The development and evaluation of a PDA-based method for public health surveillance data collection in developing countries

Ping Yu, Maximilian de Courten, Elaine Pan, Gauden Galea, Jan Pryor

School of Information Systems and Technology, University of Wollongong, Wollongong, 2522, Australia
Department of Epidemiology and Preventive Medicine, Monash University, The Alfred, Commercial Road, Melbourne, VIC 3004, Australia
Royal Prince Alfred Hospital, Sydney, Australia
Health Promotion, World Health Organization, Geneva, Switzerland
Fiji School of Medicine, Fiji

Background and purpose: EpiData and Epi Info are often used together by public health agencies around the world, particularly in developing countries, to meet their needs of low-cost public health data management; however, the current open source data management technology lacks a mobile component to meet the needs of mobile public health data collectors. The goal of this project is to explore the opportunity of filling this gap through developing and trial of a personal digital assistant (PDA) based data collection/entry system. It evaluated whether such a system could increase efficiency and reduce data transcription errors for public surveillance data collection in developing countries represented by Fiji.

Methods: A generic PDA-based data collection software eSTEPS was developed. The software and the data collected using it directly interfaces with EpiData. A field trial was conducted to test the viability of public health surveillance data collection using eSTEPS. The design was a randomised, controlled trial with cross-over design. 120 participants recruited from the Fiji School of Medicine were randomly assigned to be interviewed by one of six interviewers in one of the two ways: (1) paper-based survey followed by PDA survey and (2) PDA survey followed by paper-based survey. Data quality was measured by error rates (logical range errors/inconsistencies, skip errors, missing values, date or time field errors and incorrect data type). Work flow and cost were evaluated in three stages of the survey process: (1) preparation of data collection instrument, (2) data collection and (3) data entry, validation and cleaning. User acceptance was also evaluated in the two groups of participants: (1) data collectors and (2) survey participants.

Results: None of the errors presented in 20.8% of the paper questionnaires was found in the data set collected using PDA. Sixty-two percent of the participants perceived that the PDA-based questionnaire took less time to complete. Data entry, validation and cleaning for the PDA-based data collection from 120 participants took a total of 1.5 h, a 93.26% reduction of time from 20.5 h required using paper and pen. The cost is also significantly reduced with PDA-based protocol. Both data collectors and participants prefer to use PDA instead of paper and pen.

A R T I C L E   I N F O
Article history:
Received 23 June 2008
Received in revised form 5 March 2009
Accepted 5 March 2009

A B S T R A C T
Background and purpose: EpiData and Epi Info are often used together by public health agencies around the world, particularly in developing countries, to meet their needs of low-cost public health data management; however, the current open source data management technology lacks a mobile component to meet the needs of mobile public health data collectors. The goal of this project is to explore the opportunity of filling this gap through developing and trial of a personal digital assistant (PDA) based data collection/entry system. It evaluated whether such a system could increase efficiency and reduce data transcription errors for public surveillance data collection in developing countries represented by Fiji.

Methods: A generic PDA-based data collection software eSTEPS was developed. The software and the data collected using it directly interfaces with EpiData. A field trial was conducted to test the viability of public health surveillance data collection using eSTEPS. The design was a randomised, controlled trial with cross-over design. 120 participants recruited from the Fiji School of Medicine were randomly assigned to be interviewed by one of six interviewers in one of the two ways: (1) paper-based survey followed by PDA survey and (2) PDA survey followed by paper-based survey. Data quality was measured by error rates (logical range errors/inconsistencies, skip errors, missing values, date or time field errors and incorrect data type). Work flow and cost were evaluated in three stages of the survey process: (1) preparation of data collection instrument, (2) data collection and (3) data entry, validation and cleaning. User acceptance was also evaluated in the two groups of participants: (1) data collectors and (2) survey participants.

Results: None of the errors presented in 20.8% of the paper questionnaires was found in the data set collected using PDA. Sixty-two percent of the participants perceived that the PDA-based questionnaire took less time to complete. Data entry, validation and cleaning for the PDA-based data collection from 120 participants took a total of 1.5 h, a 93.26% reduction of time from 20.5 h required using paper and pen. The cost is also significantly reduced with PDA-based protocol. Both data collectors and participants prefer to use PDA instead of paper and pen.
paper for data collection. The trial results prove that eSTEPS is a feasible solution for public health surveillance data collection in the field. Several deficiencies of the software were also identified and would be addressed in the next version.

Conclusion: eSTEPS offers the potential to meet the need for an effective mobile public health data collection tool for use in the field. The eSTEPS field trial proves that PDA was more efficient than paper for public health survey data collection. It also significantly reduced errors in data entry. The later benefit was derived from the software providing its users with the flexibility of building their own constraints to control the data type, range and logic of data entry.

1. Introduction

Despite the numerous publications describing the potential advantages the personal digital assistant (PDA) may offer compared with traditional paper and pen based health data collection method [1–5], only nine studies were identified by Lane et al. to satisfy the rigorous, randomised controlled design for the evaluation of effectiveness [6]. Wu and Straus also commented that there was limited evidence of the effectiveness of handheld electronic medical records after only identifying two studies that satisfied the inclusion criteria for their systematic review [7]. Shirima et al. conducted a study at the point of data collection for large household survey in remote areas of rural southern Tanzania [8]. They found that the use of PDAs eliminating the usual time-consuming and error-prone process of data entry and validation and increasing data quality. However, the validity of this study is questioned because no direct comparison with paper-based method was conducted. It was suggested that further research was required on the use of PDA in different populations, particularly in comparison with paper-based practice [8].

