



HHS Public Access

Author manuscript

J Registry Manag. Author manuscript; available in PMC 2016 January 14.

Published in final edited form as:

J Registry Manag. 2015 ; 42(3): 111–114.

Open-source, Rapid Reporting of Dementia Evaluations

Rasinio S. Graves, BA

University of Kansas Alzheimer's Disease Center rgraves@kumc.edu

Jonathan D. Mahnken, PhD

Department of Biostatistics, University of Kansas Medical Center jmahnken@kumc.edu

Russell H. Swerdlow, MD

University of Kansas Alzheimer's Disease Center rswerdlow@kumc.edu

Jeffrey M. Burns, MD, MS

University of Kansas Alzheimer's Disease Center jburns2@kumc.edu

Cathy Price

Department of Biostatistics, University of Kansas Medical Center cprice@kumc.edu

Brad Amstein

University of Kansas Alzheimer's Disease Center bamstein@kumc.edu

Suzanne L Hunt

Department of Biostatistics, University of Kansas Medical Center shunt@kumc.edu

Lexi Brown

Department of Biostatistics, University of Kansas Medical Center abrown8@kumc.edu

Bhargav Adagarla

Division of Medical Informatics, University of Kansas Medical Center badagarla@kumc.edu

Eric D. Vidoni, PT, PhD

University of Kansas Alzheimer's Disease Center evidoni@kumc.edu

Abstract

The National Institutes of Health Alzheimer's Disease Center consortium requires member institutions to build and maintain a longitudinally characterized cohort with a uniform standard data set. Increasingly, centers are employing electronic data capture to acquire data at annual evaluations. In this paper, the University of Kansas Alzheimer's Disease Center reports on an open-source system of electronic data collection and reporting to improve efficiency. This Center capitalizes on the speed, flexibility and accessibility of the system to enhance the evaluation process while rapidly transferring data to the National Alzheimer's Coordinating Center. This framework holds promise for other consortia that regularly use and manage large, standardized datasets.

Keywords

Alzheimers; reporting; registry; database management system

Introduction

In the last decade, there has been a significant push for electronic data capture (EDC) and storage of clinical research information.¹ The use of EDC for research results in efficiency gains in entry time and query resolution, reduced cost, and improved turnaround time to a completed, validated dataset.^{2,3} Accuracy rates between EDC and paper-based methods are comparable, and EDC cuts the data recording workload by half.³ Recognizing the value of EDC for clinical research, the Food and Drug Administration has published recommendations for use of EDC systems in clinical trials.⁴

Beyond being employed effectively by many clinical trials to manage research operations, EDC offers potential benefits for management of observational studies or registry operations. In this paper, the University of Kansas Alzheimer's Disease Center (KU ADC) reports on the implementation of one such system, based on open source software, that integrates EDC with a storage, retrieval, and reporting methodology at the KU ADC. The reporting methodology was developed, and maintained by our investigative team, which has had approximately 10 years of experience using this evaluation protocol (7 on paper, 3 using EDC).

The KU ADC is responsible for recruiting and sustaining longitudinal data on over four hundred individuals (Clinical Cohort). Enrollees into the Clinical Cohort agree to have their health and contact information made available to investigators conducting studies, serving as both a longitudinal study of aging and a source of well-characterized potential study participants. All Alzheimer's Disease Centers funded through the National Institute on Aging are required to capture standard data, called the Uniform Data Set (UDS), at annual evaluations. This data from all centers is then pooled for general researcher use at the National Alzheimer's Coordinating Center (NACC).⁵ Every member of the Clinical Cohort receives an in-depth clinical evaluation guided by the Clinical Dementia Rating (CDR), a semi-structured interview of a participant and informant knowledgeable about the participant's daily activity.⁶ Participants also undergo a standard neuropsychological battery and have blood drawn. All data collection forms and fields are standardized by NACC to allow for data aggregation across centers.

