Survey Methodology, December 1995
Vol. 21, No. 2, pp. 179–184
Statistics Canada

An Assessment of the Use of Hand-Held Computers During Demographic Surveys in Developing Countries

D. FORSTER and R.W. SNOW[1]

ABSTRACT

Although large scale surveys conducted in developing countries can provide an invaluable snapshot of the health situation in a community, results produced rarely reflect the current reality as they are often released several months or years after data collection. The time lag can be partially attributed to delays in entering, coding and cleaning data after it is collected in the field. Recent advances in computer technology have provided a means of directly recording data onto a hand-held computer. Errors are reduced because in-built checks triggered as the questionnaire is administered reject illogical or inconsistent entries. This paper reports the use of one such computer-assisted interviewing tool in the collection of demographic data in Kenya. Although initial costs of establishing computer-assisted interviewing are high, the benefits are clear: errors that can creep into data collected by experienced field staff can be reduced to negligible levels. In situations where speed is essential, a large number of staff are involved, or a pre-coded questionnaire is used to collect data routinely over a long period, computer-assisted interviewing could prove a means of saving costs in the long term, as well as producing a dramatic improvement in data quality in the immediate term.

KEY WORDS: Hand-held computers; Demographic surveys; Psion.

1. INTRODUCTION

Large scale surveys involving tens of thousands of respondents, such as national censuses, demographic or health surveys, are routinely conducted in developing countries. Their intention is to provide rapid, up-to-date information on population and health issues for evaluation and planning purposes. Their wide scope necessitates numerous personnel comprising trainers, interviewers, supervisors, data entry staff and data managers. Examples of such questionnaire-based surveys include the World Fertility Survey (WFS 1986) and national Demographic and Health Surveys (DHS Kenya 1989). Published dates for the commencement of the WFS surveys in 12 African countries and the dates the first country reports were produced (Table 1) illustrate the time required before data was available for planners to act upon (WFS 1986). On average it took 45.6 months before the final report was released. Survey logistics in developing countries undoubtedly contribute to delays in provision of completed data; so do the mechanics of data processing. The recent Demographic and Health Survey conducted in Kenya required five data entry clerks, two data entry supervisors and a control clerk to process 8,343 household interviews; data collection began in February 1989 and the first draft of the final report was ready for circulation seven months later (DHS Kenya 1989).

Table 1

Summary of Chronology of 12 African WFS Surveys (Source: WFS 1986)

Country	Number of Interviews	Date Survey Started	Date of First Report	Number of Months from Survey Start Till Report Date
Benin	4,018	12/1981	06/1984	30
Cameroon	8,219	01/1978	04/1983	63
Ghana	6,125	02/1979	06/1983	52
Ivory Coast	6,270	08/1980	12/1984	52
Kenya	8,100	08/1977	06/1980	34
Lesotho	3,603	08/1977	12/1981	52
Mauritania	3,500	01/1981	06/1984	41
Morocco	5,800	04/1980	05/1984	49
Nigeria	9,727	10/1981	09/1984	35
Senegal	3,985	05/1978	07/1981	38
Sudan (North)	3,115	12/1978	04/1982	40
Tunisia	4,123	05/1978	06/1983	61

[1] D. Forster, Department of Tropical Medicine, University of Oxford, John Radcliffe Hospital, Headington OX3 9DU, England; R.W. Snow, CRC - Research Unit, Kenyan Medical Research Institute, P.O. Box 230, Kilifi, Kenya.

Surveys of this size involve multiple levels of checking and coding of data collected in the field providing another source of delay. As speed underpins rapid health evaluation (Anker 1991; Vlassoff and Tanner 1992), reducing the time at this check and code stage is a major advantage to the survey process. Advances in computer hardware have led to the development of microcomputers suitable for use in field situations. Together with improved software designed for questionnaire specification and administration, computer-assisted interviewing is now a viable option. National statistics offices in industrialised countries have evaluated the use of this technique, and some now use them on a regular basis (Nicholls and Groves 1986; Lyberg 1985; Denteneer *et al.* 1987; Bench *et al.* 1994). The advantages of these systems are that it reduces recording errors by simplifying skip modules and refusing inappropriate, illogical or inconsistent entries. Furthermore, large numbers of interviews can be stored and simply downloaded to a central computer at the end of every interview session, circumventing the need for data entry clerks.

