance, the IBM CSUQ [13] was used. On each of the 19 questions of the questionnaire, the study assistants were asked to give feedback ranging from strongly agree (1) to strongly disagree (7). The best result received question seven ("It was easy to learn to use this system.") and 14 ("The information is effective in helping me complete my work.") with a mean of 3.43 (n=7). The worst result received question 12 ("It is easy to find the information I need.") with a mean of 5.71 (n=7). The single question scores could be summarized into an overall satisfaction score (mean of 4.39, question one to 19), a system usability score (SYSUSE; mean of 4.40, question one to eight), an information quality score (INFQUL; mean of 4.35, question nine to 15) and an interface quality score (INTERQUL; mean of 4.33, question 16 to 18). Results are shown in figure 1.

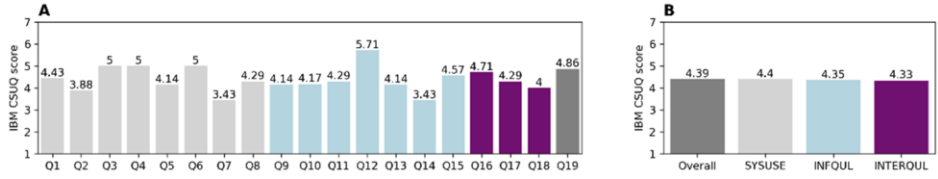

Figure 1. Distribution of single IBM computer system usability questionnaire (CSUQ) scores (A) and summarized scores (B) in quantitative user experience evaluation. (A) Single question scores are means from all given answers per question: n= 8 for question two, n= 6 for question 10 and n= 7 for all remaining questions. For IBM CSUQ questions, see [13]. Scores: (1) strongly agree, (2) largely agree, (3) agree, (4) neutral, (5) disagree, (6) largely disagree, (7) strongly disagree. Q= Question, SYSUSE= System usability, INFQUL= information quality, INTERQUL= interface quality.

3.3. Defining REDCap user requirements using qualitative user experience feedback

A semi-structured interview was used to qualitatively evaluate the use of the EHR system and REDCap for daily task completion. The requirements were formulated as user stories and verified as well as prioritized by the study assistants. Results are shown in table 1.

Table 1. Software requirements (R) formulated as user stories with prioritization by the study assistants (n=6).

Priority	User story
1	R3: As a study assistant, I would like REDCap to be able to print out entered data and laboratory results so I can hand them over to the patients.
1	R4: As a study assistant, I would like REDCap to not have to enter repetitive, identical examination entries to save time.
2	R2: As a study assistant, I would like REDCap to display patient information, findings, and laboratory results so that I have all the information I need for patient examination.
2	R8: As a study assistant, I would like REDCap to let me complete examinations and erroneous entries only when data have been entered completely and correctly, so I can be sure all entered data are correct.
3	R10: As a study assistant, I would like REDCap to not block examinations other assistants are entering data in for me, so I can work simultaneously on the same examination.
4	R6: As a study assistant, I would like REDCap to have all thematically similar examination forms organized in a folder structure, so I have a better overview of the examination forms.
5	R7: As a study assistant, I would like REDCap to have better support on suggested system adjustments, so I can use the system in an optimized and updated version.
5	R9: As a study assistant, I would like REDCap to answer all examination variables automatically by no, if the first question regarding the execution of an examination is answered by no so I can save time and effort.
6	R1: As a study assistant, I would like to see the language in REDCap standardized to German so that I can understand the system errors and warning messages better and that I can find a way to handle errors by myself.
7	R5: As a study assistant, I would like REDCap to give me the option of entering free text in the event of special incidents during the examination and to always be able to make appropriate entries, including those relating to the termination of the examination.

3.4. Conceptualized HCHS IT infrastructure

Figure 2 shows a comparison of the current IT infrastructure described in 2.1 with a proposed IT infrastructure created upon our user evaluation integrating REDCap.