At the KU ADC, findings of the semi-structured interview and the neuropsychological test battery are presented at a weekly consensus diagnostic conference of clinicians and evaluators. Each case is presented and discussed to arrive at a consensus on dementia severity rating and diagnosis. The consensus determination is the final research diagnosis for the individual for that year. Data from each KU ADC evaluation must quickly be made available to support the consensus conference and to give recruitment staff and investigators the latest cognitive status information for each member of the Clinical Cohort.

A challenge faced by the KU ADC is managing the data produced by up to fifteen weekly evaluations, including entry, retrieval, validation and storage. Paper versions of the evaluation instruments total over one hundred pages per individual, per visit. Manual data entry from paper source documentation into a database incurs significant labor costs.

Creating paper summary case report forms for distribution to up to ten evaluators at the consensus conference requires additional investment of time and resources.

Timely and efficient data flow from first contact with the patient, to a final consensus diagnosis is important to participants interested in feedback, researchers considering participants for their studies, and for reporting to the NACC. Inefficiencies inherent in a paper-based system do not allow for a swift flow of information through the KU ADC and thus we prioritized the development of an EDC system to enhance efficiency (launched June 1, 2012). We subsequently developed a retrieval and reporting system that markedly reduced our time to completion of these activities, and decreased the effort allotted to the reporting process (launched October 27, 2012). Our EDC system meets the requirements for ADCs, and we have found this approach to be more efficient in terms of time, effort, and resources than our previous data capture and retrieval system.

This paper presents our method of EDC and reporting on data that requires timely interaction and evaluation. The innovation is in the use of freely available and open source software, developed through a collaboration of both professional data managers and researchers without formal programming training. A experienced data management team, researchers in the KU ADC Data Management and Statistics Core, built the EDC on the framework of the Research Electronic Data Capture (REDCap) web application.⁷ Recognizing the need for more timely data evaluation of the participant data, researchers in the Clinical and Outreach, Recruitment and Education Cores, developed the reporting tools.

Methods

Evaluation and Data Collection

The KU ADC is required to conduct annual evaluations on the Clinical Cohort for longitudinal study and potential referral to specific clinical research projects. Visits occur in research-dedicated, clinic-style evaluation rooms equipped with standard personal computer workstations that have access to a secure network connection. Two individuals, an intake coordinator and clinician, perform the research evaluation interview and complete all research records using REDCap. At a separate visit, a psychometrician completes all neuropsychological testing on paper forms, and enters test scores into REDCap after completing the visit.

At the first point of contact, each participant is given a unique identifying alpha-numeric value, which follows them throughout their participation in all KU ADC-supported studies. The KU ADC follows standard NACC protocols for all data collection to meet the UDS requirements. This standardizes the language, data elements and relative order of the data acquisition forms. We also collect additional measures unique to the mission of the KU ADC (e.g. waist/hip ratio, physical activity), and scan, upload and save paper consent forms. All Clinical Cohort evaluations and findings are entered in a single REDCap project, set up to allow longitudinal reuse of electronic forms (E-Forms). E-Forms guide the phases of the semi-structured interview while maintaining consistency with the Uniform Data Set. Variable names, data types, and acceptable formats and ranges used in REDCap are standard and were defined by the NACC, with the exception of those fields specific to the KU ADC.

An audit log is kept of all data entry, imports and exports. A typed explanation is required for any data change.

Data Storage and Access

REDCap uses two secure servers, an application server and a database server, hosted by the university. Both these servers are virtual machines that run the on SUSE Linux Enterprise 11 (64 bit) operating system. The application is implemented using PHP and the database uses MySQL.

We evaluated three methods for data access: automated data export, Application Program Interface (API) for real-time query of the database, and manual export via the REDCap interface. With automated data export, all project-specific data captured in REDCap is exported as comma separated volumes (CSV) to a secure server nightly using a scheduled, secure file export (i.e. a cron job). Along with the CSV, the user can export a second script that assigns labels to the values in each field. The REDCap API gives the user the ability to automatically acquire the data or generate a report with the entire data set from a specific project. We decided the API was unnecessarily programming intensive, given that we require data weekly for our consensus diagnostic conferences. The third option, manual export via the REDCap interface, is available but it was deemed too labor intensive for our needs.

