Original Paper


Atinkut Alamirrew Zeleke1,2*, Abebaw Gebeeyehu Worku3, Adina Demissie2, Fabian Otto-Sobotka4, Marc Wilken1, Myriam Lipprandt1, Binyam Tilahun2+, Rainer Röhirg1+

1 Carl von Ossietzky University of Oldenburg, Department of Medical Informatics, Oldenburg, Germany
2 University of Gondar, Institute of Public Health, Department of Health Informatics, Gondar, Ethiopia
3 University of Gondar, Institute of Public Health, Gondar, Ethiopia
4 Carl von Ossietzky University of Oldenburg, Department of Epidemiology and Biometry, Oldenburg, 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: +491732587251
Email: atinkut.alamirrew.zeleke@uni-oldenburg.de

+ These authors shared senior authorship
Abstract

Background: Periodic demographic health surveillance and surveys are the main sources of health information in developing countries. Conducting survey in requires extensive use of paper-pen and manual work and lengthy processes to generate the required information. Despite the rise of popularity in using electronic data collection systems to alleviate the problems, sufficient evidence is not available to support the use of electronic data capture tools in interviewer administered data collection processes.

Objective: The purpose of this study was to compare data quality parameters in the data collected using mobile electronic and standard paper-based data capture tools in one of health and demographic surveillance sites in northwest Ethiopia.

Methods: A randomized controlled crossover Healthcare IT-Evaluation was conducted from May 10 to June 3rd, 2016 in demographic and surveillance site. Twelve interview administrator, as two individuals in six groups (one with a tablet computer and the other with paper-based questionnaire), were assigned in the six towns of the surveillance premises. Data collectors switch data collection method based on computer generated random order. Data was cleaned using Mysql program and transferred to SPSS and R statistical software for analysis. Descriptive and mixed ordinal logistic analyses employed. The qualitative interview audio record was transcribed, and the usability of this open data kit (ODK) based system was assessed using system usability scale (SUS) and interview mapping in the isometric dialogue principles for system usability.

Result: From the submitted 1,251 complete records/questionnaires in each tools, 42 % (522) of the paper and pen data capture (PPDC) and 29 % (285) of the electronic data capture (EDC) tool questionnaires had one or more types of data quality errors. The overall error rates were 1.67% and 0.6% for PPDC and EDC respectively. The chances of more errors on PPDC tool were multiplied by 1.015 for each additional question in the interview compared to EDC. The SUS score of the data collectors was 85.6. In qualitative data response mapping, EDC had more positive suitability of task responses with few error tolerance characteristics.

Conclusion: EDC possessed significantly better data quality and efficient compared with PPDC, explained in fewer errors, instant data submission, and easy handling. The EDC proved to be a usable data collection tool in the rural study setting. The consistent power source and decent
internet connection, standby technical support, and security assurance for the device users suggested by the data collectors before full-fledged the implementation of the system.

Key words: Public health, maternal, childhood, surveillance, survey, electronic data capture, tablet computer, mhealth, Ethiopia
1 Introduction

1.1 Scientific Background
In most of the low and middle-income countries, millions of people are born and die without registration in any legal and statistical records. Health and social policies projection performed with assumptions for unrecorded lives [1, 2]. The lack of functional civil registration system in those countries, make them to relay on other data sources such as Health and Demographic Surveillance Systems (HDSS), Demographic and Health Surveys (DHS) programs and facility based sources to understand population-level health determinants [3, 4]. Conducting any survey in those countries requires extensive use of paper-pen and manual processes to manage the data [5]. It is not an easy task to carry out a larger field-based survey as pre-, during- and post-collection processes require huge human and material resources. Also, the inherent drawbacks of paper based system could hamper the quality of the data. Paper Data collection processes are labor intensive, time-consuming, susceptible to errors; incur high printing and running costs and cumbersome and uncomfortable to field data collection [6, 7]. The need to support the paper process and the recent advanced popularity of mobile devices fortified the development and use of electronic data collection methods in community health and clinical research works. Electronics devices; such as personal digital assistants (PDA) [8-10], mobile phones [11, 12] and tablet computers are noticeably tested for their potential role of intent for replacement of the standard paper-based tools [2,3]. The anticipation is that electronic data capturing tools may overcome substantial limitations of the paper-based system regarding saving time, improving the data quality, and minimizing overall cost [6, 7, 9, 13, 14]. Evaluations of such systems require periodic and setting context evidence to support the growing claim on their efficiency, effectiveness, and impacts [15, 16].

1.2 Rationale for the study
The recent emerging research outputs are expediting the use of electronic data collection methods in lower and middle-income countries. However, most of the available evidence cannot surpass the common critics on the quality of mhealth evidence. The critic shares two major points: Firstly, there is no or little field-based study conducted for quantifying the interaction of data quality and data capture technologies. Secondly, the available articles with rigorous
scientific methodologies such as randomized control trials are few [17-20]. The Majority of the research articles are work experience reports, simple descriptive one-arm studies [21-23]. Moreover, the comparative studies are from research conducted at a different point in time [7, 9, 11] or are not field based research but rather carried out in the hospital and clinic settings [10, 12, 24]. There are experiences of using mhealth for data collection in Ethiopia [7, 25, 26]. As to the authors, literature review and expertise, there is a lack of comparative trial evaluation studies on the effect of data capture tools on the quality of the data in surveys conducted in demographic and surveillance sites.

