Electronic data collection in epidemiological research

The use of REDCap in the Pelotas birth cohorts

Cauane Blumenberg[1]; Aluísio JD Barros[1]

[1] Postgraduate Programme in Epidemiology, Federal University of Pelotas, Pelotas, Brazil

Keywords

Epidemiology, public health informatics, electronic data capture, medical informatics, translational research

Summary

Objectives: This paper describes the use of Research Electronic Data Capture (REDCap) to conduct one of the follow-up waves of the 2004 Pelotas birth cohort. The aim is to point out the advantages and limitations of using this electronic data capture environment to collect data and control every step of a longitudinal epidemiological research, specially in terms of time savings and data quality.

Methods: We used REDCap as the main tool to support the conduction of a birth cohort follow-up. By exploiting several REDCap features, we managed to schedule assessments, collect data, and control the study workflow. To enhance data quality, we developed specific reports and field validations to depict inconsistencies in real time.

Results: Using REDCap it was possible to investigate more variables without significant increases on the data collection time, when comparing to a previous birth cohort follow-up. In addition, better data quality was achieved since negligible out of range errors and no validation or missing inconsistencies were identified after applying over 7,000 interviews.

Conclusions: Adopting electronic data capture solutions, such as REDCap, in epidemiological research can bring several advantages over traditional paper-based data collection methods. In favor of improving their features, more research groups should migrate from paper to electronic-based epidemiological research.

Correspondence to

Cauane Blumenberg
Postgraduate Programme in Epidemiology, Federal University of Pelotas, Pelotas, Brazil
1160 Marechal Deodoro St. – 3rd floor
Pelotas, RS, 96020–220, Brazil
Phone: +55 53 32841300
Email: cauane.epi@gmail.com

Appl Clin Inform 2016; 7: 672–681
http://dx.doi.org/10.4338/ACI-2016-02-RA-0028
received: March 4, 016
accepted: June 14, 2016
published: July 13, 2016
Citation: Blumenberg C, Barros AJD. Electronic data collection in epidemiological research: The use of REDCap in the Pelotas birth cohorts. Appl Clin Inform 2016; 7: 672–681
http://dx.doi.org/10.4338/ACI-2016-02-RA-0028

Funding

This article is based on data from the study „Pelotas Birth Cohort, 2004" conducted by Postgraduate Program in Epidemiology at Universidade Federal de Pelotas, with the collaboration of the Brazilian Public Health Association (ABRASCO). The 2015 follow-up assessment was supported by a grant from the Brazilian Scientific Development Council (CNPq) with resources from the Science and Technology Department, Brazilian Ministry of Health (DECIT/SCTIE/MS). From 2009 to 2013, the Wellcome Trust supported the 2004 birth cohort study. The World Health Organization, National Support Program for Centers of Excellence (PRONEX), Brazilian National Research Council (CNPq), Brazilian Ministry of Health, and Children's Pastorate supported previous phases of the study.

1. Introduction

Epidemiological research contributes to the development of several fields related to the populations' wellbeing. It can be used, for instance, to identify disease etiology, to guide and evaluate health interventions, and to monitor the health conditions of populations [1]. Some public health achievements, such as vaccination, seat belt use, sanitation, and healthy behaviors were only possible due to epidemiological studies [2]. Conducting these studies typically involves large numbers of people, complex designs and high costs. Among several epidemiological designs, birth cohort studies are the most complex, requiring repeated follow-up of the participants in order to collect information along their lifecycle. This makes the cohort design the most suited study to evaluate incidence, as the time of event occurrence is clearly defined. Some difficulties regarding this longitudinal design arise, such as the enrollment and retention of the participants along the cohort follow-ups [3].

Our research group has a long-standing experience in conducting birth cohort studies. The first started in 1982, enrolling all newborns from mothers residing in the urban area of Pelotas (a mid-sized city located in Southern Brazil) in that year. Since then, another three birth cohorts were initiated, in 1993, 2004, and 2015, involving nearly 20,000 births [4]. Considering the long follow-up time and the large number of subjects, these studies require large fieldwork teams and typically involve long and diverse questionnaires.