2. Background

Due to a lack of sufficient empirical evidence to support the use of PDA, the common practice of large public health surveillance data collection remains using paper and pen, the data are then entered into computer by specially trained data entry staff at a centralised office [1,5,8–11]. The World Health Organization (WHO) STEPwise approach to surveillance of risk factors for non-communicable diseases (NCD) is not exceptional to this practice, despite its well-developed and highly standardised nature. The majority of its responses are gathered through structured multiple choice questions.

As data entry is essentially duplicated in paper-based surveys, first as data collected on paper, then entered into computer, the data collection/entry task is not only time-consuming, but also subject to a range of data errors and discrepancies commonly encountered in this process. In the WHO STEPwise approach, the creation of the data entry template for the survey started from the creation of the questionnaire file (.QES file) in Open Source software EpiData (ref to www.epidata.dk). The questionnaire file was then printed as a paper-based questionnaire survey tool. The data collector used this tool to collect data in the field. The electronic version of the questionnaire file (.QES) was converted into record file (.REC file) in EpiData so that the data collected in the field could be entered into this file. A check file (.CHK file) could also be generated to check the validity of data entered to ensure that out-of-range errors can be automatically detected, with alerts provided to the data entry staff.

Improving quality of data through reducing errors in data collection and entry is essential for the success of any large public health survey. Data acquisition and entry into central electronic data storage simultaneously by electronic means is an important strategy to achieve this goal. As data collectors need to go from door to door to collect NCD surveillance data in the WHO NCD STEPwise survey, a web-based or desktop computer based solution is not suitable to use [12]. Tablet personal computers (Tablet PCs) were also considered too expensive and heavy to carry for this purpose. Previous research findings suggest that handheld computers (PDA) are small, light-weight devices that can be easily carried around; and that data collection/entry using PDA is an effective alternative to paper and pen modes of data collection [12–17]. It is found to be favoured by health care providers over the paper and pen mode of data collection in areas of documentation, medical reference, and access to patient data [18].

3. The aims of the study

The aims of this study were to determine which type of data collection system is more effective in public health survey data collection. The objectives of the research are: (1) to develop a PDA-based public health survey data collection tool; (2) to evaluate the cost–benefit of this PDA-based data collection tool in comparison with a paper-based method; and (3) to provide evidence to assist in deciding which type of data collection system should be used in future WHO NCD STEPwise surveys.

4. Methods

4.1. The development of eSTEPS

An object-oriented, component-based approach was undertaken to develop the PDA-based data collection software eSTEPS. The freeware EpiData software package is utilised in conjunction with eSTEPS (ref: http://www.epidata.dk). A number of Microsoft software packages, all of which are available at no cost to the user, are used. Microsoft ActiveSync and Microsoft .NET Framework must be installed in the desktop

A sample size of 120 participants was considered to be sufficient (i.e., at a 95% confidence level and 80% power, assessment of the same data entered by two different computers). While none of the selected data collectors had any previous experience with PDAs, they were all regular users of desktop PC operating systems. A Pocket PC handhold device (PDA) running the Pocket PC 2002 SDK and its later version is required for the eSTEPS Pocket PC component to function. Microsoft .NET Compact Framework is required on PDA before the eSTEPS Pocket PC component will function correctly. All these software packages are available free of charge from the Microsoft web site (ref: http://www.microsoft.com/).

4.2. The study design for the randomised controlled trial

A comparison between the two data collection methods, the PDA-based and the paper-based method, was conducted using randomised controlled trial with cross-over design incorporating a 7-day wash-out period between data collection methods. The data was gathered from records auditing, observations, satisfaction (questionnaire) survey and structured interviews. Following a 2-day training activity covering the knowledge of NCD-STEPS and sampling methods with both paper-based and eSTEPS methodologies, a pilot data collection activity was conducted to ensure that the data collectors were competent in both data collection methods. Data collection for the NCD-STEPS questionnaire took place as arranged by the individual data collectors in approximately 4 weeks in June–July, 2004.

The same “core” items from the NCD-STEPS Instrument Version 1.4 (http://www.who.int/chp/steps/en/) were used in both paper-based and PDA-based questionnaires. The order of data entry in the PDA version of the questionnaire was exactly the same as that in the paper-based instrument. As mentioned above, the PDA-based interview questionnaire was developed on a PC using the eSTEPS questionnaire designer. The questionnaire was then downloaded to a PDA with Microsoft Pocket PC operating system installed. The PDA-based interview also had additional features to enable just-in-time data verification including consistency checks, range checks and automated skips based on entered data.

4.3. Study participants

There were two types of study participants: (1) data collectors: those collecting the NCD-STEPS interview data; and (2) survey participants: those from whom the interview data were collected. Both the data collectors and the participants were recruited from students at the Fiji School of Medicine (FSM) in Suva, Fiji. Six data collectors were recruited through advertisement on FSM notice boards and through the student email list. While none of the selected data collectors had any previous experience with PDAs, they were all regular users of desktop computers.