Data Management

For data access, cleaning and reporting, we chose two open source packages: R⁸ for computation and reporting, and Perl for string manipulation.

R is a free statistical computing and graphics environment that supports numerous open source packages and extensions with diverse capabilities. We recommend using RStudio (RStudio⁷, Boston, MA www.rstudio.com) as a free integrated development environment (IDE) for developing and implementing R. RStudio can run on most Windows, Linux and OS X machines. We install R and RStudio on workstations of staff who interact with the reporting process. Alternatively, RStudio can be implemented as a Linux server-based application accessed through a web browser to allow access and version standardization for all staff. For most Linux and OS X systems, Perl comes standard with the operating system. For Windows we used the Perl IDE “Padre” (<http://padre.perlide.org/>). Perl is primarily implemented to take advantage of its strengths in string manipulation.

Reporting is implemented through a collection of in-house R scripts that prompt the user to select from several options using the command line interface. Currently, the user can generate a consensus conference report, or produce a preformatted and addressed letter of appreciation, scheduling reminders, or physician notification letter for any individual in our Clinical Cohort. Functions from the ‘knitr’⁹ package use a combination of R and a markdown formatting language from the R package ‘rmarkdown’¹⁰ to assemble a hypertext markup language (HTML) document based on a template. At the consensus diagnostic conference, these individualized HTML reports are reviewed in a web browser app created using the ‘Shiny’ package (RStudio, Boston, MA). The Shiny package is a web application framework developed for R to allow for interactive and reactive data viewing via a web

browser. Standardized neuropsychological test scores¹¹ for the most recent visit are also displayed. Using Shiny, the reports are easily listed in a dropdown menu and displayed via projector for all staff to see as each case is presented. Individualized HTML documents can also be generated and securely emailed to individuals attending the consensus conference via telephone.

Electronic vs. Paper Source Accuracy Comparison

We conducted a comparison trial of EDC versus paper data capture followed by entry into an electronic database. Four participants were randomly selected for the comparison. An evaluator performed the usual evaluation with direct entry of data into the electronic record, using REDCap. A second individual sat in the evaluation room and recorded the findings on paper, confirming the ratings with the evaluator. A third person then entered the data into the database from the paper source. We compared the matching fields for agreement on one hundred twenty-four NACC variables. The Compare procedure within SAS 9.3 (SAS Institute Inc⁹, Cary, NC) was used to perform this analysis. We also assessed improvements in time to completion of consensus diagnosis and transfer of the record to NACC between the original paper based method and the EDC and R-based reporting method.

Results

Electronic vs. Paper Source Accuracy Comparison

Out of the 124 variables reviewed, fourteen had true missing value differences, where the data was missing for one of the two data capture methods but should have been entered. For 124 variables across four patients, there were a total of 496 pairs of fields to compare. Of these 496 pairs, there were thirty-one total pairs, (about 6.3%) that were discordant between data capture methods.

Time to Consensus Diagnostic Conference

Use of our R-based automated reporting improved time to completion of individual evaluations. After Alzheimer Disease Center designation (Aug 15, 2011) and prior to implementation of the R-based reporting system (October 27, 2012), the median number of days between the first evaluation and consensus conference was 31. Using the R-based reports, the median time to consensus diagnosis was reduced to 17 days.

Electronic vs. Paper Source Time to Completion Comparison

Use of EDC and automated reporting improved time to completion of individual records. After Alzheimer Disease Center designation (Aug 15, 2011) and prior to the R-based reporting system (October 27, 2012), the average time from first visit to record completion and NACC submission was 211.3 days. Using EDC and our R-based reporting method, the median time to record completion is 95.1 days. This improvement means that verified data are entering the final NACC data set more quickly for use by researchers.