Therefore, in this study, we evaluated mobile electronic data capture tool and compared it to the traditional standard paper-based system to increase the quality of recorded responses from respondent rural community households.

1.3 Objective

The purpose of this study was to compare data quality parameters in the data collected using mobile electronic and standard paper-based data capture tools in Dabat Health and demographic surveillance sites in northwest Ethiopia.

The study’s primary research question was

- Is the error rate of data collected by an electronic data capture tool less compared to paper and pen recorded data?

The secondary questions were:

- Is there a learning effect in the use of EDC through the course of the data collection period?
- Is an electronic mobile based data collection tool usable to the data collectors in Dabat health and demographic surveillance sites?

2 Study context

2.1 Organizational Setting
This evaluation research implemented at Dabat Health and Demographic Surveillance System (HDSS), also called the Dabat Research Center (DRC) in northwest Ethiopia. The surveillance site is a member of the INDEPTH global network of HDSSs which have 42 health research centers as members and 47 HDSS field sites in 18 low- and middle-income countries (LMICs) in Africa, Asia and Oceania [27]. The surveillance center established in 1996 aims to generate community-based representative evidence through a continuous longitudinal data collection process. The surveillance runs by the College of Medicine and Health Sciences, which is one of the colleges/faculties of the University of Gondar in Ethiopia. Dabat district, the surveillance site, is one of the 21 districts in North Gondar Administrative Zone of Amhara Region in Ethiopia. According to the INDEPTH Network, Dabat HDSS is working with 69,468 participants [28].

2.2 System details and system in use

A 30 pages paper questionnaire, in eight sub-thematic sections, was developed by a team of researchers at the Institute of Public Health at the University of Gondar, Ethiopia. The comprehensive survey called “integrated survey on maternal, childhood, nutrition, and disability in Dabat Health and Demographic Surveillance System (HDSS).” The themes addressed pregnant women, young females, children less than five years of age and people with disabilities. The number of questions in each sub-thematic questionnaire varies from 15 to 63, and each interviewer can use at least one of them or a combination of sub-thematic questionnaires based on the respondent's category of cases. Most of the overall items (88.25%) were closed-ended questions.

Open Data Kit (ODK) was chosen to develop the EDC system using the open-source application. ODK has an open suite of tools, used to design the form, collect, and aggregate the data. Similar experience observed in a couple of recent studies [7, 29]. The exact copy of the paper questionnaire converted to its electronic replica form using Excel and XLSForm. After passing the technical validity, the form was uploaded to the server. Each data collector can get a new blank form at the first time with an authorized login to the server. For multiple-choice questions, one question per screen is displayed on the tablet screen, whereas for tables based arranged questions many questions per screen populated to the screen. Only respondents households, ID, personal ID, and data collectors and their supervisor name were required items. Though there
were skipping, calendar date and number only function incorporated in the system. No other error controlling functionalities were incorporated to avoid unnecessary confusions between the two data collectors. The electronic system might induce discrepancies while both are interviewing a single person at the same time.

The data entry fields were restricted to option buttons, check boxes, or empty fields (Fig 1).
Figure 1. Screenshot examples of the type of questions (Multiple choices (A), Number (B), Single select (C) and Date (D)) presented in EDC tool used for the survey in Dabat Demographic and Health Surveillance site, northwest Ethiopia. June 2016
After completing each interview, the data collector can use ‘Edit Saved Form’ function for any valid correction and can use ‘Send Finalized Form’ function to send the form to the server instantaneously. The third generation (3G) mobile internet network was used to connect each tablet computer to the server. In the case of limited network connectivity, submission of the saved done when the data collectors are in a good network coverage area.

Locally, in Ethiopia, manufactured tablet computer with an in-built Amharic language font called TechnoPhantom7 used for this research. The battery life of the tablet computer was approximately 48 hours. We have restricted the functionalities of many applications in the devices to longer battery lifetime. To test the natural course of electricity infrastructure in the areas, we have not used extra battery or power banks as a reserve power source to charge the tablets.

3 Methods

3.1 Study design

We conducted a field-based, prospective, randomized, controlled, crossover trial to investigate the error rate as an indicator of data quality. Moreover, usability evaluation using system usability scale (SUS), and semi-structured questionnaire(S1 File) interviews were conducted to see the user impression in use of EDC.

The evaluation study started after the approval by the Ethical Committee of the University of Oldenburg (vote-no. 148/2016, chair Prof. Griesinger) and the ethical vote of the University of Gondar, Research & Community Service Gondar, Ethiopia (vote-no. O/V/P/RCS/05/501/2015, signed by T.T. Adefris). The data kept confidentially and the electronic copy of the data stored in the Department of Medical Informatics of the University of Oldenburg

3.2 Study Participants

Oral consent was obtained from the participants and data collectors after the objective of the research explained briefly. Data collectors, who permanently employed to Dabat Demographic and Health Surveillance site and contractually employed for this integrated survey, were our research participants. The data collector’s main task was to travel house to house and do a face-
to-face interview with the persons living in selected Kebeles (smallest administrative unit) of Dabat district.

From thirteen towns chosen for the overall survey, six towns selected for our specific research based on accessibility of internet coverage and electric power supply in the town or nearby towns. Based on this, 12 data collectors who assigned to work in six towns were involved. The participants are randomly selected to have either tablet computer or paper questionnaire based on computer generated orders of tool users in the study period.