Paper-based questionnaires were the norm in epidemiological studies until recently. However, applying long paper questionnaires in many subjects brings many difficulties, such as controlling the application of all questionnaires to every participant, ensuring high data quality, perform timely data quality checks, just to mention a few. Additionally, the period between data collection and analysis is longer with paper-based questionnaires [5]. These difficulties can be minimized by electronic data capture (EDC), helping to check for missing data, controlling the questions that have to be applied, checking for invalid answers in real time, and making the questionnaires smarter by the reuse of data [6, 7]. Hence, it is possible to reduce survey time and the period to generate data ready to analyze.

EDCs are usually part of an extended data collection platform that allows to efficiently control who participated and when. Other benefits such as integrated interview scheduling and workflow control can be achieved through such platforms. Despite these benefits, shifting from paper to electronic-based environments still encounters some resistance [8, 9, 10]. This happens for two main reasons: the lack of personnel with enough knowhow to build, structure, and manage EDC systems, and the high financial and time investments needed to develop local EDC solutions [11].

There exist several ready-to-use EDC applications available, of which the Research Electronic Data Capture (REDCap), developed by Vanderbilt University [12], is one. REDCap is a web-based platform that goes beyond data collection, since it involves several features that help in the research environment, such as interview scheduling and custom data reports. Additionally, it is distributed free of charge, making it a good alternative for research centers, large or small, to conduct epidemiological studies. This paper describes the use of REDCap to support the conduction of a follow-up wave of the 2004 Pelotas birth cohort, when 3,566 cohort participants (86.6% of the original cohort) and their mothers were assessed. We also comment about the advantages, difficulties, and limitations of using this EDC.

2. Background and Significance

2.1 History and context of our research needs

The 2004 Pelotas birth cohort accompanied all the births occurring in the city that year and recruited the participants to a perinatal study. We re-assessed the same babies at 3 months, and at 1, 2, 4, 6, and 11 years of age. The topics studied are varied including growth, child development and cognitive ability, nutrition, body composition, mental health, and non-communicable diseases. Most analysis focuses on the effect of early life determinants on outcomes of interest, or on how life-long exposures affect health in a longitudinal approach.

Each assessment involves several questionnaires (e.g. socioeconomic data, health status, lifestyle, food consumption) and physical exams such as anthropometry, body composition (dual-energy X-ray absorptiometry and air-displacement plethysmography), body imaging (3D photonic scanner), lung function (spirometry and the diffusing capacity of the lungs for carbon monoxide), pulse wave velocity, and carotid intima-media thickness. We also collect biological samples such as blood and saliva. All this implies in a large amount of data from many different sources and formats to be collected, recorded, organized, and documented.

During the follow-up of the 2004 Pelotas birth cohort when children were 11 years old, 3,566 cohort participants and 3,533 mothers were assessed and interviewed, achieving 86.6% follow-up rate. This cohort wave was mainly conducted at our research clinic, which is organized in three main sectors (reception, interviewing, and physical examination), each including several specific stations. Each station has one or more staff specially trained for the activity. A small team manages the flow of participants along the stations in order to minimize the time spent in the research clinic. In this context, we adopted the REDCap software with the intent of computerize every step of the study, from interview scheduling to data collection and verification.

2.2 REDCap and its functionality

The Research Electronic Data Capture (REDCap) software is a web-based platform that aims to simplify the development of electronic data capture forms to be used in research [12]. Today, over 250,000 research projects in 99 countries are conducted using the REDCap platform [13].

Apart from electronic forms, REDCap has a set of tools including a calendar, data reports, and data statistics. The calendar can be used to schedule interviews and to control the status of completeness of the forms. Data reports are used to display a list of records that match a user-defined criterion, and data statistics provide real time statistics about the data collected. REDCap also guarantees confidentiality by controlling the users' access rights to each form, and by masking information that could identify the survey participants.

REDCap improves data quality by employing branching logic, which takes into consideration answers from previously applied questions to prevent the application of questions that are not relevant for the respondent. Additionally, field validation reduces the likelihood of inconsistencies (e.g., a woman reporting a sum of deliveries and abortions larger than the previously reported total number of pregnancies), invalid values (e.g., a textual answer on a numerical field), and out of range values (such as an age of 150 year). In section 4 we describe how we used these REDCap features to manage the most recent 2004 birth cohort wave in our research clinic.