Using a sample size calculation for self-paired variables (i.e. assessment of the same data entered by two different collection methods) at a 95% confidence level and 80% power, a sample size of 120 participants was considered to be sufficient. Subsequently, 120 participants were recruited wherein each of the six data collectors recruited 20 participants on a voluntary basis from FSM students.

4.4. Variables to be measured

Variables of interest were both quantitative and qualitative, and included items related to data quality, work flow, logistical issues, related costs, and both the data collectors’ and participants’ acceptance (see Table 1).

4.5. Data collection

Following a 2-day training activity covering the knowledge of NCD-STEPS and sampling methods with both paper-based and eSTEPS methodologies, a pilot data collection activity was conducted to ensure that the data collectors were competent in both data collection methods. Data collection for the NCD-STEPS questionnaire took place as arranged by the individual data collectors in approximately 4 weeks in June–July, 2004.

The same “core” items from the NCD-STEPS Instrument Version 1.4 (http://www.who.int/chp/steps/en/) were used in both paper-based and PDA-based questionnaires. The order of data entry in the PDA version of the questionnaire was exactly the same as that in the paper-based instrument. As mentioned above, the PDA-based interview questionnaire was developed on a PC using the eSTEPS questionnaire designer. The questionnaire was then downloaded to a PDA with Microsoft Pocket PC operating system installed. The PDA-based interview also had additional features to enable just-in-time data verification including consistency checks, range checks and automated skips based on entered data.

Of the five PDAs used in the study, four were ViewSonic Pocket PCs and one was a Hewlett-Packard (HP) IPAQ h1940.
4.6. Participant randomisation

The interviewers were provided with a Data Collector Tracking Form that was used in the recruitment of participants, with each interviewer having the responsibility to recruit and collect data using both methods for each of 20 participants. The initial data collection method for the first participant for each interviewer was randomly chosen and assigned. Subsequent participants had their initial data collection method alternate between paper-based and PDA-based such that one-half of the participants began with one method and the other half began with the other. Per the cross-over design of the study, following a 7-day wash-out period, data collection was repeated for each of the participants using the data collection method not used in the first instance. Strict adherence to the cross-over design and wash-out period were managed through centralised scheduling of all activities.

Also, to limit the level of knowledge about the survey content prior to the first data collection session for each participant, instructions were given to all study individuals, both interviewers and interviewees, to not divulge details of the survey or the study to anyone else until the entire data collection period was over. This instruction was reinforced regularly throughout the data collection period.

4.7. Data entry and processing

The data collected on the paper-based forms were entered into a data entry template designed with Epion version 2.1b that included automatic skips and range checks, and a limited capacity for consistency checks. This was conducted by the interviewers who undertook special training. A second round of data entry and validation was done using Epi Info version 6.04d (www.CDC.gov). This protocol for data entry/validation from paper-based questionnaires is currently standard practice for NCD-STEPS in WHO and is designed to identify and eliminate data entry errors.

The data collected through the PDA did not require any further data entry or validation. However, the verified dataset from both methods underwent a separate phase of consistency and range checks as part of a final data cleaning activity.

4.8. Time measurement

Time for various tasks was measured using log files which documented starting and ending times for the various process tasks measured against the quantity of work accomplished (e.g., the time in hours and minutes needed to enter data from a certain number of paper-based questionnaires).

4.9. Data analysis

A statistical comparison of error rates was made between the two data collection methods in each of the areas as listed in Table 1 for the data quality variables. p-Values of less than 0.05 were taken as being statistically significant for this and all other statistical calculations.

A comparison of mean durations was made between the two data collection methods with differences tested as appropriate using analysis of variance calculations. Also, an analysis was undertaken to investigate any differences in mean durations of interview times relative to the order in which the data collection methods were used for each participant.

Responses to the participant self-administered evaluation forms were analysed to compare the two methods. Emerging themes from the interviews of data collectors and other study personnel were identified for semantic associations.

4.10. Ethical considerations

Submission of the study proposal was made to the Fiji National Health Research Committee for their endorsement. Subsequently, ethical approval of the study was also secured through the standard application process with the Fiji National Research Ethics Review Committee.

5. Results

5.1. The PDA-based data collection software eSTEPS

The eSTEPS software, as shown in Fig. 1, consisted of three major components: eSTEPS questionnaire designer (Desktop), eSTEPS Manager (Desktop) and eSTEPS Pocket PC (Pocket PC). A new questionnaire could be created, or an existing one is edited using the eSTEPS questionnaire designer in a desktop computer; then being exported to an EpiData.qes file. This EpiData.qes file could then be converted into an EpiData.rec file. Using the eSTEPS Manager, this EpiData.rec file could be exported to a PDA. The data collector entered the answers into the relevant fields in the questionnaire on the PDA using the eSTEPS Pocket PC program. The answers were saved to an eSTEPS.rml record file. The records in the .rml file could then be exported from the PDA to the EpiData.rec file using the eSTEPS Manager. This can occur directly from the PDA, or the .rml file can be transferred to a computer where the records are located.