### 3.3 Study flow

We used the bigger survey, meant to address larger population for a relatively long time, (approximately four months) to evaluate our system as nested in it. The evaluation trial started after the technical team at the University of Gondar department of health informatics in Ethiopia managed the form creation, ODK Collect app installation on the tablet and server configuration. Following, we trained data collectors for two days on the basic tablet computer use and EDC functionalities. After incorporating comments from the participants on the functionality of the system, we performed pre-tests of the system for two days other than our study area. The pre-test phase showed some critical deficits of the system, and we have learned unseen system errors, such as unnecessarily required items which significantly delaying the data collection time. They were removed, and many decimal points items were corrected. Mini user manual using the local language was prepared and given to the interviewer along with working procedures.

Each interview was conducted with two data collectors one who administered the interview and entered data using either a tablet computer or a paper questionnaire; a second one only entered data on either of the tools. In a given interview, one of the data collectors leads and asks the questions, but both of them fill in the data independently. None of the data collectors knew in advance their order of data collection methods. Every morning their supervisors instructed the data collectors order of tool use (PPDC or EDC) based on computer generated random order of tool users to switch the device between the interviewers in the same group.

The actual data collection period was from May 10 to June 3rd, 2016. At the end of the field based data collection period, the usability questionnaire was given to each data collectors.
Moreover, their satisfaction and opinion during the system use were asked by using semi-structured interview questionnaire (S1. File).

### 3.4 Outcome measures or evaluation criteria

Overall data quality errors (missing and inaccuracy): The existence missing or inaccuracy errors regardless of the tool they used for data collection. The overall data quality errors calculated the existence of one and more errors from the total records (how many of the records have one or more items with errors from the total records). For the pen and paper questionnaire, we used the original responses recorded by the data collectors before data entry and cleaning or correction from the supervisor.

**Missing value:** An absent value of response of in a given questionnaire.

**Inaccuracy:** Any problematic items or incompatible values, which include, decimal point errors, invalid date, or text values unreadable values.

**Quantitative: SUS:** 10 Item questionnaires, the system considered usable if the overall SUS score is > 67 from (0 to 100) [30].

**Qualitative isometric analysis based on ISO 9241-110 dialogue principle:** the seven dialogues based on the description in the studies [31, 32]. Transcribed interview response was related to the six dialogue principles and summarized with a “+” and “−” notation if the interview response fits positively or negatively respectively.

### 3.5 Methods for data acquisition and measurement

The data collectors with paper questionnaire handover completed questionnaires to their respective supervisors daily. To identify the potential errors in a given paper-based questionnaire, the supervisors used error extraction sheets and recorded all identified errors before giving correction, comments, and suggestions back to the data collectors. For EDC system, the completed form/questionnaires were directly submitted to the server. The administrator privileged to access and monitor the process daily checked submissions from all sites and replay for any technical inquiries by the data collectors with other standby technical members.

Electronically submitted data was downloaded on a weekly basis in Excel format from the server and sent to an actual research site where the counter paper-based questionnaire remained. Two independent health informatics technical assistants compared the data errors recorded on the
extraction sheets, and double checked each item in the questionnaire for further error identification and noted the errors in electronically submitted data.

Anonymized data from the study brought to a department of the Medical Informatics at the Carl von Ossietzky University of Oldenburg. MySQL was used for cleaning and preparation of data for analysis by two researchers.

To evaluate the software usability, we used the Amharic language translated System Usability Scale (SUS) questionnaire [30]. The ten-item questionnaire is based on a five-point Likert scale and produces a maximum score of 100 on the user’s impression on the electronic data capture tool. We chose the SUS because of its shortness and the understandable phrasing of questions even for non-IT skilled persons like our data collectors. The results of our SUS questionnaire showed a Cronbach Alpha of 0.841 for reliability.

We also used semi-structured questionnaires to perform face-to-face interviews with data collectors who were involved in this trial. Tape-recorded and noted data then transcribed for further analysis based on the dialogue principles, defined in the ISO 9241-110 [31].

### 3.6 Methods for data analysis

Univariate analyses, to quantify errors in data capture tools, performed by IBM SPSS Statistics software (Version 24.0 IBM, USA) and R-statistical software used to perform ordinal logistics mixed regression model to see the effect of some variables on the quality of the data. As the data set contains two values for each observation, a number of errors for the tablet questionnaire and a number of errors for the paper version, there are a lot of ties or dependencies between the observations. To reduce these ties, we constructed a new variable as the difference between the two numbers of errors. However, this variable contains many zeros. We chose to resolve this problem by categorizing the variable into the following three categories: fewer errors in the paper version than in the tablet version, zero difference, and more errors in the paper version than on the tablet. We assumed a dependence between the error rates and the overall length of the questionnaire as well as differences in performance between groups of six interviewers. Consequently, we constructed a mixed ordinal regression model to explain the category of error differences along the length of the questionnaire. The model contains random effects for the individual performance of groups of interviewers. At the same time, we estimated models that
specialized on the influence of the different types of questions and their quantities. For every model, we used a regression equation with the following structure:

\[
\text{cat Error} \cdot b_1 \text{covariate} + a_{G1} \text{covariate} + \ldots + a_{G6} \text{covariate} + e
\]