3. Objectives

This paper describes the use of REDCap to conduct one of the follow-up waves of the 2004 Pelotas birth cohort. The aim is to point out the advantages and limitations of using this electronic data capture environment to collect data and control every step of a longitudinal epidemiological research, specially in terms of time savings and data quality.

4. Methods

4.1. User groups and access rights

The most recent follow-up from the 2004 Pelotas birth cohort was conducted when the children were 11 years old. The data collection lasted eight months, where we performed examinations and interviews in 7,099 individuals (cohort members and their mothers) who provided informed consent. Considering the large amount of data collected, and the daily number of interviews (up to 76), we needed a sizeable team to conduct the field work. A total of 45 people were involved in the field work, their roles described in ▶Table 1, along with their user group.

The access rights to each REDCap instrument was given on user group level. ▶Figure 1 presents the use case diagram that graphically describes the access rights to the REDCap instruments and projects. Our study was composed by two projects: Contact Information and Data Collection. Each project had its own set of instruments, illustrated by the ellipses. The dolls, known as actors, represent the user groups, and the association lines depict the actors' access rights (edit or view) to the entities. The diagram is hierarchical in the sense that an actor closer to the instrument inherits its parent access rights (but not the other way round).

4.2 Managing a birth cohort follow-up with REDCap

The scheduling team, composed of three members, was responsible for scheduling the assessment of each participant. They registered the appointments in the calendar from the Contact Information REDCap project, which held the contact information on all the cohort participants, including their close family and friends. Using the calendar from this project, the reception team prepared in advance the materials and identification tags needed for the interview, and could update the contact information. Among the materials there was a wristband, containing a barcode with the participants' identification number. This number was used to access each REDCap instrument, and the barcode was scanned on each data collection station to avoid typing errors. The use of an identification number instead of any personal information also guaranteed the anonymity of the participants.

We developed two REDCap projects, allowing us to use one calendar per project. The calendar from the Contact Information project was used to schedule the participants' assessment, while the calendar from the Data Collection project served as a check-in control. This was important since not all the scheduled participants would come to the research clinic on the scheduled date and time. Thus, the receptionists created a calendar event on the Data Collection project to signal the workflow control team that the participant had arrived and was ready to be assessed.

The workflow control team was responsible for keeping track of each participant in the research clinic, and directing them through the stations in a way to minimize their time in the clinic. In order to monitor which stations the participant had already completed, we used a traffic light approach to indicate its current condition: not yet started = white; on-going = red; interrupted = yellow; completed = green. The status was updated in real time for each station. This approach made possible to ensure that each participant had completed all steps before dismissal.

The data collection team comprised thirty members, divided into general interviewers, psychologists, and physical examiners. Interviewers were responsible to apply twenty-three questionnaires, as shown in ▶Table 2 (further details about the questionnaires are published elsewhere [14]). The psychologists applied two mental health tests, while physical examiners performed four examinations, including anthropometry, body composition (two tests), and body image. Except for the anthropometry, each exam ran its own software, making it impossible to record the results directly into REDCap. For this reason, key exam results were manually recorded in order to generate a result letter to be delivered to participants at the end of their assessment.

To deal with special situations and to guarantee that the research process was being well conducted, we had three field supervisors with research experience. The members took daily turns to follow all the steps in the research clinic. To facilitate their work we created fifteen data reports to detect odd situations during the participants' assessment. One of the reports grouped the participants that had refused to undergo any of the examinations, so the supervisors could try to encourage their participation. Another report listed the participants that had any incomplete instrument, avoiding the occurrence of erroneous missing data.

The supervisors also checked the comment logs to verify if the data collection team had pointed out any special situation, aiming to solve these issues in real time. This enhanced data quality and simplified database post-processing. When there were recurrent errors, they were discussed with the study coordinators to adequate the instruments to avoid further errors. The REDCap administrator was responsible to perform any modifications on the instruments, and coordinated the report creation, the server stability, and the database post-processing. Only the system administrator had enough rights to export the data already collected in order to avoid data leak. For this reason, he was in charge of keeping track of the data collection statistics to notify the study coordinators about the production of the research clinic.