There is surprising reluctance to adopt this technology in developing countries despite its apparent advantages. There are several possible reasons for this. Firstly, the initial costs may seem daunting and the application deemed inappropriate in countries with scarce resources. Secondly, there have been few attempts to validate their use under field conditions providing little quantifiable evidence of their limitations or advantages over traditional data collection techniques (Reitmaier 1985; Ferry and Cantrelle 1988; Forster *et al.* 1991). This paper presents the results of a comparative study of two methods of field data collection and processing conducted during a demographic survey on the Kenyan Coast.

2. THE ADULT MORTALITY SURVEY

The study was carried out as part of ongoing demographic and epidemiological studies of 60,000 people living on the Kenyan coast. The study population and survey methods employed to monitor demographic events has been described elsewhere (Snow *et al.* 1994). In brief, following an initial census of the population all vital events are monitored by means of 6-weekly house-to-house visits and bi-annual re-censuses of the entire population. During a re-enumeration of the population in November 1993, a survey was undertaken to estimate adult mortality using indirect demographic methods (Timaeus 1991). All women aged between 25 and 44 years were interviewed using the structured questionnaire as shown in Figure 1. The format used precoded closed questions, with logical skips and a consistency check.

Twenty-four field staff, all secondary school leavers, were involved in the survey. All were familiar with survey and census procedures, having had previous formal training in field survey techniques and between 1 and 5 years of field experience. Two days was spent on additional training on the administration of the adult mortality questionnaire. During the survey field staff were divided into two teams, each supervised by a senior fieldworker. Questionnaires completed at the end of each day were checked by field supervisors then passed to the computer staff for data entry. This was done using a screen design reflecting the structure of the paper questionnaire in FoxPro (version 2.0). The same data was independently entered by two data entry clerks and the two completed files compared to identify entry errors, which were subsequently corrected. The completed file was then subjected to logical, range and consistency checks; these included for example, the identification of missing data, incorrect coding (*i.e.*, not using "Y" or "N"), dates inconsistent with the ages of the women and the date of the survey (questions 5 and 6 in Figure 1) and checks that the sums of questions 7, 9, 11 and 13 are consistent with question 15 as shown in Figure 1.

3. COMPUTER DATA COLLECTION TEST

3.1 Computer Hardware and Software

An earlier version of questionnaire-based software was developed for the Psion Organiser II (Forster *et al.* 1991). This model had a limited screen size, 16 characters by 2 lines, but had a fully operational keyboard. The Psion Series 3, used during the present study, offers new possibilities: the screen is much larger, with 40 characters by 8 lines, and integrated graphical capabilities. The machine remains small (165mm by 85mm by 22mm), and weighs 265g including 2 AA sized batteries. The storage devices can store up to 1 megabyte. The keyboard is a 58 key, QWERTY layout. Communications between the Psion Series 3 and a PC entails a simple copy operation between the two storage media.

The software was developed using Psion's in-built programming language, OPL. The paper questionnaire is represented in a structured format in a text file, according to a prescribed format. The questionnaire definition includes a mixture of questions and commands such as skips and range checks. The internal range checks included those developed for the inconsistency checks for the data entered using FoxPro described above. Data entered on the Psion is stored in a separate file, one line for each interview.

To specify a question correctly it must include a question number, the question text and the answer type, which can be a list option, a character input or a number. The definition should also indicate what position in the line the corresponding data entry should be stored and how long the entry is. Numeric answers can also include a prespecified number of decimal points. A range of acceptable inputs

Figure 1. The Adult Mortality Questionnaire

Questionnaire on the survival of relatives
(For all women aged 25-44 years)

Names
Date ID ⌊__⌋⌊__⌋-⌊__⌋⌊__⌋⌊__⌋-⌊__⌋⌊__⌋

I would like to ask you some questions about your natural parents and about your brothers and sisters who have the same mother as you.

1. Is your mother alive? (1 = yes, 2 = no) ⌊__⌋
2. Is your father alive? (1 = yes, 2 = no) ⌊__⌋
   INTERVIEWER: If both parents alive (Q1 and Q2 = 1), go to Q7.
3. Have you ever given birth? (1 = yes, 2 = no) ⌊__⌋
   INTERVIEWER: If she has never given birth (Q3 = 2), go to Q6.
4. Was (*MENTION ALL PARENTS NOT ALIVE NOW*) alive at the time that you gave birth to your first child?