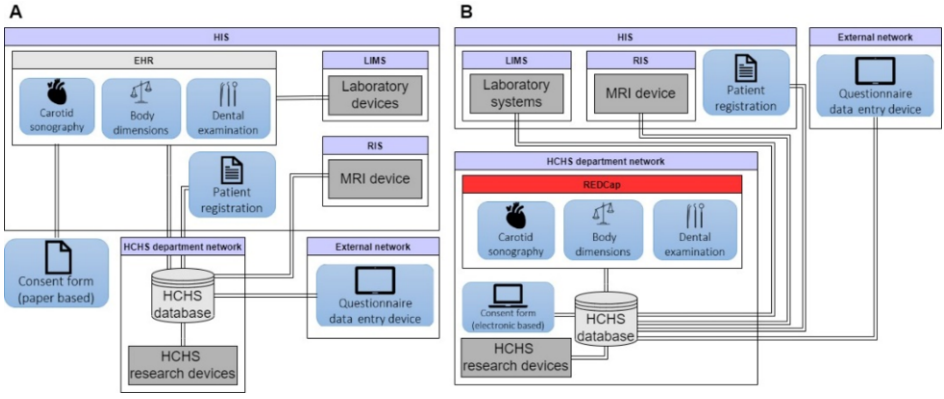

Figure 2. IT infrastructure concepts. (A) The current HCHS IT infrastructure. The HCHS database derives a large proportion of data from non-HCHS department network components (EHR, LIMS, RIS, external network). Exemplary examinations are shown. (B) Proposed IT infrastructure concept including REDCap as the central data collection tool. Comparably to the current IT infrastructure, an increased amount of data is derived directly from the HCHS department network. HIS= Hospital information system, LIMS= Laboratory information and management system, RIS= Radiology information system, EHR= Electronic health record, HCHS= Hamburg City Health Study, MRI= Magnetic resonance imaging, REDCap= Research Electronic Data Capture.

Our proposed it concept is the result of the HCHS student assistant evaluation study. This leads to user stories, which in turn have led to the proposed IT concept using REDCap as the central data collection tool instead of the proprietary EHR system. Consequently, an increasing amount of data collection can be centralized within the HCHS department network. REDCap is giving more flexibility or independence then proprietary systems, allowing more possibilities for customization and management.

4. Discussion

In this user evaluation, we found that REDCap can be successfully used to enter and collect examination data for large clinical studies such as the HCHS and thereby supplement an established, proprietary EHR system with a free-to-use EDC tool. We found that REDCap application was rated with a medium acceptance by the study assistants of the HCHS. In total, five study assistants completed all evaluation tasks and fully responded the IBM CSUQ and the interview questions. This response might explain the retrieved medium acceptance rating. There can be varying reasons for an interruption or a termination of the task completion despite a previous voluntary commitment. On the one hand, the tasks were performed by native German speakers only. Although all tasks and the REDCap eCRFs were translated into German, the basic REDCap interface is English which reportedly led to dissatisfaction. On the other hand, tasks were performed on regular workdays. Time pressure and hectic situations can also lead to dissatisfaction and consequently to an interruption or a termination of the evaluation or a moderate rating. Two other studies validated REDCap as a new data collection tool and rated its usability using the IBM CSUQ. In both studies a greater user acceptance was obtained [12,14]. The main difference to our research is the study size used for REDCap implementation. With increasing study size, the system requirements grow. Large studies, such as the HCHS, call for additional functions highlighted with our user stories and their prioritization. Since REDCap is a tool with a high flexibility, it offers great prerequisites for requirement implementations. A positive example is R1, standardization of the interface to German. Corresponding requirements can be adjusted by using modules offered by the REDCap community, e.g. Multi-Language-Module. However, other requests need REDCap-independent solutions, e.g. R2 to include patient information, findings, and laboratory results to REDCap. This could be solved with an integration of REDCap into the HCHS infrastructure. In general, we expect an increase of acceptance among the users after the implementation of all requirements. Additionally, training and support structures for users should be established to improve the impeccable usage of the system. Next steps would be the integration of REDCap into the HCHS IT infrastructure to elevate the HCHS department network centralization. This would consequently increase the independence of the study from external network components and ensure administrative sovereignty. After all, adjustments to the data collection can be made faster, more efficient, and flexible by using REDCap which consequently leads to better data quality.

