Available online at www.sciencedirect.com

SCIENCE DIRECT®

Applied Nursing Research 18 (2005) 182–185

Applied Nursing Research

www.elsevier.com/locate/apnr

Research Briefs

A comparison study: paper-based versus web-based data collection and management

Bryan A. Weber, ARNP, PhD*, Hossein Yarandi, PhD, Meredeth A. Rowe, RN, PhD, Justus P. Weber

University of Florida, College of Nursing, Gainesville, FL 32610-1097, USA

Received 25 September 2004; revised 10 October 2004; accepted 18 November 2004

Abstract The purpose of this pilot study was to compare the cost, accuracy, and efficiency of a web-enhanced handheld computer data collection system with those of the traditional paper-based data collection and management system and to increase awareness/knowledge of researchers on these two data collection and management methods. This is an important topic because funding resources are diminishing and high startup costs associated with automated data collection systems may give researchers pause when faced with these financial expenditures. Hence, this information will position grant writers and funders to make intelligent decisions regarding the feasibility and advantage of web-enhanced electronic data collection strategies.

1. Introduction

Researchers in the social sciences traditionally collect data using paper-based instruments that have inherent limitations and risks (Birkett, 1988; Cummings & Masten, 1994). Data can be lost between the field and the research office. They must be manually coded and double checked for accuracy, which adds both time and the potential for human error (Reynolds-Haertle & McBride, 1992). The subsequent data entry into a spreadsheet or statistical software by either research assistants (RAs) or a professional data entry firm also requires double checking for accuracy and potentiates error. Professional firms historically enter data with fewer errors than most research staff, but data may be lost during transit and these professional services are costly (Roberts, Anthony, Madigan, & Chen, 1997; Weber & Roberts, 2000).

Computerized data collection and management have been successfully used in clinical trials (Weber et al., 2004) and are innovative alternatives to paper-based strategies that reduce the inherent risk common to traditional strategies (Birkett, 1988; Crombie & Irving, 1986; Irving & Crombie, 1986; Weber & Roberts, 2000). Data are collected on computers and automatically coded and stored in a spreadsheet or statistical software for analysis. Limitations for this method of data collection have been the size of the computer (making it difficult to use in some field situations), the use of visual analog scales that require touch-screen technology, the available random access memory (or working memory) and storage memory, the computer screen size and resolution, and the volatile nature of magnetic media used to transfer information. However, the relatively new handheld computers and tablet PCs incorporate touch-screen technology and are compact and easy to transport to research sites. In addition, recent advances in data collection software and wireless technology resolve memory limitations by allowing the use of web-based data collection forms in combination with wireless internet connectivity and encryption technology to transfer data from point of entry in the field to powerful computer servers where they can be periodically backed up. With such a system, the number of staff required to manage the data after they have been collected is minimized and the opportunity for human error is "maximally minimized." Moreover, the large geographic areas associated with many research studies as well as the need to maintain the integrity of clinical trial data make web-based data collection even more appealing.

* Corresponding author. Tel.: +1 352 273 6327; fax: +1 352 273 6531.
E-mail address: bweber@ufl.edu (B.A. Weber).

0897-1897/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.apnr.2004.11.003

Researchers are often reluctant to make the initial time and money investments to get started with such a system and/or have misconceptions about difficulty or security issues. The purpose of this study was to compare the cost, accuracy, and efficiency of a web-enhanced handheld computer data collection system with those of the traditional paper-based data collection and management system and to dispel the reluctance and/or misconceptions that researchers have about computerized data collection and management systems. This study was guided by the tenets of survey methodology that aim to minimize nonsampling error in research.

