Data quality and cost-effectiveness analyses of electronic and paper-based interviewer-administered public health surveys: protocol for a systematic review

Atinkut Alamirrew Zeleke¹, Tolga Naziyo¹, Fleur Fritz² Rainer Röhrig ¹

¹Carl von Ossietzky University of Oldenburg, Department of Medical Informatics, Oldenburg, Germany
²Institute of Medical Informatics, University Münster, Münster, Germany

Corresponding Author:
Atinkut Alamirrew Zeleke, MPH-HI
Department of Medical Informatics
Carl von Ossietzky University of Oldenburg
Ammerländer Heerstraße 140,
26129 Oldenburg, Germany
Phone: +49 (441) 798 - 2871
Email: atinkut.alamirrew.zeleke@uni-oldenburg.de
ABSTRACT

Background: Population-level survey (PLS) is an essential standard method used in public health research to quantify sociodemographic events and support public health policy development and intervention designs with evidence. Though all components can contribute to generating information with right data quality parameters, data collection mechanisms are usually the sources of significant blessing and curse in data quality. The use of electronic devices such as smartphones and tablet computers improve the quality and cost-effectiveness of public health surveys. However, there is lack of systematically analyzed evidence to show the potential impact of electronic-based data collection tool for improving data quality and cost reduction in interviewer-administered surveys.

Objective: This systematic review aims to evaluate the impact of interviewer-administered electronic device data collection methods concerning data quality and cost reduction in PLS compared to the traditional paper-based methods.

Methods: A systematic search will be conducted on MEDLINE, EMBASE, CINAHL, PsycINFO, AMED, Global Health, TRIP, ISI and Web of Science for studies from 2007 to 2018 to identify relevant studies. The review will include randomized and non-randomized studies that examine data quality and cost reduction outcomes. Moreover, usability, user experience, and related user parameters from the same study will be summarized. Two independent authors will screen title and abstract and the quality of the data with third person involvement for disagreement. If the studies are considered to be combinable, meta-analysis will be performed.

Results: The preliminary search in Pubmed and Web of Science search showed 1,491 and 979 hits respectively. The review protocol will be registered with the International Prospective Register for Systematic Reviews (PROSPERO).

Conclusion: The systematic review will inform policymakers, investors, researchers, and technologists about the impact of electronic-based data collection system for data quality and cost reduction. In particular, it will focus on metrics to determine whether electronics data collection system support the data collection process, namely decreasing the error rate, increasing the work and workers efficiency and save costs.
KEYWORDS

Electronic data collection; Demographic and Health Survey; Tablet computer; smartphone

Introduction
Population-level survey (PLS) is an important method of public health research. It helps monitor the sociodemographic events and support policy development and intervention designs with evidence [1]. Most developing countries conduct a census or periodic demographic and health surveys to determine national and regional estimates for population-level epidemiologic indicators and to identify determinants of mortality and morbidity. Though all components of the data collection and management processes impact data quality, the mechanism of data capture is the main determinant of data quality [1-3].

Conducting surveys requires extensive use of paper-pen and manual processes to manage the data collection and reporting processes [4]. It is often a cumbersome task to carry out broader field-based surveys as pre-, during- and post-collection processes require enormous human and material resources. Also, the inherent drawbacks could hamper the quality of the data. Paper data collection processes are labor-intensive, time-consuming, and susceptible to errors; incur high printing and running costs, cumbersome and uncomfortable to field data collectors [5, 6]. This resources demanding nature of paper-based data capturing technique affect the quality of data takes longer time and costly [7, 8].

The growth of information and communication technologies coverage in the last few decades has helped to minimize the challenges encountered in paper-based data collection systems. Implementation of a tablet or smartphone-based data collection is becoming increasingly popular in public health surveys[9, 10]. The potential of electronic data collections systems varies according to their intended area of intervention (disease or healthcare event), country setting, mode of administration (self or interviewer-administered), type of tool ( paper or electronics) and type of research studies (clinical trial or survey). For example, comparison among computer assisted self-interview (CASI) with face-to-face, or telephone interview mode of administration conducted for specific conditions in drug abuse[11], sexual health/HIV[12-14]. The findings showed that CASI tool is a preferable mode to have more significant reporting of potentially stigmatized drug, sex, and HIV risky.
In a clinical trial, the paper-based clinical case report form (CRF) and electronic clinical report form (eCRF) management systems were compared [15-17]. EDC found to be advantageous in broad, low-risk studies and could contribute to improving the data quality and reduction of cost. A recent review also showed that the use of ICT tools like EDC tools for clinical research viewed as cost-effectiveness and means of improving research efficiency and data quality [18].

As a component of eHealth and subcomponent of mHealth, the potential of mobile devices in surveys are noticed in DHS surveys [3, 19] general surveys [20] and longitudinal surveys [21]. The studies that asserted that electronics data collection method has potential to improve data quality, work efficiency and save cost. However, those study the following shortcomings. The studies embedded the impact of the mobile device for data collection in eHealth or mHealth research outcomes. This embedding may compromise the self-standing effect the device for the data collection methods [9, 22, 23]. Therefore the impacts of the electronic data collection mechanism in surveys need to be separately aggregated and reported.