5. Conclusion

The proposed concept for the use of REDCap as the central component in the HCHS data collection is a feasible solution for large-scale clinical studies. Overall, the study

assistants consider the system to be usable for the HCHS. Nevertheless, requirements emerged that REDCap could not meet without extensions. It is also necessary to integrate REDCap into the existing study infrastructure for use in HCHS. The approach could reduce dependence on proprietary software solutions and facilitate the technical design of large-scale research studies. Furthermore, REDCap provides a secure, user-friendly, cost-effective, and flexible solution for data management. The implementation of this concept should help the HCHS team to achieve the long-term objectives of the study and could serve as a template for other research-based infrastructures.

Declarations

Conflict of Interest: The authors declare no conflict of interest.

Contributions of the authors: All authors were involved in the planning of the project. YH, AH, AS, UH provided the HCHS metafiles. NS, KG, AH, AS, UH were involved in the REDCap eCRF building process. NS, KG planned, executed, and evaluated the quantitative and qualitative user evaluation process. YH provided the initial HCHS infrastructure ideas. NS, KG expanded the initial HCHS infrastructure ideas. LT, FÜ critically revised the manuscript. All authors read and approved the final manuscript.

References

[1] Helms RW. Data Quality Issues in Electronic Data Capture. Drug Inf. J. 2001 Jul;35(3):827–37. doi: 10.1177/009286150103500320

[2] Welker JA. Implementation of electronic data capture systems: Barriers and solutions. Contemp Clin Trials. 2007 May;28(3):329–36. doi: 10.1016/j.cct.2007.01.001

[3] Walther B, Hossin S, Townend J, Abernethy N, Parker D, Jeffries D. Comparison of Electronic Data Capture (EDC) with the Standard Data Capture Method for Clinical Trial Data. PLoS ONE. 2011 Sep 23;6(9):e25348. doi: 10.1371/journal.pone.0025348

[4] Shah J, Rajgor D, Pradhan S, McCready M, Zaveri A, Pietrobon R. Electronic Data Capture for Registries and Clinical Trials in Orthopaedic Surgery: Open Source versus Commercial Systems. Clin Orthop Relat Res 2010 Oct;468(10):2664–71. doi: 10.1007/s11999-010-1469-3

[5] 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. doi: 10.1016/j.jbi.2008.08.010

[6] Jagodzinski A, Johansen C, Koch-Gromus U, Aarabi G, Adam G, Anders S, et al. Rationale and Design of the Hamburg City Health Study. Eur J Epidemiol. 2020 Feb;35(2):169–81. doi: 10.1007/s10654-019-00577-4

[7] Manolio TA, Weis BK, Cowie CC, Hoover RN, Hudson K, Kramer BS, et al. New Models for Large Prospective Studies: Is There a Better Way? Am J Epidemiol. 2012 May 1;175(9):859–66. doi: 10.1093/aje/kwr453

[8] 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. doi: 10.1016/j.jbi.2019.103208

[9] Reichold M, Heß M, Kolominsky-Rabas P, Gräßel E, Prokosch HU. Usability Evaluation of an Offline Electronic Data Capture App in a Prospective Multicenter Dementia Registry (digiDEM Bayern): Mixed Method Study. JMIR Form Res. 2021 Nov 3;5(11):e31649. doi: 10.2196/31649

[10] Zhengwu Lu. Electronic Data-Capturing Technology for Clinical Trials: Experience with a Global Postmarketing Study. IEEE Eng Med Biol Mag. 2010 Mar;29(2):95–102. doi: 10.1109/MEMB.2009.935726

[11] Lewis JR. Measuring Perceived Usability: The CSUQ, SUS, and UMUX. Int J Hum–Comput Int. 2018 Dec 2;34(12):1148–56. doi: 10.1080/10447318.2017.1418805

[12] Dunn WD, Cobb J, Levey AI, Gutman DA. REDLetr: Workflow and tools to support the migration of legacy clinical data capture systems to REDCap. Int J Med Inform. 2016 Sep;93:103–10. doi: 10.1016/j.ijmedinf.2016.06.015

[13] Lewis JR. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. Int J Hum–Comput Int. 1995 Jan;7(1):57–78. doi: 10.1080/10447319509526110

[14] Ndlovu K, Mauco KL, Grover S, Kovarik C, Williams VL. Evaluation of REDCap as an alternative data management and quality improvement tool for cutaneous lmphoma care in Botswana. In: Proceedings of the 12th Health Informatics in Africa Conference; 2019 Nov 20-22; Port Elizabeth, South Africa; p. 51–57.