“Covariate” stands for the different measures of the length of the questionnaire and \(a_{G1}\) to \(a_{G6}\) denote the random effects for each group. For the estimation, we used a cumulative logistic mixed regression model. Estimation was performed within the software R using the add-on package “ordinal.” We tried to include additional covariates in a forward selection process though they were eliminated on the basis of the Akaike Information Criterion (AIC). For each regression coefficient we can construct a test for statistical significance of the influence of the corresponding covariate. We may test the hypotheses \(H_0: |\beta| = 0\) vs. \(H_1: |\beta| \neq 0\) by constructing a confidence interval for the level \(1-\alpha\) and check whether it includes the value from \(H_0\). Hence, we can conclude whether we find a significant increase in the error rate in either of the two methods.

For qualitative analysis, two computer scientist experts and one public health informatics background professional independently mapped the qualitative responses to six of the seven dialogue principles, and then later approved the category together. When there is a difference arise a physician and medical informatics expert consulted for an agreement.

The dialogue principles of ISO 9241-110, namely: Suitability for the task, suitability of learning, controllability, self-descriptiveness, conformity with user expectation, and error tolerance used to map the transcribed interview result. To include many of the respondent views we have expanded the dialogue principles, mainly suitability for the tasks, focussed from the software and hardware context to include social and environmental contexts.

4 Results

4.1 Data collectors Sociodemographic

At the end of our data collection period, a total of 1,246 respondent’s data recorded by interviews using conventional paper and pen data capture tools whereas 1,251 interview records observed in the electronic database. During data cleaning, we found out that five of the submitted by EDC
were empty. The counter check-up revealed that those data were also not found in the paper counter of the same group.

Twelve interviewers, four male and eight female and age, ranges from 21 to 32, participated in the study. Among these, four nurses, three information technologists, one social science background person by their profession, the rest of the data collectors were with no specific background in their profession. Regarding their educational status, the majority (8/12) have completed high school and vocational schools certificate or diploma. Their experience as a field worker ranges from two months to seven years with a mean of 2.6 years. Four of the interviewers have experience of using smartphones as a personal phone. None of the interviewers have previous experience of using a smartphone or tablet computers as a data collection tool. Regarding their educational status, the majority (8/12) have completed high school and vocational schools certificate or diploma.

4.2 Unexpected events during the study

Though six of the groups of data collectors included in the study period, unfortunately, one team terminated the study after two weeks because of an incident of personal conflicts in the area. This conflict causes the termination of data collection in that particular area as the situation could not be resolved during the data collection period. Therefore, we have completed the study with the remaining five groups.

4.3 Study findings and outcome data

4.3.1 Errors per questionnaire

From the complete 1,246 submitted questionnaires in both PPDC and EDC tools, 522 (42%) of the PPDC questionnaires had one or more errors. From this, the majority with 175 (33.5%) of the questionnaire had only one error followed by 112 (21.4%) questionnaires with two errors and three 55 (10%).
At EDC, 385 (29%) of the questionnaires had one or more errors. Thus, the majority of 241 (62.5%) had only one error followed by 76 (19.7%) questionnaires with two errors and 41 (10%) with three errors (Fig 2 and S1 Table).

Figure 2. Frequency of error comparison among EDC using the tablet computer and paper and pen data capture (PPDC) tools during a survey in the Demographic survey site in 2016, Dabat Northwest Ethiopia.

4.3.2 Error rate

The overall error rate, computed from the total error count over the total number of asked items, was 1.14%, from which 0.73% was missing, and 0.4% of the rate was inaccuracy errors. The paper and pen data capture error rate was 1.67 (missing 0.92% and inaccuracy 0.75%) while the electronic data capture error rate was 0.6% of which 0.54% was missing and 0.064% inaccuracy errors (Table 1).
Table 1. Error rate by types of error and the tools used during a survey in Demographic survey site in 2016, Dabat Northwest Ethiopia.

<table>
<thead>
<tr>
<th>Error type</th>
<th>Overall</th>
<th>Asked</th>
<th>Paper</th>
<th>Asked</th>
<th>Items</th>
<th>Tablet</th>
<th>Asked</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items</td>
<td>22,110</td>
<td>11,069</td>
<td>6</td>
<td></td>
<td></td>
<td>11,041</td>
<td>5*</td>
<td></td>
</tr>
<tr>
<td>Errors count (%)</td>
<td>Errors count (%)</td>
<td>Errors count (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>1,620 (0.73%)</td>
<td>1,020 (0.92%)</td>
<td>600 (0.54%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inaccuracy</td>
<td>901 (0.40%)</td>
<td>830 (0.75%)</td>
<td>71 (0.064%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All errors</td>
<td>2,521 (1.14%)</td>
<td>1,850 (1.67%)</td>
<td>671 (0.60%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NB * The item difference results from the extra asked items in the paper questionnaire due to asking items that have to be skipped accordingly. This applies to other tables, too.

4.3.3 Error rate over time/learning effect

Though the figure shows no smooth pattern of mean error rate increasing or decreasing over the study time, there is a visible difference observed between the two tools regarding the error rate. The EDC has a constant error rate swinging below 1%, whereas the overall error rate and PPDC move around slightly above 1% and 1.5%, respectively. There were random peaks of the error rates at different points of time in the study period; the overall error rate is not showing a constant trend of decreasing or increasing over time (Fig 3.)
Figure 3. Means of the overall error rates trend of Electronics data collection tool (EDC) using tablet and paper and pen data capture (PPDC) tools used during the survey in Demographic survey site in 2016, Dabat Northwest Ethiopia.