Discussion

We describe the data capture process for longitudinal measures of our Clinical Cohort and the several advantages of using EDC along with an open source system for query and display. The data is immediately tracked, locked and available, resulting in more secure and ultimately more readily-available data compared to paper capture. This is the major benefit for the KU ADC Clinical Cohort. Another advantage of EDC is the ability to validate and mandate data entry. REDCap, as with most data capture and management systems, provides a means to require field entry and restrict ranges or formats. These controls potentially increase completeness and accuracy of the data compared to paper capture, which is more susceptible to human error. The ability to access information on a particular patient with little delay, results in significantly reduced time to completion of evaluation. The Center benefits with improved screening time for trials and reduced data entry burden. The participant benefits from earlier feedback on their visit.

This investigative team has a broad range of experience in data management. The individuals who built and maintain the REDCap-based EDC system each have many years of experience in database development and EDC. The individuals who developed the R-based reporting system had hobby programming experience only and no formal training. This underscores the utility and ease of use for open source platforms of data management. Investigative teams with a broad range of professional experience in data management can effectively build on open-source tools to improve operational efficiency.

Adoption of this method has important limitations that should be considered. There is some evidence that EDC evaluations require more time on the part of the evaluator.³ A second limitation is the lack of source information to cross check against the electronic database. In a standard paper source to electronic database arrangement, the electronic database can be checked against the paper source for accuracy. When the electronic source is the only capture record, validity checks must be robust enough to catch the mistake. As reported, we found that there were minimal entry errors when we directly compared the EDC and paper sources. However, because we did not record evaluations we were unable to ascertain which method was most accurate. One study looking at EDC of clinical trial data recently found significantly higher error rates with EDC early in the study.³ This may have been a result of lack of familiarity with the technology. An extended training period may be warranted before “going live” with EDC.

Conclusion

There are many options when selecting software to utilize for data capture. We chose to use open source software to facilitate our EDC and reporting needs of the KU ADC, and have been served well by these systems. Further, the use of open source tools provides a sharable platform for multi-institution collaborations. Based on our experiences, other research teams and investigators with varying levels of data management and programming experience may opt for similarly applying open source technologies to collect, store, and utilize their project information.

References

1. El Emam K, Jonker E, Sampson M, Krleza-Jeric K, Neisa A. The use of electronic data capture tools in clinical trials: Web-survey of 259 Canadian trials. *J Med Internet Res*. 2009; 11(1):e8. [PubMed: 19275984]
2. Litchfield J, Freeman J, Schou H, Elsley M, Fuller R, Chubb B. Is the future for clinical trials internet-based? A cluster randomized clinical trial. *Clinical Trials*. 2005; 2(1):72–79. [PubMed: 16279581]
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; 6(9):e25348. [PubMed: 21966505]
4. United States Food and Drug Administration. Guidance for Industry: Computerized Systems Used in Clinical Investigations. United States Department of Health and Human Services; 2007.
5. Weintraub S, Salmon D, Mercaldo N, et al. The Alzheimer's Disease Centers' Uniform Data Set (UDS): the neuropsychologic test battery. *Alzheimer Dis Assoc Disord*. 2009; 23(2):91–101. 2743984. [PubMed: 19474567]
6. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*. 1993; 43(11):2412b–2414. [PubMed: 8232972]
7. 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*. Apr; 2009 42(2):377–381. [PubMed: 18929686]
8. *R: A Language and Environment for Statistical Computing* [computer program]. R Foundation for Statistical Computing; Vienna, Austria: 2013. URL: <http://www.r-project.org>
9. *knitr* [computer program]. 2014. URL: <http://rmarkdown.rstudio.com>
10. *R Markdown* [computer program]. 2015. Version 2 URL: <http://yihui.name/knitr/>
11. Shirk SD, Mitchell MB, Shaughnessy LW, et al. A web-based normative calculator for the uniform data set (UDS) neuropsychological test battery. *Alzheimers Res Ther*. 2011; 3(6):32. 3308021. [PubMed: 22078663]