5. Results

5.1 Time savings

During the follow-up, we collected 2,081 variables for each pair of participants (950 variables provided by the cohort member and 1,131 by the mother). This represented 59.7% more variables collected than the previous cohort follow-up, when we investigated 838 variables. However, this increase did not add extra time to the data collection process. Using REDCap took, on average, 3.2 hours (SD 0.8). This is similar to the 3-hour duration from the previous 2004 Pelotas birth cohort wave [14].

Using REDCap, we also saved time to achieve a closed database. After the last assessment we obtained a database ready to analyze in only 2 weeks. When using paper-based questionnaires, it was necessary, on average, 5 months to achieve a closed database. This extra time was generated by the need to visually check inconsistencies and to double type the questionnaires, actions that are automatically performed in REDCap.

5.2 Reduced data entry errors and increased data quality

REDCap automatically registers as missing the questions skipped by the branching logic. These are not erroneous missing, since they are different from missing data generated by accidental question skips. Using REDCap, we identified only 80 erroneous missing data among 3,566 cohort member assessments. Considering that each cohort member generated 950 variables, this type of error was negligible in contrast to the amount of data collected.

We also enhanced data quality using the field validation feature. We analyzed two types of errors, invalid values and out of range values. Among all the assessments, there were no invalid values and 10 out of range values. The out of range values occurred in the anthropometry station, where the body measurements were erroneously registered. However, the research coordinators identified these errors during data collection, correcting them before the participant dismissal.

5.3 High follow-up rates

Our birth cohorts are historically known for their high follow-up rates [4, 14, 15], but these elevated rates are not easy to achieve. During the first 25 weeks of study, we interviewed, on average, 250 participants per week. However, in the remaining nine weeks, the production reduced to an average of 94 participants per week. This happens since some participants do not want to attend to the research clinic. Aiming to bypass this drawback, from the twenty-sixth week of study we started to conduct home interviews along with the research clinic interviews, as show in ▶Figure 2.

We conducted 105 home interviews, being 50 of them applied to cohort members and 55 to their mothers. This was not an elevated number of interviews, but contributed to achieve 86.6% of follow-up rate in the end of study.

6. Discussion

6.1 Advantages of using REDCap

Using REDCap and comparing to a previous follow-up from the 2004 Pelotas birth cohort, we could investigate almost 60% more variables without spending extra data collection time. This is very beneficial for the study, as more associations and hypothesis may be investigated. This EDC also contributed to enhance data quality, since we identified a negligible number of errors during the eight months of follow-up. We could not compare the number of errors to previous follow-ups, as we do not have this documented. Although, we perceived better data quality using REDCap as the complexity to achieve a closed database was lower. When using paper-based questionnaires it was necessary 5 months to achieve a closed database, whereas using REDCap it was necessary only 2 weeks.

Having a closed database contributes to the process of inconsistency check. Inconsistencies in epidemiological research may occur by two different sources, due to errors in the questionnaire or due to interviewer's error. Sometimes the only way to solve the inconsistent data is contacting the interviewee to check the information provided. In order to get reliable information, the time between the interview and the checking should be the smallest possible. This is another advantage of using REDCap, as we can avoid the process of data typing and translation, which made the gap between data collection and information checking wider when using paper-based questionnaires.

During the course of the study, it was possible to check interim results. Hence, we could evaluate the characteristics of the population already followed, the frequency of variables of interest, and the production of the research clinic (number of participants interviewed). This was crucial to determine when the home interviews should start in order to keep the pace of interviews, and to achieve a high follow-up rate.

Another benefit of using electronic data capture solutions is the reduction of study costs. Migrating from paper to EDC environments might apparently seem unworthy as the initial investment is not negligible. The acquisition of data collection devices, i.e. tablets, notebooks, smartphones, and the employment of a computer expert to set up this environment, contributes to the initial elevated costs. On the other hand, EDC solutions avoid the costs for paper, printing questionnaires, and data entry. These savings make the costs per interview and the overall research costs lower when compared to the paper research methods [5, 16, 17]. Additionally, the bigger the study populations, higher are the financial benefits [18]. These benefits can be even better through internet data collection, since no interviewers are necessary [19].