4.3.4 Frequency of Errors per item type

Looking at the overall frequency of errors nearly half (47.4%) of the errors in PPDC were found in questions which had multiple option answers followed by questions with single choice answer (25.2%). At EDC, two-thirds of the errors shared by multiple options, number and date questions type with 33.6% and 30.1%, respectively.

Looking at the relative error rate, in the paper questionnaire, questions, which had multiple answer options had a relatively higher rate of errors with 3.65%, followed by free text and the number and date answers with 3% and 1.9%, respectively. In EDC, questions with free text options had a relatively higher error rate with 2.1%, followed by numbers and date options answers (Table 2).

Table 2 Error by item type and the tools used during a survey in demographic survey site in 2016, Dabat Northwest Ethiopia.

<table>
<thead>
<tr>
<th>Item types</th>
<th>Paper Total Asked Items</th>
<th>Number of errors (rate %)</th>
<th>Electronics Total Asked Items</th>
<th>Number of errors (rate %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single select</td>
<td>46,660</td>
<td>466 (0.99%)</td>
<td>46,427</td>
<td>145 (0.31%)</td>
</tr>
<tr>
<td>Single select tabular</td>
<td>19,062</td>
<td>84 (0.44%)</td>
<td>19,062</td>
<td>42 (0.22%)</td>
</tr>
<tr>
<td>Multiple selects</td>
<td>24,303</td>
<td>877 (3.64%)</td>
<td>24,302</td>
<td>226 (0.92%)</td>
</tr>
<tr>
<td>Numbers and dates</td>
<td>18,022</td>
<td>343 (1.9%)</td>
<td>17,980</td>
<td>202 (1.12%)</td>
</tr>
<tr>
<td>Free text</td>
<td>2,644</td>
<td>80 (3.0%)</td>
<td>2,644</td>
<td>56 (2.1%)</td>
</tr>
<tr>
<td>All item errors</td>
<td>110,691</td>
<td>1,850 (1.67%)</td>
<td>110,415</td>
<td>671 (0.60%)</td>
</tr>
</tbody>
</table>
4.4 Regression results
In each estimated model, we find that the chance for a higher number of errors in the paper version (in comparison to the tablet version) increases with the length of the overall questionnaire. For each additional question in the interview, the chances of more errors by using the paper tool were multiplied by 1.015. We have obtained similar findings for the different subsets of the questionnaire, where we find an increase in the chances for additional errors in the paper version in each type of the questions. The increase is slower for single item questions. The increase is strongest in the time and date questions with an odds ratio of 1.084 per additional question. None of the 95% confidence intervals include an estimate of zero and an odds ratio of one, respectively (Table 3). Hence, we can see that each variable measuring a number of items has a significant effect on the comparative error rate.

Table 3. Mixed model effect, ordinal logistic regression for the data collected by EDC and PPDC tools used during a survey in the Demographic survey site in 2016, Dabat Northwest Ethiopia.

<table>
<thead>
<tr>
<th>covariate</th>
<th>Regression coefficient b_1</th>
<th>Exp (b_1)</th>
<th>SD</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of items</td>
<td>0.015</td>
<td>1.015</td>
<td>0.0035</td>
<td>(0.008;0.02)</td>
</tr>
<tr>
<td>Number of single items</td>
<td>0.029</td>
<td>1.030</td>
<td>0.0074</td>
<td>(0.014;0.04)</td>
</tr>
<tr>
<td>Number of multiple items</td>
<td>0.042</td>
<td>1.043</td>
<td>0.0094</td>
<td>(0.024;0.06)</td>
</tr>
<tr>
<td>Number of single table items</td>
<td>0.023</td>
<td>1.023</td>
<td>0.011</td>
<td>(0.001;0.04)</td>
</tr>
<tr>
<td>Number of time and date items</td>
<td>0.081</td>
<td>1.084</td>
<td>0.017</td>
<td>(0.048;0.114)</td>
</tr>
</tbody>
</table>

4.4.1 System Usability Scale test

The usability of the system was measured based on the System Usability Scale (SUS). The global scores can range between 0 and 100, with 100 reflecting the highest usability. The result yielded individual SUS scores between 67.5 and 100 and a total score of 85.6. The graph for
each SUS questions depicted that majority of the data collectors response categorized from “agree” to “strongly agree” for positively articulated questions. Likewise, “strongly disagree” to “disagree” for the negatively articulated questions (Fig. 4 and 5).

Figure 4. Usability response for negatively articulated questions in SUS used during a survey in Demographic survey site in 2016, Dabat Northwest Ethiopia.
Figure 5. Usability questionnaire response for positively articulated questions in SUS used during a survey in Demographic survey site in 2016, Dabat Northwest Ethiopia.

