Evaluation of a computerized field data collection system for health surveys

D. Forster,¹ R. H. Behrens,² H. Campbell,³ & P. Byass⁴

A customized field data collection system (FDCS) has been developed for a hand-held computer to collect and check questionnaire data. The data quality, preparation time, and user acceptability of the system were evaluated during a malaria morbidity survey in Bakau, the Gambia. Eight field-workers collected data with either the FDCS or on paper questionnaire forms in alternate weeks over a 6-week period. Significantly fewer item errors occurred with the FDCS, and by the end of the survey period interview times were significantly less with the FDCS than with the paper and pencil questionnaire.

Advanced appropriate technology may have a useful role in providing accurate and rapid information, particularly in overcoming bottlenecks in data processing, and in obviating the need for costly expertise and equipment. In developing countries this could help to improve the quality of data on health care.

Introduction

The first application of computers for the direct collection of data was for studies of telephone interviews, conducted in the 1970s (1). Recent advances in computer technology and increasing miniaturization have resulted in portable computers that now make possible computer-assisted personal interviewing (2). The facilities offered on such systems include the ability to provide range and consistency checks at the time of the interview, guided interviews with predefined routing, error messages, feedback messages, and interpretation of the information collected. Computer-assisted collection provides data in a computer-ready format, thereby eliminating the labour-intensive and error-prone process of data entry, and validation. For this purpose, we designed a custom-written field data collection system (FDCS) to provide validated data entry. Following the development stage, the system underwent field evaluation within an established malaria study at the MRC laboratories in the Gambia (3). The system was compared with a traditional paper questionnaire for improvements in the quality of data, user acceptability, and robustness in daily use. Although computer-based information systems in primary health care have already been established (4), innovative applications should help overcome some of the many difficulties in planning and administering health in developing countries.

Materials and methods

The field data collection system

The system hardware comprised a Psion Organiser II XP hand-held computer, weighing 250 g (Fig. 1), with two removable data chips. The customized

Fig. 1. The Psion Organiser II XP, with a 16 × 2 character screen.

² Senior Lecturer, Centre for Human Nutrition, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, England. Requests for reprints should be sent to Dr Behrens at this address.
³ Clinical Scientist, MRC Laboratories, Gambia. Present address: Acute Respiratory Infections unit, World Health Organization, Geneva, Switzerland.
⁴ Computer Scientist, MRC Laboratories, Gambia.

Reprint No. 5152
software, written in OPL (a BASIC-like language), is held permanently on a 64 Kbyte EPROM chip and collected data are stored on a 32 Kbyte semi-volatile random access memory (RAM) chip. The software provides a versatile and user-friendly method of transferring questionnaires onto the hand-held computer. The construction of the questionnaire involves formulating a question prompt, an item type, and an explanatory note for each question. The item types are defined as numeric, logical, date, or character, and can be validated by range checks on numeric fields, code validity on single-coded fields, branch conditions, or consistency checks against previously recorded data. If required, records can be automatically date-, time- and identity-stamped, and access to the system restricted by password control. Customized routines can also be incorporated into the questionnaire to use child-growth reference standards and produce anthropometric data in a standardized format, or include cohort identity lists for use as a cross-reference during data entry. Several questionnaires can be defined and, when selected for data entry, supported by editing, optional help, and in-built error-message displays. Other utilities available in the data entry mode include data summary and record review. Data are transferred via a standard RS232 interface to any IBM-compatible microcomputer and this step is controlled from within the FDCS.

**Study design**

A 45-item questionnaire from an ongoing field study of malaria morbidity in Bakau, Gambia, was defined on the FDCS. The broad structure of the questionnaire is shown in Table 1. Following 5 days' training in questionnaire administration and the use of FDCS, the field-workers completed weekly questionnaires on their attitudes to the use of computers in surveys. During the first week the eight field-workers were paired and allocated randomly to either the FDCS or the paper questionnaire. They then alternated between the two methods within pairs, over the rest of the 6-week study period, and conducted approximately 500 interviews per week.

The data collected on the paper questionnaires were routinely double-punched, verified, and corrected using a previously published procedure (3). An independent observer estimated the duration of each stage of this process. Once on the data base, the two sets of data (FDCS and paper questionnaires) were thoroughly examined for any logical or range errors made during collection or entry. This routine cross-referenced census data against the identification variables, checked the range of numeric, coded, and date variables, and carried out a plausibility routine on interrelated questions. The duration of the interviews with the paper questionnaires was calculated from the times recorded by field-workers at the start and finish of each interview. With the FDCS the time-stamping facility was selected and times were automatically added to each record. Analysis of variance was used to determine the statistical significance of interviewer, method, and learning effects.

**Results**

The error rates for the two methods are compared in Fig. 2. Use of the FDCS resulted in a significantly lower overall error rate, which when analysed according to the types of variable, showed a 30% improvement for the collection of identification data (name, survey number, and week) and, as would be expected, a 100% improvement for dates and times (Fig. 3). There was also a significant reduction in inter-individual variability in the data accuracy, throughout the study. For example, the error rate of the most accurate and of the least accurate field-worker, respectively, was 0.18 versus 0.58 errors per 100 items (a 70% difference), using the paper questionnaire, and 0.15 versus 0.31 errors per 100 items (a 50% difference), using FDCS. The mean interview duration was significantly less by the third week of the study, and ranged from 4.7 minutes with FDCS to 6.8 minutes for the paper questionnaire, i.e., by the third week the average interview times were 31% shorter for field-workers who used the FDCS ($P = 0.007$, paired Student's $t$-test; Fig. 4). The estimated times taken to enter, verify, and correct a week's data are shown in Table 2. The time taken to transfer data from the FDCS onto a microcomputer averaged 30 minutes per week.