	Yes	No	D/K
Woman's mother	1	2	9
Woman's father	1	2	9

5. In what year was your first child born? ⌊__⌋⌊__⌋
6. In what year (*MENTION ALL PARENTS NOT ALIVE NOW*) die?
   Woman's mother ⌊__⌋⌊__⌋
   Woman's father ⌊__⌋⌊__⌋
7. How many living sisters, born to your mother, do you have? (ALIVE NOW) ⌊__⌋⌊__⌋
   INTERVIEWER: If no living sisters (Q7 = 0), go to Q9.
8. How many of these living sisters are less than 15 years old? ⌊__⌋⌊__⌋
9. How many of your sisters, born to your mother, have died? ⌊__⌋⌊__⌋
   INTERVIEWER: If no dead sisters (Q9 = 0), go to Q11.
10. How many of these dead sisters died before age 15 years? ⌊__⌋⌊__⌋
11. How many living brothers, born to your mother, do you have? (ALIVE NOW) ⌊__⌋⌊__⌋
    INTERVIEWER: If no living brothers (Q11 = 0), go to Q13.
12. How many of these living brothers are less than 15 years old? ⌊__⌋⌊__⌋
13. How many of your brothers born to your mother, have died? ⌊__⌋⌊__⌋
    INTERVIEWER: If no dead brothers (Q13 = 0), go to Q15.
14. How many of these dead brothers died before age 15 years? ⌊__⌋⌊__⌋
    INTERVIEWER: Sum Q7, 9, 11 and 13:
    Q7 = ⌊__⌋⌊__⌋
    Q9 = ⌊__⌋⌊__⌋
    Q11 = ⌊__⌋⌊__⌋
    Q13 = ⌊__⌋⌊__⌋
15. I want to make sure that I have this right. Apart from you, your mother had . . . . . ⌊__⌋⌊__⌋ children altogether? Is that correct?
    INTERVIEWER: In the case of any inconsistency, probe and correct Q7 to Q14 if necessary.

    INTERVIEWER: Please thank the woman for her co-operation.

    Fieldworker code ⌊__⌋⌊__⌋

is an optional specification for numeric or character answers and will include a minimum, a maximum or both. List options can be used to specify codes and their values.

Command actions can be embedded in question texts, so that they are evaluated at the time of questionnaire administration. For example, the final cross-check question in Figure 1 requires an addition. The syntax allows this instruction to be included within the main body of the question text. Other commands can contain instructions on skipping a question, conducting a cross-check between answers or for moving to a different question.

Thus the software uses a flexible way of defining a questionnaire, which is generally applicable. It incorporates manipulation of entered information, integrating arithmetic functions into questions or command lines. The next step forward would be to design an interface for questionnaire specification which removes the burden of constructing a syntactically correct questionnaire definition. The software is available from the authors.

3.2 Test Design

An additional day's training on the use of the Psion was provided for the two team leaders. This involved an explanation of the hardware and software, as well as practice sessions in the field. Both supervisors had had no previous computing experience. Both conducted interviews using either the Psion or paper questionnaires on alternate days, and these formed the basis of the comparison between the methods.

Errors made by all the 22 fieldworkers using the paper questionnaire were counted and tabulated to estimate the background error rate using this method of data collection. Times taken to check the forms once they had been brought back from the field, and for the data to be entered, verified and corrected following range and consistency checks were recorded throughout the survey. Similar timing assessments were made for the Psion data collection procedures.

4. TEST RESULTS

4.1 Time

The average length of interviews conducted on paper was 5.1 minutes, and for those on computer 5.0 minutes, demonstrating no difference between the two methods (Table 2; note only 215/234 interviews were timed). The length of interviews varied considerably from 1 to 18 minutes and increased time related not simply to the number of skips made on each interview, but whether the respondent gave clear, non-contradictory answers.

Team supervisors required between 2-3 hours per day to check each teams questionnaires. The average time taken to enter 500 records (approximate number of interviews completed per week) was 3 hours 40 minutes per data entry clerk; double entry required 7 hours 20 minutes (Table 3). Verification required an additional 2 hours 23 minutes for the same number of questionnaires. The completed files were edited twice to reflect corrected errors and then verified; this took on average two hours 30 minutes for 500 records.

Table 2

Comparison Between Paper Questionnaire and Psion Series 3 Data Collection Methods

	Length of Interview in Minutes				
	Minimum Overall	Maximum Overall	Average Overall	Average Leader A	Average Leader B
Paper	1	16	5.1 (215)*	5.2 (128)	4.9 (87)
Computer	1	18	5.0 (363)	5.5 (190)	4.5 (173)

* Number of timed interviews stated in brackets.