With eSTEPS, the paper-based NCD-STEPS questionnaire is computerised on a PDA, allowing data collection and data entry to occur simultaneously during administration of the interview by survey personnel in the field. Data is first stored on the PDA and later uploaded to a host computer, which contains the NCD-STEPS database in EpiData format. For the purpose of reducing data errors, inconsistencies and data discrepancies, eSTEPS also includes the functionality to provide real-time logical range and consistency checks, guide interviews with predefined routing (skips) and provide error messages.

The following section will discuss the results of the randomised controlled trial comparing the PDA-based and paper-based data collection methods.

---


---
5.2 Data quality

As stated in Section 4, data quality is measured by errors in data entry. Twenty-five of the 120 paper-based questionnaires (20.8%) had at least one error requiring a decision or action by a data analyst. In total there were 31 errors in these 25 paper-based questionnaires. The types of errors include: (1) 11 questionnaires (9.2%) had logical range errors in 14 different fields; (2) 6 questionnaires (5%) had skip errors; (3) 6 questionnaires (5%) had missing values; (4) 2 questionnaires (1.7%) had a missing time field and 2 questionnaires (1.7%) had a missing date field; and (6) 1 questionnaire (0.8%) had an incorrect data type.

As range checks, skips and required fields were confined at real-time by the program automatically, none of the PDA-based questionnaires had these errors. The difference in error rates between paper-based and PDA-based data collection is thus significant (p < 0.001).

5.3 Work flow of NCD STEPwise data collection

The activities for NCD STEPwise data collection include preparation of data collection instrument; data collection in the field; and data entry, validation and cleaning in the centralised office.

5.3.1 Preparation of data collection instrument

The paper-based data collection instrument took 9 h to prepare. It covered the tasks of (1) designing the actual paper-based questionnaire from a previous NCD-STEPs template; (2) designing the EpiData and Epi Info database templates to accommodate double data entry; and (3) printing, collating, stapling, transporting and distributing the paper-based questionnaires. In comparison, the preparation of the PDA-based instrument took 6.5 h, 2.5 h less than the paper-based method to prepare. It comprised of three tasks: (1) designing the questionnaire and EpiData database using the PC-based eSTEPS Editor; (2) converting the questionnaire into EpiData compatible form; (3) exporting the completed questionnaire to the PDAs through the software eSTEPS Manager; and (4) distributing the PDA-based instrument to the data collectors.

5.3.2 Data collection

According to the cross-over study design, the same questionnaire survey was repeated for each participant, once on paper, then on PDA, or vice versa. As the participant would have heard all the questions the second time, it is not surprising that data collection from participants took longer on average the first time than the second time. This was true whether the paper questionnaire was used first (p < 0.001) or whether the PDA was used first (p < 0.001). As noted in Table 2, this longer average time to complete the survey the first time using either data collection method, was also confirmed by the participants’ subjective perception of which method took longer to use.

However, the cross-over design of the study provides some insight into which data collection method actually took longer to use. While the mean duration for the paper-based questionnaire was only marginally longer (1.1 min) than that of the PDA-based questionnaire (p = 0.300) when being used the first time, it was found that the paper-based questionnaire took significantly longer (3.0 min) as the second method used compared to the PDA-based questionnaire (p = 0.001). Furthermore, as a whole, the mean duration of 13.8 min (n = 120) to complete the paper-based questionnaire was significantly longer than the mean duration of 11.7 min (n = 120) to complete the PDA-based questionnaire (p = 0.015). This suggests that data collection takes less time with the PDA.

The above findings were reinforced by the perception of 62.4% of the 117 participants (n = 73) that the PDA-based questionnaire took less time to complete than the paper-based one. Notably, as shown in Table 2, for the 37.6% of the participants who felt that the PDA-based questionnaire took longer (n = 44), all but one of these (2.3%) occurred when the first method used for interviewing the participant was PDA-based. In other word, all but one participant was happy with the speed of data collection when PDA was used in the repeat data collection. This forms a sharp comparison with 16 participants (27.12%) who were not happy with the speed of data collection when paper was used in the repeat data collection.

5.3.3 Data entry, validation and cleaning

Data entry, validation and cleaning for the paper-based data collection from the 120 participants took a total of 20.5 h. It included the tasks of (1) data entry using the prepared Epi-Data template; (2) repeat data entry and validation using the prepared Epi Info template; and (3) completion of all checks for range errors, skip errors and missing data.

In comparison, data entry, validation and cleaning for the PDA-based data collection from the 120 participants only took a total of 1.5 h. This included the tasks of (1) uploading and merging the files from the PDAs to the PC (30 min) and (2) completion of all checks for missing data (60 min).

Table 2 – The survey participants’ responses about which data collection method takes longer to complete (in comparison with data collection method used first).

<table>
<thead>
<tr>
<th>Data collection method used first</th>
<th>Number of participants reporting the data collection method as being more time-consuming</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper-method (n = 73, 62.4% of participants)</td>
<td>PDA-method (n = 44, 37.5% of participants)</td>
</tr>
<tr>
<td>Paper-method</td>
<td>57</td>
</tr>
<tr>
<td>PDA-method</td>
<td>16</td>
</tr>
</tbody>
</table>

5.4. User acceptance

As mentioned in Section 4, user acceptance was considered for both data collectors and survey participants. The reasons behind the responses given were also explained.