6.2 REDCap limitations

The longitudinal design of the cohort studies requires the participants to be periodically followed-up. Although some information varies on a regular basis, others may not vary in a similar pace and might be the same between different cohort waves. Thus, a good practice would be to import some information given by the participant on a previous follow-up. This would make possible to check the coherence between the previous and the current answer; i.e. one could not say that it has less years of study than it was stated on a previous follow-up. This could also accelerate the data collection process since the interviewee would not need to provide answers to all the questions.

The REDCap software implements a longitudinal design that allows reuse of data from different follow-ups. This design requires the same data to be collected on different moments. However, many longitudinal studies (including our birth cohorts) apply different questionnaires across the follow-ups, being necessary to create one project for each wave. REDCap projects are independent, making it difficult to reuse data between them. Thus, it would be interesting to develop an easy method to reuse data from different, in order to exploit the above-mentioned benefits.

Setting up the REDCap server is not a simple task [20]. It requires a professional with enough expertise to install and configure the necessary tools to get the server running. Hence, epidemiological research teams might need to employ a computer expert to perform this task. The employment of a computer expert might be associated with high salaries, and this may contribute to the resistance to migrate from paper to electronic-based data collection tools.

7. Conclusions

The features that implemented by the current electronic data capture solutions enable their use not only to assist the data collection process, but also to support the management of the entire study. Agile data collection, better data quality, higher control of the research process, and faster results are some of the benefits obtained by using EDC solutions. Adopting REDCap to conduct a birth cohort follow-up, we could collect more data without increasing the data collection time. Additionally, we identified negligible inconsistencies and errors during the research process. However, there are some features to be considered by REDCap, such as data reuse, which could potentially increase data quality and reduce data collection time. Thus, we encourage research groups to adopt EDCs in order

to exploit their benefits, contribute to their development, and to strengthen the growing field of electronic data capture.

Clinical Relevance Statement

Epidemiological researches are key factors to support the employment of new actions to benefit the populations' wellbeing. Adopting electronic data capture solutions to conduct epidemiological research bring several advantages over the traditional research methods, i.e. reduced data collection time, reduced costs, higher data quality, and less time to achieve a ready to analyze database. Hence, we encourage study coordinators to shift from paper to electronic-based solutions to enhance the epidemiological research process.

Conflict of Interest

The authors declare that they have no conflicts of interest in the research.

Protection of Human Subjects

The study protocol was analyzed and approved by the Medical Research Ethics Committee of the Federal University of Pelotas. We also obtained written consent from the participants.

Acknowledgements

We would like to thank the Faculty of Medicine of Ribeirão Preto (FMRP) from the Universidade de São Paulo (USP), which introduced to us the REDCap environment and provided us a testing platform.

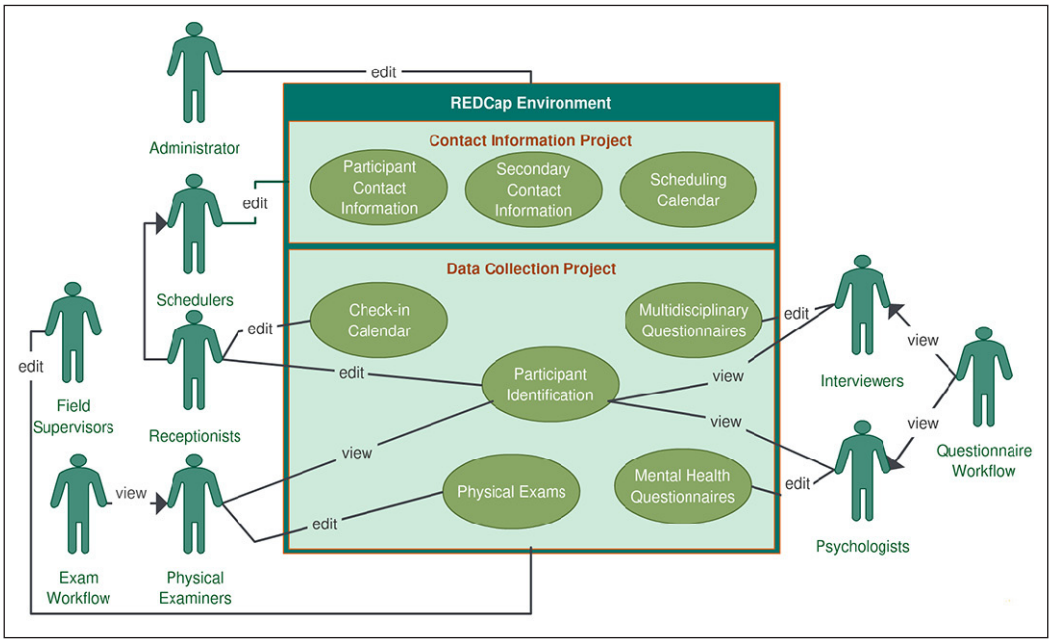

Fig. 1 Use case diagram representing the user groups (dolls) and their access rights (lines) to REDCap projects (rectangles) and instruments (ellipses).