2. Methods and materials

The local institutional review board (IRB) approved the research protocol. The unit of measure (cases) was 20 newspaper articles. Data were extracted by student RAs from 20 newspaper articles that reported cases of older adults diagnosed with Alzheimer's disease who died as a result of becoming lost in the community. Data were gathered from American newspaper reports published from 1998 to 2002 that described incidents in which persons with dementia died as a result of becoming lost in the community. We used the following internet search engines to locate appropriate reports: NexusLexis Academic, Dow Jones Interactive, Google.com, and Teoma.com. Most articles came from daily newspapers of or near the city where each individual was found. Inclusion criteria for the reports were that the person described had a dementing illness, had become lost in the community on his or her own, and had subsequently been found dead. The greatest yield was obtained with the search terms *Alzheimer's*, *missing*, and *dead*. Synonyms of these terms were also used (*dementia*, *lost*, *deceased*, etc.). In addition, the report had to provide information on at least 50% of the variables in the study. These included age, sex, residence, when and how the individual left, when and where the individual was found, who found the individual, how long had the individual been missing, and how far away was the individual from the place he or she was last seen. All the stories that met these criteria were used. We retrieved the original newspaper articles for data collection.

For the first 10 cases, RAs collected and coded data using a traditional paper-based system. Research assistants were also asked to record the amount of time it took to collect, code, and enter the data for accounting purposes and to record the number of errors found through double coding and double entry. For the next 10 articles, RAs used the web-based system to collect data. The participants were not aware at the outset that the data collection strategy would change and were told that the new system was initiated because funding became available for the equipment. This strategy was expected to control for altered behavior on the part of the RAs (Hawthorne effect), who might have otherwise thought that they were competing with the computer. Training to use the handheld computer was minimal because RAs already had experience with using personal digital assistants (e.g., Palm Pilot). As the RAs entered the data, the computerized system automatically timed, coded, and securely transferred the data to a secure server that housed the web-based SQL database.

The web-based forms for this study were created in hypertext markup language using Adobe Accelio Capture Form Flow and were designed for optimal viewing on a handheld computer with 64 MB of random access memory that used a Windows graphical based operating system. The computer was equipped with a cellular wireless modem, so data collection took place in the community without the need for telephone hookup. SQL scripting connected the forms to a database housed on a server and validated the data as they were entered to ensure accuracy and minimize error. Automatic calculations, database lookups, and custom macros were used to simplify form completion and improve accuracy. Forms were customized so that scientific and government standards regarding subject enrollment and IRB protocols would be monitored automatically, thereby minimizing infrastructure costs.

Forms were deployed to a web server hosted by the university health science center. Because both the forms and the database resided on the server, the requirements for storage memory in the handheld computer were minimized. The web server had both public and private areas that were accessed via the internet using web browsers such as Microsoft Internet Explorer. The database and private areas of the web server were secured and only accessible to those with a username and password. The server met government guidelines for research subject data collection and storage. The security of the data transfer was enhanced via encrypted secure socket layer technology. Those individuals to whom access was granted had access to the forms and database from anywhere in the world where there was internet connectivity. Data files were redundantly backed up on local storage devices and via the university's electronic data archiving system.

2.1. Measures

2.1.1. Time

Time in minutes to extract, code and double code, and enter and double enter the data into the statistical software was measured for the traditional paper-based strategy. Time was also automatically calculated by the computerized system.

2.1.2. Error

Research assistants recorded the number of errors discovered when double coding and double entering the data. Errors included incorrect data coding, incorrect data

Table 1
Average time, variable cost, and error: paper-based data collection and management system versus computer-based system

Category	Paper-based (n = 10)			Computer-based (n = 10)		
	Mean time per case (min)	Mean cost per case (dollars)	Errors [frequency (percent)]	Mean time per case (min)	Mean cost per case (dollars)[a]	Errors [frequency (percent)]
Data collection	7.7	1.03	–	4.4	0.59	–
Coding	2.6	0.35	17 (8.5)	0.00	0.00	0.0
Entry	7.0	0.93	13 (6.5)	0.00	0.00	0.0
Mean total per case	17.3	2.31	15%	4.4	0.59	0.0

[a] Includes the cost of cellular connectivity that would be omitted if internet access were available.

entry, failure to add data to the coding sheet, and failure to add data to the spreadsheet for analysis.