5.4.1. Data collectors’ acceptance

All but one of the data collectors were either “quite confident” or “extremely confident” in their use of the PDA. They all expressed a similar level of satisfaction with the level of participant interaction using the PDA-based method, as well as with its speed and ease of use. Although the level of satisfaction among the data collectors using PDA were greater than those using paper, the small number of data collectors (six people) is not adequate to draw conclusive result.

Four of six data collectors expressed a preference for the PDA-based data collection method primarily for its advantage of automated skips at real-time of data collection. The other two data collectors had “no preference” and, while highlighting certain advantages of the PDA-based data collection method, they also noted advantages of the paper-based data collection method (e.g. reliability and ability to more easily enter “out of range” values).

5.4.2. eSTEPS technical issues that may attribute to the data collectors’ acceptance

It is important to note that there were three instances of PDA failures during the field data collection, two among data collectors who expressed “no preference” and one among a data collector who preferred the PDA. It is likely that the PDA failures led to these two data collectors’ attitude of “no preference”. Two of these instances involved the PDA turning off without warning, and the third instance involved a “frozen” PDA that could not be re-set. The former two failures required the PDA be re-set, resulting in a loss of all installed programs, and the need to reinstall the eSTEPS program and to restart the PDA-based data collection for the involved participant. The third failure was resolved by swapping the memory card into a spare PDA. Due to the fact that all saved data was stored on memory cards, there were no instances in which a PDA failure resulted in the loss of stored data.

The HP platform had a significantly faster processor as indicated by a much shorter period required for initial loading of the eSTEPS questionnaire.

The PDA failures indicate some instability in the overall eSTEPS platform but it is unclear as to whether this was due to the software alone or whether it was perhaps influenced by the relatively slower processor speed in the ViewSonic Pocket PCs.

Another point to note is the requirement to reinstall the eSTEPS program in the event of the need to re-set the PDA due to a failure. This may be solved by installing the eSTEPS software onto the memory card rather than in the PDA’s in-built memory. It should technically be achievable and could be provided in a further revision of the software. Another specific problem with the eSTEPS software was also identified: it allowed duplicate records to be produced during uploading from the PDA to the PC because the original record on the PDA was kept instead of being erased. This issue could also be resolved in a further revision of the software.

5.4.3. Participants’ acceptance

The PDA-based and paper-based data collection was equally acceptable by the participants. A considerable majority (71.4%) of the participants (p = 0.000) expressed a preference for the PDA-based data collection method, primarily citing its ease of use, speed and the perception that its use improved the confidentiality of the data being collected. Only 12.9% of the participants expressed a preference for the paper-based method, and 13% had no preference for either method.

A significantly greater proportion of the participants (p < 0.001) felt more comfortable in being interviewed with the PDA (94.4%) compared to being interviewed with the paper-based questionnaire (77.1%). However, a similar proportion of participants (p = 0.083) felt that the data collector was equally confident in use of the PDA (92.7%) and the paper-based questionnaire (85.3%).

5.5. Cost

Using PDA for public health surveillance data collection includes the following cost: (1) the production or purchase of the software; (2) desktop computers and PDAs; (3) preparation of the questionnaire; (4) training of questionnaire designer and data collectors; (4) data collection; and (5) data entry, validation and cleaning.

5.5.1. The production of the software

The software was programmed by a group of the third-year computer science students as their software project at the University of Wollongong under the support of a software consultant and the first author. Therefore, no salary was paid to the development team, although the WHO provided funding to sponsor the cost of the consultant and the travel cost for the first author to travel twice from Australia to Fiji for consultation with the end users. Desktop computers were needed whether the survey is PDA-based or paper-based.

5.5.2. Cost of PDA or production and transport of paper-based questionnaires

Table 3 lists the costs associated with either PDA-based or paper-based data collection for a standard sample size of 2000 questionnaires. If the cost of PDAs can be depreciated in 3-year term, and a PDA is used in repeat surveys in the 3 years, its depreciated cost per questionnaire would be close to negligible.

5.5.3. Preparation for survey

Preparatory tasks for the paper-based or the PDA-based survey relate to the development of the relevant data collection instrument and making it ready for the initiation of the data collection process. Although it takes slightly longer to prepare the paper-based survey instrument as compared to the PDA-based method, this difference is fairly minimal assuming this task would be accomplished by an experienced data professional in Fiji at a cost of US$7.50 per hour. The 9h of preparatory work for the paper system would thus cost US$67.50, marginally more than the cost of US$48.75 to cover the 6.5h of preparing for the PDA-based survey (Table 3).
Table 3 – Comparison of the relevant costs by data collection method.