A recent review by Cochrane Collaboration compared the impact of apps and another alternative delivery mode such as paper, laptop computer, and tablet computer for the survey [10]. The review considered only studies which reported self-administered survey methods hence, conducting a systematic review considering interviewer-administered based data collection may complement the evidence. Also, adding interviewer in the interaction of interview tools, the interviewee may influence the quality of data. As to the knowledge of the investigators, there is no systematic review evidence which compared the data quality and cost-effectiveness of an electronic and paper-based interview-administered public health surveys. This systematic review will evaluate comparative evidence of data quality and cost-effectiveness in interviewer-administered public health surveys using electronic and paper-based collection methods. Hence, the proposed systematic review will answer the following questions: What is the data quality evidence available for the data collected by electronics compared to paper-based methods in interviewer-administered public health surveys? What is the cost-effectiveness evidence available for the data collected by electronics of methods compared to paper-based methods in interviewer-administered public health surveys?
Methods
The protocol will be registered in PROSPERO, the International Database of Prospectively Register of Systematic Reviews. This protocol follows the PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 guideline [24].

Study design
We will include parallel randomized controlled trials (RCTs, quasi–RCTs, Controlled Clinical Trials – CCTs and crossover randomized control trial), paired repeated measures, cohort and case-control studies and comparative cross-sectional studies that compared the electronic interviewer-administered survey with paper-based methods.

Study Participants /data
This review depends on the data obtained from the data collectors who used either of the data collection modalities during public health survey. For user preference and usability related information, the potential eligible for this reviews are data collectors/community workers or face to face interviewers who are assigned to undertake a face to face interview with a household resident in the survey area. We will also include data supervisors of the interviewer or data managers for the performed survey included in the reviews.

Types of Interventions
Any mobile device data collection tool which was designed to support interview administered data collection process of public health survey regardless of the setting and country.

Types of Technology
Electronics based data collection in our review refers to portable, wireless digital devices usually supported by networked mobiles or satellite communications infrastructures, such as cell–phones, smart–phones, personal digital assistants, and tablet computers. The support includes data capture and instant/store and forwards submissions to the research center. Within the scope of this review, we will include all applications of such technologies which meant directly supporting the data collection process designed to enable data collectors/interviewer to collect and send data for survey research at the population level or enable supervisors/data manager to monitor the data collection process.
Types of comparisons

Studies comparing mobile device based data collection intervention with the paper-based data collection tools.

Outcomes

The primary outcomes in this review are: Data quality indicators such as data completeness and data accuracy (Accuracy was thereby defined as the absence of typographical errors, decimal point faults, and illogical values). The evaluation includes nonsampling errors resulted from the incorrect recording of the interviewer errors or questionnaire design errors, e.g., missing and inaccuracy. We will compare the proportion of errors or problematic items between two modes for delivering the same survey questionnaire [10]. Cost-effectiveness outcomes will be measured using resource costing method, which includes provider perspective direct cost which compares the cost of conducting a survey using paper and pen method and electronics tools. The secondary outcomes include work efficiency, usability, user experience[19] and user acceptability descriptions.

Study Setting

Geographic location will not limit the setting of the study. All countries and research facilities where interviewer administered electronic device used and compared with paper base tools regardless of the socioeconomic status of the country will be included.

Exclusion Criteria

The following study types will be excluded from the review: All studies which compare electronic and paper-based tools in a self-administered survey; Studies which are performed in settings other than a house to house field surveye.g. electronic medical record, electronic case report form; Studies not performed on human subjects; (4) studies reported before January 1, 2007 (5)Studies with experience report, letters, reviews, commentaries, and editorials; (6) non-English language publications.

Information source

A systematic key-word search will be conducted on electronic databases such as MEDLINE, EMBASE, CINAHL, PsycINFO, AMED, Global Health, TRIP, ISI Web of Science and The Cochrane Library. In addition, reference lists of eligible studies and citations to included articles will be screened for additional eligible published studies. Unpublished and in progress studies
will be identified from the following trial registries: www.clinicaltrials.gov; www.controlled-trials.com; www.anzctr.org.au; www.who.int/ictrp/en/.

The search restriction follows to articles published in English, from 2007 to mid-2018 (devices that became available during this time are compatible with the mobile operating system (OS) framework that focuses on apps[10]).

Search strategy

The search strategy will consider three themes namely: the technology terminologies (e.g. mobile device, mobile phone, mhealth or EDC), data collection terminologies and related terms (e.g. data collection, demographic and health survey or large-scale surveys) and the outcome of interest (e.g. data quality, missing and cost effectiveness). We will connect all the similar terms in the same group with boolean operators “OR” and “AND” (Table 1).

Table 1. Preliminary search hits from PubMed and web of science.