2.1.3. Cost

Fixed cost, variable cost, total cost, and total average cost were measured. Fixed cost does not vary with the number of subjects, thus the average fixed cost decreases as the number of subjects increases. With the paper-based system, fixed costs were the costs of preparing the data collection instrument and the coding keys using word processing software. With the computer-based system, fixed costs were the capital investment in dollars for the purchase of equipment and software and the cost of developing, testing, and deploying the electronic forms on a computer server.

Variable cost is the cost that is incurred as a direct result of the operation of the research. It thus varies in proportion to research productivity (i.e., number of subjects). The variable costs in this study were associated with data collection, data entry, and data coding for the paper-based system. The variable costs for the computerized system were associated with RA time spent entering the data.

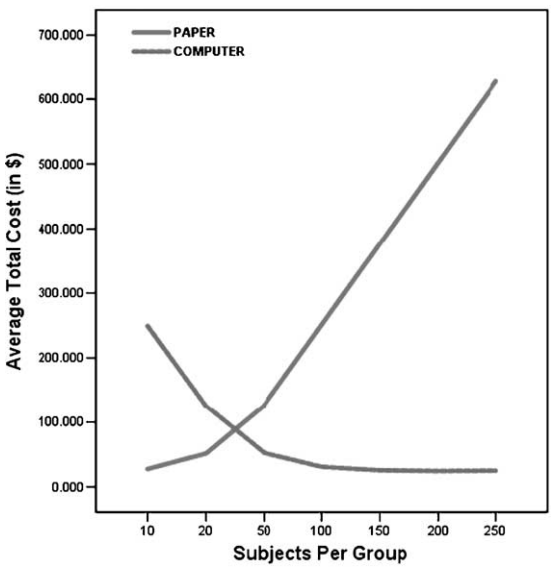

Fig. 1. Average total cost distribution for paper-based and computer-based data collection.

Total cost is measured in dollars as the sum of variable and fixed costs. The average total cost is calculated by dividing the total cost by the number of subjects (or in this study, cases).

3. Results

Research assistants using the paper-based system took 77 min to collect the data, 26 min to code and double code them, and 70 min to enter and double enter them into the statistical software—totaling 173 min. Using the computer-assisted system, the entire process took 44 min.

There were 17 errors (8.5%) in the data collected and managed through the paper-based system. Errors included omission, incorrect coding, and incorrect entry into the statistical software. There were 4 errors that occurred during data coding and 13 during data entry. There were no errors detected in the data managed through the computer-assisted system (see Table 1).

Variable costs totaled US $23.15 for the paper-based method and US $5.90 for the computer-based method (see Table 1 for cost per case). Fixed costs totaled US $24 for the paper-based method and US $2,475 for the computer-based method (an investiture cost that will not recur). Research assistants were paid US $8 per hour for data entry (variable cost) in both systems, for a cost of US $23.05 in the paper-based portion and US $5.86 in the computer-assisted portion of the study. Variable costs for the paper-based system also included US $2 to photocopy 10 copies of the instrument. Fixed costs for the paper-based method included the cost of 3 hr of RA time (or US $24) to prepare the data collection and coding instrument. Development, testing, and deployment of the electronic forms used in the computer-assisted method required 24 hr of RA time, for a cost of US $192.00. Additional fixed costs associated with the computerized web-based data collection strategy included hardware (US $568), software (US $1,305), and cellular connectivity (US $420/year).

The average total cost, as indicated in Fig. 1, steadily increased for the paper-based system as the number of subjects increased. In contrast, there was a decline in average total cost associated with the computer-based system as the number of subjects increased. Average total

costs for the two systems intersected at 32 subjects per group.

4. Discussion

As in other studies that have addressed similar issues of data management (Roberts et al., 1997; Weber & Roberts, 2000), computerized data collection systems allow investigators to maintain maximum control over data, thus minimizing the possibility of lost data during transit when outsourcing strategies for more accurate data management. Moreover, data collection and management processes can be combined in computerized systems to include tasks such as automatic coding, reverse coding, and data entry into secured SQL databases. Databases using SQL protocols are attractive means of storing data because they can be accessed using a variety of front end software packages such as Excel, SPSS, and SAS, among others.