<table>
<thead>
<tr>
<th>Item</th>
<th>Paper-based</th>
<th>PDA-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparation of data collection instrument</td>
<td>$67.50</td>
<td>$48.75</td>
</tr>
<tr>
<td>Printing of questionnaires</td>
<td>$990</td>
<td></td>
</tr>
<tr>
<td>Purchase of PDAs and memory cards</td>
<td>$993.80</td>
<td></td>
</tr>
<tr>
<td>Training of 10 data collectors</td>
<td>$200</td>
<td>$250</td>
</tr>
<tr>
<td>Data collection process</td>
<td>$1820.00</td>
<td>$1670.00</td>
</tr>
<tr>
<td>Data entry and validation process</td>
<td>$2284.36</td>
<td>$31.96</td>
</tr>
<tr>
<td>Data cleaning process</td>
<td>$312.50</td>
<td>$125.00</td>
</tr>
<tr>
<td>Total</td>
<td>$5674.36</td>
<td>$3129.51</td>
</tr>
</tbody>
</table>

Assumptions: Standard sample size of 2000; data cleaning by an experienced data professional at US$7.50 per hour; data entry personnel hired at US$3.85 per hour; data collector field cost at US$10/day/person; depreciation of PDAs and memory cards over 3 years.

5.5.4. Training of the questionnaire designer and data collectors

The cost of training personnel to use eSTEPS questionnaire designer to design the questionnaire was not able to be measured in this study, because this role was undertaken by the third author of this paper, who also undertook the role of system analyst and tester for the software development.

The PDA-based data collection method requires some additional training in the use of PDAs. The standard training activity for the paper-based method required approximately 1 day. It was observed that the use of the PDA requires approximately another half-day of a comprehensive training.

5.5.5. Data collection

As stated in Section 6.2, the PDA-based method gives a marginal advantage of 11.7 min to collect information from one participant compared to 13.8 min using the paper-based method. This equates to a decrease in data collection time of approximately 15%. As ViewSonic Pocket PC was found to be much slower than HP IPAQ, this is likely to be an underestimate if a PDA with a faster processor is used.

Based on the daily salary of US$10 per day for a data collector in Fiji, such a person using the paper-based method for an actual NCD-STEPs survey could complete approximately 11 questionnaires daily, and ten such data collectors could collect data on 2000 participants in approximately 18.2 days at a cost of US $1820 in Fiji. This compares to data collectors using the PDA-based method who could complete approximately 12 questionnaires daily, with ten such data collectors collecting data on 2000 participants in approximately 16.7 days at a cost of US $1670. The comparative advantage of the PDA-based method relative to the actual data collection process equates to less than 2 days of time or US $200 in costs.

5.5.6. Data entry, validation and cleaning

As mentioned above, the data entry and validation process is essentially eliminated with the PDA-based method, and that the data cleaning process is greatly simplified. The cost summary listed in Table 3 also suggests that the PDA-based survey underpinned by the eSTEPS technology provides a cost-efficient alternative to the paper-based questionnaire.

The data entry and validation process for the paper-based questionnaire in an actual NCD-STEPs survey would take between 8 and 9 weeks at a cost of US $2284.36 with data cleaning taking approximately another week, at a further cost of US $312.50 for a total cost of US $2596.86 (based on the hourly salary rate of $3.85 for a data entry personnel). In contrast, the PDA-based method would require approximately 1 day of data entry plus 2 days of data cleaning time at a total cost of US $156.96. This equates to a 94.0% reduction of cost shifting from the paper-based method to PDA-based protocol.

6. Discussion

The goal of eSTEPS is to provide a mobile data collection tool that is convenient to use in field data collection. This tool directly communicates with EpiData, an existing desktop application widely used by public health agencies in data entry, documentation and error detection. Together with another popular Open Source software Epi Info [19], EpiData are often used by public health agencies around the world, particularly developing countries, to meet the needs of public health data management. eSTEPS thus provides a promising solution to fill the need for a mobile component compatible with current Open Source public health data collection technology.

We have compared the cost-benefits of eSTEPS-based data collection with the paper-and-pen method in a public health survey data collection process in voluntary student participants in Suva, Fiji. The similar study conducted by Shirima et al. drew similar conclusions about the use of PDAs eliminating the usual time-consuming and error-prone process of data entry and validation and increasing data quality [8]. There were several similarities between these two studies: (1) both were targeted at large size public health data collection; (2) both used software programs to build automatic control mechanisms into the questionnaire to prevent errors, such as logical checks, routing patterns and alerts; (3) training of interviewers proved straightforward in both studies; (4) error rates of data were low, validating and cleaning of the full dataset was much faster with PDA compared with paper method; and (5) there was a very high level of acceptance by both the interviewers and the survey participants.

Shirima et al. realised that the big limitation of their study was that there was no direct comparison with a paper-based method [8]. In contrary, our study takes a comprehensive randomised controlled trial with cross-over design. It provides sound, empirical evidence about the advantages of the PDA-based method for public health surveillance data collection in developing countries, Int. J. Med. Inform. (2009), doi:10.1016/j.ijmedinf.2009.03.002
based data collection in comparison with the paper-based method. Thus our study complements the study conducted by Shirima et al. [8]. We will compare the results of our study with that of the previous researchers for each attribute we have evaluated.