4.4.2 Qualitative interview: ISO 9241 - 110 dialogue principles mapping
According to the interviewee’s response, the advantage of paper-based systems expressed by the perceived limitation of the EDC tool. Accordingly, being familiar with the PPDC tool, easy to correct mistakes manually, being free from technical deletion errors, not dependent on electricity and power supply for operation, were referred as the advantage of PPDC tool. Thus, EDC has less conformity with user expectation, controllability, and suitability of task. In the same view, PPDC being prone to unreadability errors during correction of respondent answers, questionnaires become out of stock in the field paper being heavy and easily wasted during
improper handling as well as storage and incurring higher duplication and operational costs for larger surveys were considered as a disadvantage of PPDC. Thus, the issues listed above were mapped more error tolerance and suitability of task of EDC, respectively.

The benefit of EDC tool explained as increasing efficiency data collection and the possibility of instant submission of the data to the research center and then helps to monitor the real time working progress. Also, the EDC device being small and easy to handle, the inclusion of automatic functions to reduce potential errors and the opportunity of introduction to the EDC technology to improve the IT skill of the data collectors are considered as a benefit. Hence, EDC possesses good suitability of task, error tolerance, controllability, and suitability of learning.

The inability to edit the electronic form in case of correction in the questionnaire, losing data with unintentional deletion errors, the difficulty of writing in keyboards and dependent on a constant supply of power and internet and hence affected by frequent interruptions of electricity and internet connections listed as a limitation of EDC. Those EDC characteristics mapped to less conformity with user expectation, error tolerance, and suitability to the task.

The interviewee views further mapped based on aspects explored during the interview which includes, efficiency/speed, data recording and entry processes, data management/operation, logistics and operation, concern, learning, infrastructure, reported and solved work interrupter challenges, satisfaction, and confidence and future improvement (Table 4).

Table 5. Data recorders perceptions and its isometric dialogue mapping, for data capture tools used during a survey in Demographic survey site in 2016, Dabat Northwest Ethiopia.

<table>
<thead>
<tr>
<th>Aspect explored</th>
<th>Response</th>
<th>EDC Dialogue mapping (+/-)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency/speed</td>
<td>Using a paper questionnaire, I can collect more data, as it is more familiar to me and easy to write on it“</td>
<td>ST+</td>
</tr>
<tr>
<td></td>
<td>Working with paper saves time, on paper, you can swap pages while working easily, ask the questions and fill it page by page finish the interviewer and store it... but with EDC, it takes time for saving or sending the data.</td>
<td>C –</td>
</tr>
<tr>
<td></td>
<td>I work relatively fast when I use a tablet computer, and this can save time as long as it is free from an internal software error.</td>
<td>ST+</td>
</tr>
<tr>
<td>Data recording and entry processes</td>
<td>In the paper questionnaire, deletion or edition in case of mistakes or incorrect answer can cause unreadability, in tablet we can easily delete or edit when it is necessary and can be free of such errors.</td>
<td>ET+ &amp;ST +</td>
</tr>
<tr>
<td></td>
<td>With the paper, we may forget the skip question pattern, but in EDC there is automatics skip function.</td>
<td>C/SD +</td>
</tr>
<tr>
<td>Data management/operation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Written work can stay longer, and it cannot be easily destroyed as in the case of format or delete in EDC. | • ST- & ET  
| Retrieving of files in EDC may be difficult or impossible after deletion, but it is easy with paper. | • ET-  
| In the paper, if there is a mistake in the items we can correct it later, but in EDC we do not have a chance to correct later as the data is sent instantly to the server. | • ST+  
| There is a possibility data loss as a result of paper torn out or wasted during inappropriate handling or seasonal variation in environmental conditions, but in EDC such problems are not frequent. | • ST+  
| The managers can daily get the completed questionnaire through EDC instantly; this may help to control the actual work progress of the data collectors. | • ST+  
| Logistics and operation                                                                 |  
| The paper questionnaire runs out while we are in the field and sometimes it takes longer to reach for us, that is not the case in EDC. | • ST +  
| The tablet is small and easy to handle, and there is the potential to work more with easy to handle device like this. However, paper questionnaire pack is heavy, and bulky carrying paper packs for long distance is difficult and affects our physical activity, and also we feel tired during writing. | • ST +  
| Concern                                                                                   |  
| As we are working in rural areas alone with this expensive device, we feel insecure that the device might be forcefully stolen or the data recorders may get attacked in the way of taking the device. | • ST-  
| Learning                                                                                 |  
| Our skill of using IT device will increase day to day as we become familiar with the system. | • SL+  
| Infrastructure                                                                           |  
| The battery capacity lasts a maximum of a one and a half day, during the study period overnight charging of the tablet computer in or around the research town was possible. | • ST+  
| I appreciate paper questionnaire if there is termination of work because of tablet battery is low or switched off or when the mobile data connection is interrupted. | • ET-  
| Internet collection was sufficient to send the data once in a day during the study period | • ST+  
| Sometimes in deep rural areas electricity may not be available to charge the battery or the internet network may not be available. | • ST-  
| Reported and solved work                                                                  |  
| The save function of ODK: Replace the actual data collection date with the data edition date and time. | • SD& C-  
| Downloading a new form at the beginning of the study or form updating was challenging.   | • SD& C-  
<p>|</p>
<table>
<thead>
<tr>
<th>challenges</th>
<th>Deleting partially filled data was challenging.</th>
<th>• C -</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Some questions were refusing to accept the point’s character.</td>
<td>• CUE -</td>
</tr>
<tr>
<td></td>
<td>Unintentionally touch some buttons, change on screen resolution settings, date and time: unable to communicate with the server.</td>
<td>• C -</td>
</tr>
<tr>
<td></td>
<td>Mobile Internet data balance may be finished while we are in the remote area.</td>
<td>• ST -</td>
</tr>
<tr>
<td>Satisfaction and confidenc e</td>
<td>Have better satisfaction and confidence to work using this device in the future.</td>
<td>• ST +</td>
</tr>
<tr>
<td>Future improvemen t</td>
<td>The questionnaire details such as point, number, legal value, should be well refined in EDC before implementation.</td>
<td>• CUE -</td>
</tr>
<tr>
<td></td>
<td>As frequent phone consultation for technical help may not be possible in case of network problems, strong training is needed to troubleshoot simple errors by EDC users without technical team consultation.</td>
<td>• –</td>
</tr>
</tbody>
</table>