<table>
<thead>
<tr>
<th>Field type</th>
<th>No. of questions</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Character</td>
<td>8</td>
<td>Identification and health card details</td>
</tr>
<tr>
<td>Logical or code</td>
<td>25</td>
<td>Yes/no and number of days with illness</td>
</tr>
<tr>
<td>Open-ended</td>
<td>2</td>
<td>To obtain further information on illness and medication taken</td>
</tr>
<tr>
<td>Numerical</td>
<td>5</td>
<td>Clinical and anthropometric measurements</td>
</tr>
<tr>
<td>Date and time</td>
<td>4</td>
<td>To record start and end times of interviews</td>
</tr>
</tbody>
</table>

Table 1: Broad structure of the questionnaire used in the study

108

Discussion

Recently, the use of computers to collect data in developing countries has been described by Ferry & Cantrelle (5) and Reitmaier et al. (6) using lap-top computers and customized software. Lun et al. have described also the use of hand-held computers to collect health data (7). The study by Ferry & Cantrelle reported improved data accuracy and decreased collection time in longitudinal demographic studies in Senegal, while Reitmaier et al. used computer-aided methods to collect data for an anthropometric survey in Cape Verde. The latter study described how the Sharp pocket computer improved data precision and, with plausibility checks, was able to plot the distribution of the anthropometric data against international standards.

An important focus of the evaluation of our results was whether computers were acceptable to the field-workers, who had been educated to secondary-school standard, and to determine the worker's capacity to deal with the technical aspects of the system, such as installing chips and replacing batteries. Our findings suggest that the field-workers found little difficulty in learning to use and maintain the computers. Their attitudes to computers were monitored throughout the study and they came to accept FDCS for routine data collection. In daily use over 7 weeks, no equipment failure was reported and the computers proved durable, despite exposure to the elements and overnight storage at field-workers' homes. However, more extensive experience in a climatically inhospitable environment is necessary to assess the long-term durability of the system.

High levels of data accuracy with error rates of 0.29 per 100 items have previously been reported for experienced field-workers in a study that used localized real-time microcomputer survey management techniques (3). However, in situations where such facilities and expertise may be unavailable, use of the FDCS could be advantageous. In the United Kingdom the cost of the Psion Organiser II XP hardware, two chips, and two nickel–cadmium rechargeable 9-volt batteries is currently £186 (US$ 350) (excluding value added tax). A cost-benefit analysis of the system must consider the capital costs of the hardware, maintenance, and replacement expenses versus the benefits accrued from improved information for management and planning.
Fig. 3. Mean number of errors detected per week, following data base entry and validation, grouped by data field types; FDCS = field data collection system.

Fig. 4. Mean interview times ± two standard deviations, for field-workers using the field data collection system (FDCS) with times stamped at first and last entry, and for paper and pencil (PP) questionnaires with times recorded at the start and end of the interview (by week 3 the difference between the methods was significant at the $P < 0.001$ level).
purposes. In conventional record-based health information systems, 40–60% of primary health care workers’ time may be devoted to collecting and managing data (5) and any method of reducing this burden should be welcomed. As shown by our results, a major contribution to improved data quality in computer-assisted data collection is disciplined data entry with enforced input, which prevents missing or contradictory entries. Health information systems with local microcomputer access may be able to employ the FDCS to reduce the burden of information collection, while at the same time improving the use and quality of data in administering health services.

Acknowledgements

Save the Children Fund (UK) and the Department of Information Systems, London School of Economics, are thanked for their financial support. We thank Toshiba, UK, Ltd., for computing facilities, Dr B. Greenwood and Dr A. Greenwood for their substantial commitment and support in initiating the field study, and Mr J. Bangali and the computer unit staff for their assistance with data processing. We acknowledge the significant contribution of the field workers in successfully completing the study. D. Forster was supported by a British Council Fellowship.

Résumé

Evaluation d’un système informatisé de collecte des données sur le terrain pour des enquêtes sanitaires

Un système de collecte des données sur le terrain a été mis au point; il consiste en un micro-ordinateur portable et un logiciel spécifiquement conçu pour cette tâche, capable de générer et de gérer un questionnaire. La qualité des données, le temps de préparation et l’acceptabilité du système informatique par l’utilisateur ont été évalués au cours d’une étude de morbidité paludéenne à Bakau (Gambie). Huit agents de santé travaillant sur le terrain ont été formés pendant cinq jours à l’utilisation du système, puis ont recueilli des données chacun à leur tour pendant une semaine en utilisant soit un questionnaire sur papier, soit le système informatique, pendant une période de six semaines. Le degré d’erreur est nettement inférieur (51%) avec le système informatique ($P < 0.001$) qu’avec le questionnaire sur papier; on a ainsi économisé 17,5 heures par semaine en entrée et validation des données. Dès la troisième semaine de l’étude, l’utilisation du système informatique a permis de réduire sensiblement (de 31%) la durée des interviews par rapport à la prise de notes au moyen d’un crayon et d’une feuille de papier.

Une technologie avancée appropriée peut ainsi s’avérer utile pour fournir une information précise et rapide, notamment en venant à bout des goulots d’étranglement dans le traitement des données. Dans la mesure où la préparation et la gestion des données nécessitent souvent des compétences et des équipements coûteux, la mise à disposition de méthodes simples et efficaces pour la collecte, le traitement et la gestion des données peut donc améliorer la qualité de l’information sur les soins de santé dans les pays en développement.

References