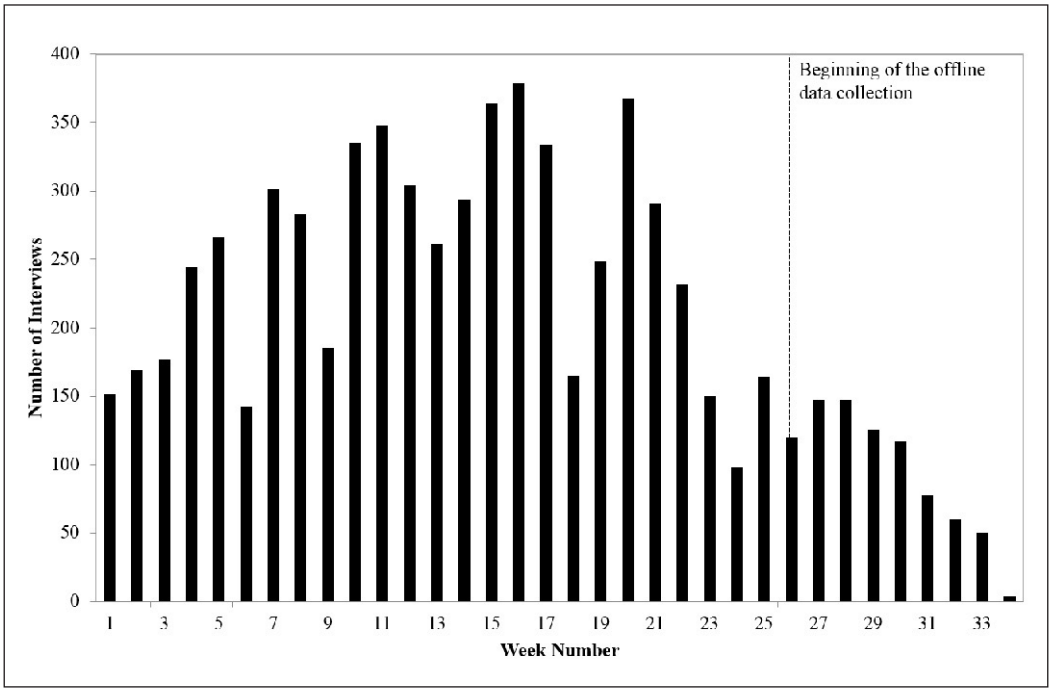

Fig. 2 Number of interviews per study week and beginning of the offline data collection.

Table 1 List of teams and user groups registered in the REDCap projects.

Team	User Group	Number of Users
Scheduling		3
Reception		4
Workflow control	Questionnaire workflow	2
	Exam workflow	2
Data collection	General interviewers	10
	Psychologists	10
	Physical examiners	10
Field supervision		3
System administration		1
Total		**45**

Table 2 List of the general interview questionnaires applied to the cohort members and their mothers.

Cohort 2004 Members	Members' Mothers
School	Adolescent's Care
School Environment	Adolescent's Health
Bullying	Domiciliary Characteristics
Physical Activity	Family Income
Sleeping Quality	Mother's Health
Diet and Coffee Intake	Postnatal Depression
Stressing Events	Parent-Child Conflict
Body Image Perception	Quality of Life
Face/Happiness Scale	Oral Health
Computer and Internet Use	Adolescent's Food Intake
Locus of Control	
Female Health	
Oral Health	