*= + means more and – means less
ST, Suitability of task; SL, suitability for learning; SD, self-descriptiveness; ET, Error tolerance; C, controllability; CUE, Conformity with user expectation; SI, suitability for individualization

4.5 Unexpected observations

Most of the interviews went smoothly except a few incidences of loss of internet connectivity, unexpected power loss, and software stacks. Recording geographic coordinates were part of the EDC questionnaire as a required input. The required option had created a substantive delay in data collection process at three groups, as they could not ask further questions without detecting the GPS signals; we have removed the requirement of GPS coordinated, and the data collection went back to normal in three of the tablets.

5 Discussion

5.1 Answers to study questions. Magnitude of errors

This study aimed to evaluate and compare the magnitude of data quality errors among paper and pen and electronic data capture technologies and the usability of the tablet computer to capture survey respondent data.
In this study, data collected by electronics pointed out to have fewer errors compared to the conventional paper and pen data collection. Nearly half of the PPDC and one-third of EDC questionnaires submitted exhibited one or more errors in the respondent answers. This kind of measurement may show inflated error occurrence and can be misleading to report higher magnitude. Thus, computing error rate, which considers the total number of questions, is important in bulky questionnaires like ours. For example, the error rates for PPDC and EDC tools were 1.67%, and 0.6%, respectively. This shows most of the submitted questionnaires with errors have very few errors from the total asked items.

The low rate might be from two possible effects; ‘being observed effect’ or ‘inter-tool interaction.’ The effects were unavoidable due to our study design nature. The former effect might be from the extra care taken by the data collectors while working. They may feel that their performance watched indirectly. The inter-tool interaction between the paired data collectors might affect the performance of one data collector over the other. The unwanted communication might neutralize the data quality of one tool over the other. Our finding supported this claim where three paired groups have similar magnitudes of errors.

SUS scores and impression of the data collectors from the interview showed high usability and satisfaction in using EDC for data collection with a usability score of 85.4. Data collectors stated that the system is easy to carry, able to directly submit the data to the server and let them be familiar with the information technology; this encourages the implementation of such systems in resource-limiting settings.

The qualitative responses triangulated with isometric dialogue principles. The conceptual focus of these dialogues is basically about the hardware and software capability system for given task. However, we tried to force some of the dialogue principles to accommodate environmental and social contexts which are of particular importance to the research sites settings. For example, responses such as, “we have a constant supply of electricity and internet,” “I work relatively fast when I use tablet computer” and “When I use a tablet, I will not worry about finishing questionnaire categorized under suitability for task dialogue principle. The responses are software neither interface issue nor hardware issues; rather they are setting context to facilitate the data collectors’ tasks. This may imply that there could be room to extend the dialogue principles on the importance local context issues where the system evaluation takes place. Though mapping the qualitative interview cannot explicitly show all the dialogue principles into
the level of equal share, relevant information can be drawn even with its limitation. For the future, a qualitative research design where the questions intended to include all of the six dialogue principles can yield a better understanding of the system usability.

The potential incidence of theft of the electronic device and subsequent physical attacks in an attempt to steal the device or fearing of replacement if the device is lost were frequently mentioned concerns by the data collectors. The source of the concern positions the fact that the routine data collection activities require the data collectors, mostly females’ cadres, to travel a very long distance alone and visits the house of respondents, which is far apart from each other with minimum security. Insecurity resulted from system user may create unintended resistance. If the system planned for scaled up implementation in the routine work, addressing their concern is important.

5.2 Strengths and weaknesses of the study
This study adopted an existing open source platform to develop the electronic questionnaire and data submission and storage configuration. This showed we could easily adopt the available resources and save the human and material resources to develop the new system from scratch. The research design in this study is a cross over randomized trial, which helps to eliminate unnecessary learning trends and show the real effect of the devices accordingly.

The other possible strength of this study is the fact that the survey showed the usability of the system using the quantitative (SUS) and the qualitative interview supported by isometric dialogues principles.

Pairing data collectors for a single interview may affect the inter-data collectors’ performance, or tools may affect each other, hence lower the magnitude of actual errors. Inability to implement all the functions of the electronics error controlling mechanism, such as required items for all, value range determination in EDC might have contributed to some preventable errors that were existed in EDC.