6.1. Data quality

Our measure of data accuracy by error incidence is not only similar with that used by the previous researchers [6], but also advanced in its comprehensive coverage of the types of errors that could occur in survey data collection. Of the six studies reviewed by Lane et al. [6], only two were qualified for inclusion in the review of the effectiveness of handheld computers versus paper method for data collection [20,21]. Lane et al. found the data collection via PDA to be more accurate [6]. In three studies accuracy was similar between the PDA and the paper method [22–24]. Lane et al. thus suggest that the PDA-based method does not necessarily bring in improvement in data quality [6]. The improved data quality is likely attributed to the carefully structured questions that only allow determinate types of responses and the use of prompts to ensure that questions are followed in sequence and cannot be skipped in these PDA-based data collection protocols [20,21]. As eSTEPS allowed all these constraint functions, such as logical range checks, skips, required fields and alerts, to be programmed by an end user, it effectively eliminated these errors. This outcome of error reduction supports the previous researchers’ prediction [20,21].

Lane et al. also suggest that PDA-based data collection methods will not result in greater accuracy of data if the source of error is in data collection rather than in data entry or transmission, or if the performance of the paper method is already high [6]. In our case, if the errors made by a data collector were not in the category that could be prevented or detected by the program’s built-in control functions, they would remain as for the paper-method; and these types of error cannot be detected later on. However, eSTEPS effectively eliminated various human errors made in the field by the constraints set up by an end user; therefore, a PDA-based data collection method underpinned by built-in data control functions could lead to greater data accuracy than the paper-based method. This outcome is also reported in the survey conducted by Shirima et al. [8]. In contrast to the suggestion made by Lane et al. [6], the results of our study and that of Shirima et al. suggest that if effective preventive functions are built into the PDA-based data collection tool, they can significantly improve accuracy of data [8].

The integrity of data is vulnerable to data entry errors [25]. Through eliminating the process of entering data recorded on paper into computer, Cole et al. suggest that PDA-based data entry could significantly reduce or eliminate data entry errors [26]. Our research findings and those of Shirima et al. [8] support this recommendation.

6.2. Work flow

Lane et al. suggest that differences between the handheld and paper and pencil instruments in data entry, handling and transfer times were seldom evaluated [6]. For the first time, this study gives detailed evaluation of work flow in data collection.

The preparation of the PDA-based data collection instrument took significantly less time in our study as the third author undertook this duty in this study. However, in the normal situation of public health data collection, the personnel undertaking this duty may need much more training. Once properly done, this activity can lead to significant reduction of human errors in the field and reduction in time for data cleaning [8].

As found by Shirima et al. another practical advantage of the PDA-based method is it eliminates the need to transport, account for and archive large volumes of paper forms [8]. However, there does not appear to be much difference in time and cost between the PDA and paper-based method in the actual data collection.

Increased efficiency at the data entry and validation stage appears to be substantial for the PDA-based data collection method compared to the paper-based one. Lal et al. found that the use of electronic instruments reduced data entry and transfer time by 23% [20]. Shirima et al. also reported that the preliminary results of their survey of 20,636 households were presented to the local leaders 2 days after the completion of the survey, a speed that can never be achieved using paper method [8]. As concluded by Lane et al., handheld computers performed better in timeliness of receipt and data handling than paper and pencil [6]. Similar results are achieved in our study.

6.3. User acceptance

This study suggests that both interviewers and interviewees generally preferred the PDA-based data collection method. The automated ‘jump’ function was seen to reduce the data collectors’ cognitive burden of paying attention to the notification of ‘jump’, thus eliminating 5% of ‘jump’ errors encountered in the paper-based questionnaire. Also the data collectors were all confident in their use of the PDA and satisfied with the level of participant interaction allowed with the PDA as well as with its speed of use.

A considerable majority of participants (71.4%) preferred the PDA-based data collection method. Some of them held the perception that the data collectors appeared more professional carrying PDAs than pen-and-paper. Therefore, the participants’ acceptance of PDA is significantly better than that of the paper-based questionnaire. This finding agrees with the findings of the previous studies [8,27–32].

6.4. Cost

Although the cost to prepare the data collection instrument for a sample size of 2000 was similar between the two data collection methods, the ability to depreciate the cost of the needed ten PDAs over a 3-year period can significantly reduce this cost in the second and third year.

While there would be a need to provide some additional content to the standard NCD-STEPS Operational Manual with a related marginal increase in production costs for the Manual, the NCD-STEPS training activity for a survey using the eSTEPS...
technology could likely be accomplished with minimal, if any, increased training costs per se.

It is in the data entry, validation and cleaning process that the PDA provides the most significant time and cost–benefit for the prospective NCD-STEPs surveys. This cost saving with the PDA-based protocol is also reported by Gravlee [25] and Shirima et al. [8].

The cost analysis suggests that the PDA-based data collection with eSTEPS technology provides a cost-effective alternative to the paper-based questionnaire, particularly if the cost of the PDA can be depreciated. The hourly labour cost in Fiji is much lower than that in most developed countries. With the increase of labour cost, it is logical to anticipate that the cost saving with PDA-based data collection will further increase.

6.5 Technical challenges for the implementation of a PDA-based questionnaire

Although the instances of PDA failures were rare (three instances, 1.3% occurrence rate) in our study, the consequence of such incidences can be severe in field conditions, as two of the instances resulting in a loss of all installed programs, and the need to reinstall the eSTEPS program and to restart the PDA-based data collection for the involved participant. It has also negatively impacted on the data collector’s acceptance of the new data collection method. These technical deficiencies exposed provided the project group with valuable knowledge about what can go wrong with PDA-based data collection in the field. This is beneficial for the improvement of technologies to increase its reliability and robustness. Furthermore, it is notable that all three PDA failures occurred with ViewSonic Pocket PCs. As such, future applications of the eSTEPS should probably be implemented on PDAs with faster, and likely more reliable, processors.