Specialized softwares allow for the creation of customized data collection forms that, again, provide the researcher with greater control over the collection and management of data. And because data are stored on a computer server instead of on paper, which can be lost or stolen, the privacy and ethical integrity of the research are enhanced. Computerized data collection systems also include dynamic features such as the generation of an e-mail message to the PI when research participants are enrolled. Moreover, scripts can set to run in the background to verify things such as if the study is operating under a current and valid IRB protocol at the start of each data collection session before unlocking the data collection tool.

Cost analysis revealed that initial investment was higher for the computer-based system than for the paper-based method, with the average total cost between the two systems being equal at 32 subjects per group. As more subjects are recruited into a study, the discrepancy between the average total costs for the two methods increases. For example, the average total cost of conducting a study with 250 subjects per group will be approximately US $628.00 for a paper-based system and approximately US $25.00 for a computer-based system. This cost divergence further increases as additional technological features are used in the computer-based system. For example, research participants could complete the web-based electronic forms themselves, thereby reducing the number of RAs needed. In addition, e-mail reminders can be sent in lieu of paper reminders and participant incentives can be immediately delivered electronically (e.g., electronic gift certificates to online retailers), thus saving money typically spent on postage and perhaps reducing attrition and enhancing the response rate.

5. Limitations

Fixed costs make a computerized data collection strategy less cost–effective for studies that have a limited number of cases or measurement tools. However, once an e-forms software is purchased, the fixed cost can be recovered over multiple research projects despite the size or complexity of data collection involved.

6. Conclusions

These findings highlight the advantages of using a web-based computerized data collection and management system. First, such a system ensures data integrity and compliance with Health Insurance Portability and Accountability Act of 1996 (HIPAA) regulations. Second, it increases the accuracy and reliability of the data by reducing the opportunities for human error. Third, as competition for funding continues to increase, strategies that maximize research dollars by reducing operating costs are becoming more important. Web-based computerized data collection and management save time in all phases of research studies. Initial startup costs associated with computerized data collection may be high, but with repeated use, large samples, or large data sets, computerized data collection becomes less expensive, more efficient, and more reliable than human data collection and management (Burns & Grove, 1993). Lastly, web-based computer data collection may reach research participants more effectively through availability 24 hr/day and 7 days/week and may reduce attrition and increase compliance with electronically delivered reminders and incentives.

Acknowledgment

Support for this research was provided by a grant from the University of Florida College of Nursing.

References

Birkett, N. J. (1988). Computer-aided personal interviewing. A new technique for data collection in epidemiologic surveys. *American Journal of Epidemiology*, *127*, 684–690.

Burns, N., & Grove, S. K. (1993). *The practice of nursing research* (2nd ed). Philadelphia: WB Saunders.

Crombie, I. K., & Irving, J. M. (1986). An investigation of data entry methods with a personal computer. *Computers in Biomedical Research*, *19*, 543–550.

Cummings, J., & Masten, J. (1994). Customized dual data entry for computerized data analysis. *Quality Assurance*, *3*, 300–303.

Irving, J. M., & Crombie, I. K. (1986). The use of microcomputers for data management in a large epidemiological survey. *Computers in Biomedical Research*, *19*, 487–495.

Reynolds-Haertle, R. A., & McBride, R. (1992). Single vs. double data entry in CAST. *Controlled Clinical Trials*, *13*, 487–494.

Roberts, B. L., Anthony, M. K., Madigan, E. A., & Chen, Y. (1997). Data management: Cleaning and checking. *Nursing Research*, *46*, 350–352.

Weber, B. A., & Roberts, B. L. (2000). Data collection using hand-held computers. *Nursing Research, 49,* 173–175.

Weber, B. A., Roberts, B. L., Resnick, M. I., Deimling, G., Zauszniewski, J.A., Musil, C., et al. (2004). The effect of dyadic intervention on self-efficacy, social support, and depression in men with prostate cancer. *Psychooncology*, *13*, 47–60.