5.3 Results in relation to other studies
The overall error rate in this study, 1.67% for PPDC tool and 0.6% is comparable with a study in Kenya which had 1% and 0.1% missing rate in paper and electronics, respectively [11]. In addition to that, a study with a similar design in India revealed with 2% PPDC and 1.99% EDC error rates [29]. However, this magnitude is lower compared to other similar studies (7% and 1%,
with no omission in EDC) [9], 3.6% PPDC and 5.2% in EDC[33]. The difference might be an item selection mechanisms: The above studies calculated their error rate from selected items in a given questionnaire while this study considered all items in the questionnaire, which is highly likely to inflate non-errors denominators for such kinds of typical huge number of asked items (over 22,000). Furthermore, random assignation of data capture tools for data collectors, while the studies compare data collected from two different points of time, by various data collectors and non-randomized methods. Non-randomized studies might possess inflated error influenced by non-homogeneous factors distributions among data collectors and other characteristics of the study.

Many EDC studies that use the maximum potential of error control mechanisms such as required items for all, logical skip, or logical input values in EDC, reported zero to less 1% missing responses [14, 34]. In our study, because of the design nature, we could not fully implement all the error-controlling mechanisms in EDC to avoid unnecessary confusion, waiting time between the two data collectors while working together. Under possible circumstances, implementing maximum potential of error preventing mechanism in the software are mandatory.

Our study and other related findings revealed there is statistically significant difference in the chance of having more errors in paper-based records than electronic data records [7, 9, 11]. However, an equivalent research report from rural India claims that there was no statistical difference in error magnitude between the two tools, where EDC is as a curate as PPDC [29]. In our study, the chance of having more errors increased as the number of asked questions increases. Our study was unique in depicting this finding, due to the nature of the survey, which has integrated all round survey where the number of the questions and the type of the questionnaire depends on the kinds of cases interviewed by the data collectors in the households. This shows the consideration of the total the number of in a questionnaire might be important when choosing the type of data capture tools during design, as bulky questionnaire might be suitable for electronic records than a paper based system.

The SUS score value 85.4, and the qualitative interview Isometric dialogue mapping shows the system users, in our case data collectors, use the system comfortably. The findings shared by other EDC user’s impressions in related researchers [11, 26, 29, 33, 34]. A task-technology fit study claims that users understand their system as tools helping or hindering performing their
tasks. Users react positively to system’s features that appreciate their mission related demands[35].

5.4 Meaning and generalizability of the study

Our study findings are in line with similar studies that used ODK app platform to develop the electronic form and the data management settings and implemented in rural and urban community household surveys, though there is a methodological difference. The quantitative and qualitative analysis depicted the EDC is usable and preferable data capture tool for the field based survey.

We believe the study might be generalized to rural community setting research sites, with limited internet and electricity access.

5.5 Unanswered and new questions

Working in a pair might put ‘inter-supervisor’ effect to the data collector’s performance. In a further study, we can assign one data collector independently to work with EDC. In this study, a single data collector had a cumulative two weeks exposure in the four weeks study period. In this short time exposure, it may be too early to measure the learning effect through time. Moreover, research considering organizational readiness evaluation, working culture alteration and its outcome measurement in using this system in resource-limited settings are tasks on the table. Further research is also necessary to validate qualitative based isometric usability questions in different linguistic and communication culture settings.

6 Conclusion

The objectivist approaches of this study conclude that data collected by using an electronic based data collection system had a significantly better quality compared to paper-pen data collection system explained by fewer errors. Implementation of electronic data collection tools, like tested in this comparative trial, was found to be usable by data collectors in the rural resource limited settings. Implementation of full error controlling function exhaustively, setting standby technical and monitoring team, assuring security concerns on the device will contribute to better implementation of the electronic data collection system in the resource limited-settings.
Stakeholders of the health information system particularly in demographic and surveillance site can adapt and use the existing open source mobile device platforms in their routine data collection and management practices.

7 Author’s contribution

AZZ had the idea and designed this evaluation study. He participated in the development of the EDC tool and performed data collection. He analyzed the data and prepared the first draft of the manuscript.
AGU, participate development of the general proposal, supervise data collection, revised the draft manuscript.
ADB, participated in the data collection and analysis, revised the draft manuscript.
FS, participated in the data analysis and revised the draft manuscript.
ML, MW participated in the data analysis and categorization of the interviews, and revised the draft of the manuscript.
BT initiated and conceptualized the overall project and the design and the development of the EDC tool, participated in the data collection and analyzed and revised the draft. He lead the overall project.
RR, supervised the study design, data collection and analysis, and manuscript writings and revised the final version of the manuscript.

8 Competing interest

The authors declare that they have no competing interests.

9 Acknowledgement

We would like to thank the University of Gondar for funding the research expenses. We also would like to acknowledge Mr. Temesgen Azemeraw ( Data manager in Dabat HDSS) for his
technical contribution in ODK form development. We would also like to thank Mr. Brendan Whelan (Dublin) and Stefanie Gacek for Copy Editing the manuscript.

10 Reference

12. M. Ganesana SP, Vincy Pushpa Mary, N. Janakiramana, Ashok Jhunjhunwalab,Nuwan Waidyanath. The Use of Mobile Phone as a Tool for Capturing Patient Data in Southern Rural Tamil Nadu, India. Journal of Health Informatics in Developing Countries. 2011;5(2).

Supporting information
S1 Table. The frequency of error in electronics data capture and the paper data capture tools.
S1 File. A semi-structured questionnaire used to interview the system users.