Table 3

Times for Data Processing 500 Questionnaires

Activity	Average time
Data checking	4 hours 8 minutes
First data entry	3 hours 40 minutes
Second data entry	3 hours 40 minutes
Verification	2 hours 23 minutes
Editing	2 hours 33 minutes
Total time	18 hours 24 minutes

4.2 Errors

The errors made were divided into two periods, to assess the effect of familiarity over time. Excluding the two team leaders, the remaining 22 fieldworkers made 1,704 errors on 1,427 questionnaires in the first period, and 1,049 errors on 1,158 questionnaires in the second period. Thus the average error rate per questionnaire in the first two weeks was 1.19 and in the third and fourth weeks was 0.90. In addition, over the entire period 37 questionnaires (1.2% of all interviews) had to be sent back to the field to be redone, as the errors found were not reconcilable in the office. These questionnaires had between 1 and 6 errors to be corrected, with a total of 61 errors. The highest number of errors were made on question 5 (17 errors) and question 6b (15 errors). Error rates per question are shown in Table 4. Fourteen out of the 22 fieldworkers redid at least one questionnaire. One fieldworker was required to redo 8 questionnaires.

Table 4

Type of Errors Made by 22 Fieldworkers Using Paper Questionnaires (for question specification see Figure 1)

	Period 1 (first fortnight)	Period 2 (second fortnight)
Identification	163	48
Question 1	6	1
Question 2	8	2
Question 3	125	92
Question 4a	201	138
Question 4b	151	93
Question 5	105	61
Question 6a	94	57
Question 6b	65	41
Question 7	14	0
Question 8	109	63
Question 9	51	10
Question 10	178	134
Question 11	13	1
Question 12	108	71
Question 13	53	3
Question 14	204	149
Question 15	19	76
Fieldworker code	37	9
Total errors	**1,704**	**1,049**
Total questionnaires	**1,427**	**1,158**

Errors were detected either manually by final checking by one of the investigators (Forster) or through the range and consistency checks performed in FoxPro on the entered data. Field team leader A made 8 errors on 144 questionnaires (0.06 errors per questionnaire), and team leader B made 18 errors on 90 questionnaires (0.20 errors per questionnaire). Most of these errors occurred in question 10 (4 errors) and question 15 (5 errors). The only errors found from the computer data were errors of respondent identification. There were 7 of these, 2 by leader A and 5 by leader B, giving errors of 0.01 and 0.03 per questionnaire respectively. Such errors could have been circumvented by pre-loading the Psion with a call list of respondents to interview.

4.3 Cost

The differential costs of a survey of this size using Psion-based and paper-based methods are given in Table 5. The Psion prices quoted are the recommended retail prices. Intense competition between retailers means that purchase prices could be up to 20% lower than those quoted here. Prices of hardware products are also decreasing. Current prices indicate that the one off cost of a Psion-based system can be recouped after 12-15 similar paper-based surveys of approximately 7,000 respondents.

Table 5

A Comparative Study of Computer-based and Paper-based Survey Methods (UK £ Sterling)

	Equipment required	Cost
Computed-based survey	20 Psion Series 3	2,539.00
	20 1 MB storage devices	2,039.00
	1 serial communications link	59.45
	80 rechargeable batteries	146.20
	1 battery recharger	15.95
	Total cost	**4,799.59**
Paper-based survey	14 reams of paper for 7,000 interviews	42.00
	Duplicating costs for 7,000 questionnaires (double-sided)	70.00
	20 pens, erasers and correcting fluid	27.40
	20 clipboards	100.00
	2 data entry clerks (two weeks)	70.00
	2 supervisors* (one month plus overtime)	85.00
	Total cost	**394.40**

* Necessary for the manual checking of forms as they come in from the field each day.

5. DISCUSSION

The lowest error rates using a traditional paper questionnaire by senior field workers with five years of data collection experience was on average 0.11 errors per questionnaire with 17 fields. This was reduced to negligible levels using the questionnaire software developed for a Psion Series 3 hand-held computer. This technique eliminated most of the errors made by fieldworkers in the routing of the questionnaire (Table 4) by using pre-defined skip modules, thus reducing the error rate by at least 90%. With the additional implementation of a call list in the software, the rate of respondent identification errors would be even lower.

The field supervisors were keen to use the computer, mastering the unfamiliar QWERTY keyboard, and learnt operating procedures quickly enough to take to the field without supervision after two days. Although no formal investigations were under taken to gauge and quantify interviewees' reactions to the Psion there were surprisingly few comments about the computer and no interview refusals.

Data processing involved two data entry clerks using two IBM machines full time for 92 hours to complete the data entry process for the entire survey. A data manager was on hand to offer assistance where necessary and design

the data entry format. The setting of the present study was such that both data entry clerks were familiar with data entry procedures and the available hardware and software. In situations where this is not the case, closer supervision and involvement by a data manager would be necessary, thus incurring an additional cost. A Psion data collection system would require much less of a data manager's time to down-load each day's data, thereby reducing this component of staff costs. Never-the-less, the initial cost of the Psion Series 3 may be prohibitively expensive when compared to the costs of paper and duplication of questionnaires if it was not envisaged that they form part of future data collection activities.