References

1. Grimes DA, Schulz KF. Descriptive studies: What they can and cannot do. Lancet 2002; 359(9301): 145–149.
2. Magnuson JA, O'Carroll PW. Introduction to Public Health Informatics. In: Magnuson JA, Fu PC, editors. Public Health Informatics and Information Systems. London: Springer; 2003. p. 3–15.
3. Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. Lancet 2002; 359(9303): 341–345.
4. Barros AJD, Santos IS, Matijasevich A, Araújo CL, Gigante DP, Menezes AMB, Horta BL, Tomasi E, Victora CG, Barros FC. Methods used in the 1982, 1993, and 2004 birth cohort studies from Pelotas, Rio Grande do Sul State, Brazil, and a description of the socioeconomic conditions of participants' families. Cad Saude Publica.2008; 24 (Suppl. 3): S371–S380.
5. Welker JA. Implementation of electronic data capture systems: barriers and solutions. Contemp Clin Trials 2007; 28(3): 329–336.
6. Hesse BW. Public Health Informatics. In: Gibbons MC, editor. eHealth Solut. Healthc. Disparities. New York: Springer; 2008. p. 109–26.
7. Sarkies MN, Bowles K-A, Skinner EH, Mitchell D, Haas R, Ho M, Salter K, May K, Markham D, O'Brian L, Plumb S, Haines TP. Data Collection Methods in Health Services Research: Hospital Length of Stay and Discharge Destination. Appl Clin Inform 2015; 6(1): 96–109.
8. Cornman JC, Glei DA, Goldman N, Chang M-C, Lin H-S, Chuang Y-L, Hurng B-S, Lin Y-H, Lin S-H, Liu I-W, Liu H-Y, Weinstein M. Cohort Profile: The Social Environment and Biomarkers of Aging Study (SEBAS) in Taiwan. Int J Epidemiol 2016; 45(1): 54–63.
9. Schooling CM, Chan WM, Leung SL, Lam TH, Lee SY, Shen C, Leung JY, Leung GM. Cohort Profile: Hong Kong Department of Health Elderly Health Service Cohort. Int. J. Epidemiol 2016; 45(1): 64–72.
10. Wahabi H, Fayed A, Esmaeil S, Alzeidan R, Elawad M, Tabassum R, Hansoti S, Magzoup ME, Al-Kadri H, Elsherif E, Al-Mandil H, Al-Shaikh G, Zakaria N. Riyadh Mother and Baby Multicenter Cohort Study: The Cohort Profile. PLoS One 2016; 11(3): e0150297.
11. Couper M. Web Surveys: A Review of Issues and Approaches. Public Opin Quart 2000; 64(4): 464-494.
12. 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; 42(2): 377–381.
13. REDCap. Nashville: Research Electronic Data Capture; [updated 2016 May; cited 2016 Apr 17].
14. Santos IS, Barros AJD, Matijasevich A, Zanini R, Chrestani Cesar MA, Camargo-Figuera FA, Oliveira IO, Barros FC, Victora CG. Cohort Profile Update: 2004 Pelotas (Brazil) Birth Cohort Study. Body composition, mental health and genetic assessment at the 6 years follow-up. Int J Epidemiol 2014; 43(5): 1437–1437f.
15. Gonçalves H, Assunção MC, Wehrmeister FC, Oliveira IO, Barros FC, Victora CG, Hallal PC, Menezes, AMB. Cohort profile update: The 1993 Pelotas (Brazil) birth cohort follow-up visits in adolescence. Int J Epidemiol 2014; 43(4): 1082–1088.
16. 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.
17. Bart T. Comparison of Electronic Data Capture with Paper Data Collection–Is there really An Advantage? Bus Brief Pharmatech 2003; 1–4.
18. Pavlović I, Kern T, Miklavčič D. Comparison of paper-based and electronic data collection process in clinical trials: Costs simulation study. Contemp Clin Trials 2009; 30(4): 300–316.
19. van den Berg MH, Overbeek A, van der Pal HJ, Versluys AB, Bresters D, van Leeuwen FE, Lambalk CB, Kaspers GJL, van Dulmen-den Broeder E. Using web-based and paper-based questionnaires for collecting data on fertility issues among female childhood cancer survivors: differences in response characteristics. J Med Internet Res 2011; 13(3): e76.
20. Klipin M, Mare I, Hazelhurst S, Kramer B. The process of installing REDCap, a web based database supporting biomedical research: the first year. Appl Clin Inform 2014; 5(4): 916–929.