One potential limitation of the PDA platform is limited battery life. This is not as crucial with the HP IPAQ as a back-up battery can readily be installed to prolong the useful field time for the PDA. This is not possible with the ViewSonic PDA.

Data security was also raised as a concern for the eSTEPS; however, this concern also exists for paper-based data collection. The project team will refine the eSTEPS software to eliminate the program deficiencies and enhance privacy protection for personal data collected.

6.6 Limitations of the study

This study assumes that there are no costs to the survey team of procuring the software. This cost was borne by WHO and the project team. The software was provided free of charge to users. The cost of developing the software further and of releasing maintenance versions could be distributed among the user community once it is released as Open Source software and attracts a sufficient number of developers.

The error rate is based on logical range errors/inconsistencies, skip errors, missing values, data or time field errors and incorrect data type. There was also possibility of choosing the wrong answer from a list of possible answers. As the comparison of the answers for each participant based on data collection method was not conducted, therefore, this type of error was not reported in this study.

As study participants, both data collectors and interviewees, were students from the Fiji School of Medicine, this population group may be different from normal data collectors and participants who will be recruited for NCD STEPwise survey. Thus this might limit the generalisability of the study findings. The study also does not take into account the cost of learning to use the software for setting up and managing surveys. As suggested by Shirima et al., time is needed to pre-program acceptable ranges for numeric variables, to program skip patterns, to check for logical consistency in responses and to thoroughly test that the routines work [8]. However, as mentioned above, the specific arrangement wherein the third author took on this role, as well as the analysis and testing of the final software product, made it impossible to accurately measure the time and cost for these activities. If the questionnaire designer is not such an expert, then time required for designing the PDA-based questionnaire may be significantly increased. Therefore, there may not be any efficiency gain in this activity for PDA-based system.

7 Conclusion

This project provides a potential Open Source solution for mobile public health field data collection. It re-enforces the previous findings that the major advantages of handheld computers are that they “can be programmed to provide determinate responses, date stamped to document times of data entry, prevent omission of data entry, and can save considerable time and labour incurred in data handling. Handheld computers are well accepted, and are more likely than paper methods to be the choice of the user” [6]. As eSTEPS provides all of the above functions, it offers substantial advantages over paper and pen based data collection method.

Besides improving data quality, our study compares in detail the differences between the handheld and paper-based instruments in the whole process of field data collection, including questionnaire design and distribution, data entry, data handling and transfer. It confirms that the PDA-based method can reduce time in data entry and validation.

For the first time, this study details the cost associated with the PDA-based and paper-based data collection methods and provides evidence to show that the PDA-based method underpinned by eSTEPS technology can save cost in large public health surveys. Also, it is the preferred data collection method by both data collectors and participants. Therefore, this study confirms that a PDA-based data collection system that provides the end user with the flexibility of building their own constraints on types, range and logic of data entry is more efficient over paper and can reduce data entry errors. The technical challenges in the field are also identified. This can facilitate the further refinement of the eSTEPS technology and protocol for use in large scope public health surveys.
Summary points
What was already known on the topic?

- EpiData and Epi Info have been widely used by the public health community in developing countries for data entry, storage, analysis and reporting.
- There is no publicly available PDA-based data collection software that is ready for use for mobile public health surveillance data collection.
- There is no convincing evidence about the benefits of using PDA-based public health surveillance data collection software.

What this study added to our knowledge?

- Introduction to an open source, PDA-based data collection software that integrates with EpiData.
- The evaluation results of using this PDA-based data collection software to collect public health surveillance data in a quasi-experimental design.
- Using this PDA-based data collection software significantly reduces cost, data error, time for data entry and cleaning.
- Both data collectors and interviewees were significantly more satisfied with using this PDA-based data collection tool than using pen and paper for public health surveillance data collection.

Acknowledgments

The authors would like to acknowledge the contribution of Stephen Trembath, Nick Harders, Admed Salam and Jeremy Moh for the development of the eSTEPS program and the voluntary data collectors and survey participants, who participated in the eSTEPS product trial. The study was partially funded by the World Health Organization and supported by the WHO Regional Office for the Western Pacific. The eSTEPS software is made freely available from the WHO STEPS team (http://www.who.int/chp/steps/contact/en/index.html).

Authors’ contributions: FY led the design and development of the eSTEPS software and took the lead role in preparing the manuscript and reviewing the selected articles; MdC conceived the software program and its application for the WHO STEPS surveys as well as the WHO STEPS survey method; and oversaw the product trial. EP led the design of the field trial and wrote the initial research report for WHO, based on which this manuscript was developed. All of the authors provided major editing of the manuscript. All of the authors read and approved the final manuscript.

References

[19] J. Ma, M. Otten, R. Kamadjeu, R. Mir, L. Rosencrans, S. McLaughlin, S. Yoon, New frontiers for health information systems using Epi Info in developing countries: structured...