QUESTOR (Ferry and Cantrelle 1988) offers a suitable software environment for computer-assisted interviewing. However, the hardware required is a portable PC, several times the costs of a hand-held Psion. Our experience demonstrates that it will be worth pursuing the development of an appropriate package using this compact PC compatible technology as a more practical alternative in the field, being easier to handle, more robust and with reduced power consumption.

There is a trade off between error rates, time and cost of a survey. The use of computer-assisted interviewing software can reduce both the error rates and the length of time for data preparation considerably. Such a collection system should reduce the unacceptable delays in first presentation of data experienced during surveys such as the World Fertility Survey (Table 1). The context of the present comparative study differs from many large scale demographic surveys where recruited fieldstaff are unfamiliar with questionnaire procedures. We feel that the results presented here therefore represent a minimum improvement that could be expected in data quality. The initial cost of setting up such a survey mechanism may be daunting, but will be proportionally less for repeated surveys, or in institutions conducting a variety of different surveys over time.

ACKNOWLEDGEMENTS

The authors wish to thank all the field staff involved during the survey especially the two supervisors, Rodgers Chisengwa and Lewis Mitsanze, the data entry clerks Robert Mutai and Monica Omondi. We also acknowledge the support of Dr. Ian Timaeus, who designed the adult mortality questionnaire and Dr. Chris Nevill, who conducted the re-enumeration. We also wish to thank Dr. Kevin Marsh for his support for the computer studies at Kilifi, the Wellcome Trust, UK for financial support; the donation of a Psion Series 3 by Psion PLC, UK; and the Director of KEMRI for permission to publish this work. Dr. Bob Snow is supported as part of The Wellcome Trust Senior Fellowship programme in Basic Biomedical Science.

REFERENCES

ANKER, M. (1991). Epidemiological and statistical methods for rapid health assessment. *World Health Statistics Quarterly*, 44, 94-97.

BENCH, J., CLARK, C., DUFOUR, J., and KAUSHAL, R. (1994). Computer-assisted interviewing for the labour force survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

DHS, KENYA (1989). Report on Kenyan Demographic and Health Survey. Institute for Resource Development/Macro Systems Inc., Columbia, Maryland, USA.

DENTENEER, D., BETHLEHEM, J.G., HUNDEPOOL, A.J., and KELLER, W.J. (1987). The BLAISE System for Computer-Assisted Processing, Automation in Survey Processing. Netherlands Bureau of Statistics, CBS Select No. 4, 67-76.

FERRY, B., and CANTRELLE, P. (1988). The use of microcomputers for collection of demographic data in the field. In *African Population Conference*, Dakar, Senegal. Liege, Belgium: International Union for the Scientific Study of Population (IUSSP), 15-30.

FORSTER, D., BEHRENS, R.H., CAMPBELL, H., and BYASS, P. (1991). Evaluation of a computerized field data collection system for health surveys. *Bulletin of the World Health Organisation*, 69, 107-111.

FORSTER, D., and SNOW, R.W. (1992). Using microcomputers for rapid data collection in developing countries. *Health Policy and Planning*, 7, 667-71.

LYBERG, L. (1985). Plans for computer-assisted data collection at Statistics Sweden. *Bulletin of the International Statistical Institute*, 45th session, Invited Papers, Volume LI, Book 3, Section 18.2.

NICHOLLS, W.L., and GROVES, R.M. (1986). The status of computer-assisted telephone interviewing: Part I - Introduction and impact on cost and timeliness of survey data. *Journal of Official Statistics*, 2, 93-115.

REITMAIER, P., DUPRET, A., and CUTTING, W.A.M. (1987). Better health data with a portable microcomputer at the periphery: An anthropometric survey in Cape Verde. *Bulletin of the World Health Organisation*, 65, 651-657.

SNOW, R.W., MUNG'ALA, V.O., FORSTER, D., and MARSH, K. (1994). The role of the district hospital in child survival on the Kenyan Coast. *African Journal of Health Sciences*, 1, 71-75.

TIMAEUS, I.M. (1991). Measurement of adult mortality in less developed countries: a comparative review. *Population Index*, 57, 552-568.

VLASSOFF, C., and TANNER, M. (1992). The relevance of rapid assessment to health research and interventions. *Health Policy and Planning*, 7, 1-9.

WORLD FERTILITY SURVEY (1986). Final report. International Statistical Institute, Netherlands.