<table>
<thead>
<tr>
<th>Category</th>
<th>PubMed search hit</th>
<th>Web of science</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1. Technology / Intervention</td>
<td>((mobile phone OR cellular phone OR Cell Phone OR cellphone OR smart phone OR smartphone OR tablet device OR Tablet Computers OR Computers, Handheld OR Computer, PDA OR personal digital assistant OR electronics data capture OR EDC OR electronic survey OR eCR OR electronic forms OR eHealth OR mHealth OR Mobile Technology OR android OR Mobile Application OR Mobile Apps OR App, Mobile OR Apps ))</td>
<td>111669</td>
</tr>
<tr>
<td>#2. Area of application</td>
<td>(Data collection OR electronic data collection OR electronic data capture OR data entry OR data capture OR data gathering OR questionnaires OR survey OR health survey OR research OR interview OR demographic OR household survey OR large-scale surveys OR population surveillance OR surveillance OR Demographic Health Survey OR DHS)</td>
<td>3,248,226</td>
</tr>
</tbody>
</table>
**Data management**

The retrieved literature from all databases will be imported to Endnote software for managing duplication and further screening. Covidence web-based screening tool will be used to import the set of de-duplicated citations and to manage the title and abstract screening process. Title and abstract will be screened for the inclusion criteria.

**Selection process**

Two review authors (AAZ and TN) will screen titles and abstracts independently and identify potentially eligible studies based on the eligibility criteria. In case of disagreements, a third author (FF) will be consulted. Full text of the eligible studies will be retrieved and further screened based on the inclusion criteria. We will note the reasons for exclusion and exclusion using flow chart. All authors will discuss and agree on the filtered list of included studies.

**Data Extraction**

Data extraction excel sheet will be used based on the inclusion criteria and the needed items to fulfil the objective of the review. To ensure uniformity across reviewers, we will conduct calibration exercises before starting the review.

**Data items and Extraction**

Two reviewers will independently extract data using customized data extraction forms. The following information will be extracted:

- Author and year,
- National affiliation of author
- Country in which the study was conducted,
• Study design,
• Healthcare/research site setting,
• Target users,
• Size of enumerated population dataset/data elements
• Type of mobile device; delivery mode; application type;
• Stated purpose of intervention
• Range of data quality outcome measures described based on our operational definition for data quality parameters,
• Range of economic evaluation outcomes used to evaluate the cost-effectiveness of EDC
• Types of economic evaluation models or outcomes assessed
• Usability, user experience, and work efficiency outcomes or descriptions
• Key findings from each included study will be summarized and tabulated.

Outcomes and prioritization
The primary outcome of this review is the data quality indicator parameters (error rates for missing (mean number of incomplete records per interview) and inaccuracy (mean number of problematic records per interview) in paper and EDC. The second primary outcome is cost-effectiveness parameters or cost related outcomes. Moreover, usability scale and qualitative user satisfaction indexes will be considered for the secondary outcomes analysis.

Risk of bias in individual studies
Quality will be assessed using parameters such as random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting; and other biases. Each parameter of trial quality will be graded as low, high and unclear risk of bias (in case of sufficient information is not reported) [25]. Agreement of reviewers on methodological quality assessment will be assessed, and discussion will resolve disagreements.

All assessments of study quality will be performed by at least two reviewers (AAZ, TN) with any disagreement resolved by consensus, or mediation with a third reviewer (FF or RR) where necessary.

Data synthesis
The analyzed data will be presented in the tabular and narrative form. Where possible, meta-analyses will be performed on methodologically comparable studies (comparable particularly with regards to the study design, and endpoint measures in the outcomes) reporting main, primary, and secondary outcomes. The choice of statistical tests will depend on the nature of the
outcome variable. Where relevant data are missing, we will contact authors. Where the number of included studies per outcome is sufficient, publication bias will be assessed. If heterogeneity is significant, we will not perform a meta-analysis; a narrative, qualitative summary will be done.

**Ethics and Dissemination**
As only previously published studies are included and reported in the review, no additional formal ethical assessment and no informed consent is required.

**Discussion**
This systematic review will identify and synthesize the available evidence on the data quality and cost-effectiveness outcome of electronics data collections tools deployed for interviewer administered surveys. The evidence supposed to complement the available evidences in the impact of mHealth under different health care common applications particularly for demographic and health care data collections. [26].

**Publication plan**
The systematic review protocol will be registered in the International Prospective Register of Systematic Reviews (PROSPERO) and findings will be summarized in a single manuscript.

**Timeline:** Reporting date 30 June 2018.

**Conflicts of Interest**
None

**Funding**
No external funding has been received for this review.

**Authors’ contributions**
AAZ initiated the idea and wrote the first draft of the protocol. RR and FF critically revised the draft of the protocol. AAZ and TN will conduct the systematic electronic searches, screening and data extraction. AAZ, FF, and RR will conduct the quality assessment, data extraction and review the results.

All authors read and approved the final version of this protocol.
Reference


26. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. (2169-575X (Print)).