

ORIGINAL ARTICLE

**EXPERIENCE WITH DIGITAL ENTRY
OF NATIONAL IODINE SURVEY DATA IN SENEGAL**

Siekmans K^{1*}, Ngnié-Teta I, Ndiaye B³ and P Berti⁴



Kendra Siekmans

*Corresponding author: ksiekmans@healthbridge.ca

¹ Health and Nutrition Specialist, HealthBridge, Ottawa, Canada

² [Formerly] Senior Program Officer, Micronutrient Initiative (MI), Ottawa, Canada;
[Currently] Nutrition Specialist, UNICEF, Haiti

³ Director, Sénégal & Sahel, Micronutrient Initiative (MI), Dakar, Sénégal

⁴ Nutrition Advisor/Deputy Director, HealthBridge, Ottawa, Canada

ABSTRACT

There is growing evidence and enthusiasm for the use of technology to enhance accuracy and speed and minimize the costs of quantitative data surveys. Handheld computers have the potential to facilitate high quality data collection and timely analysis for large complex surveys in developing country contexts. The objective of this paper is to document the experience of using handheld computers for direct data capture in a national survey to monitor progress toward universal salt iodization in Senegal. Twenty-five personal digital assistants (PDAs) were programmed and used by five teams across 13 regions for entering data from 3768 households over a period of three weeks. The health staff selected to collect the data learned how to manipulate the PDAs within a short amount of time. The PDAs contributed to improved quality of data collection due to automatic skipping of non-applicable questions and selection of individuals eligible for goitre assessment. The PDAs also were programmed to randomly select one woman and one school-age child within each household for biochemical sample collection. Data on geographic locations were collected for 82% of households surveyed using the PDA's in-built Global Positioning System (GPS) functionality, which showed the geographical dispersion of these households and which will be used for analysis of results for key indicators by location. Problems with household selection processes, identification coding and standardized interview methods were observed. While costs for printing questionnaires and manual data entry were saved, significant costs were incurred for PDA technical support by an international consultant. The use of PDAs significantly reduced data processing time. Data were stored and downloaded to a central database, with the full dataset and preliminary results available to stakeholders within one week following the survey completion. The PDAs are an excellent tool for enhancing capacity to collect complex household survey data and make it available for analysis in a timely manner. Quality training and adequate pre-testing of questionnaires are still essential. Increased use of direct data capture methods in health program design, monitoring and evaluation is recommended, along with efforts to build local technical capacity.

Key words: salt iodization, survey, handheld computer, Senegal

INTRODUCTION

Data collection through surveys is widely used in research and program evaluation. Typically, this involves data collection using paper questionnaires followed by computer data entry to capture the information in a format useful for data analysis. However, the use of handheld computers enables these two steps in data collection to be done simultaneously, a process called “direct data capture”, and this technology radically changes how survey data are collected. Handheld computers currently come in a variety of formats, including tablets and smartphones, and less commonly now, PDAs. Experience with handheld computers for direct data capture is increasing in the global health research field and the body of evidence for the benefits, costs and limitations is growing [1-4]. Personal digital assistants (PDAs) are one type of handheld computer that have been used to facilitate high quality data collection and timely analysis for large complex surveys, including in low and middle-income countries.

The benefits of using direct data capture include achieving data quality similar to paper systems or higher, without the time and cost associated with data entry of paper-based questionnaires [3]. Use of logical checks can reduce data entry errors and prevent missing values. Programmed skip patterns provide a means of improving internal consistency by ensuring that all appropriate questions are asked by the enumerator. Data records can be stored both in the PDA’s internal flash memory as well as on a secure digital (SD) card and cannot be edited after leaving each household, thus ensuring the security and integrity of the data [5].

Key concerns with the use of this technology include the complexity of programming the questionnaire, the lack of a hard copy of completed questionnaires for data verification, and the potential for data loss due to damaged, lost or stolen equipment. Handheld computers are most suitable for yes/no questions, multiple-choice questions, or questions with short numeric answers. In contrast, given the small on-screen keyboard, handheld devices are not well-suited for lengthy text answers. Data collected through iterative questioning, such as the 24-hour dietary recall, is also difficult to transfer to this format. Another frequently cited concern is cost, even though data on the cost comparison between traditional paper-based and direct data capture surveys are limited. While direct data capture surveys have additional costs related to the handheld computer hardware, software and technical support, they also benefit from significant cost savings due to the lack of printing questionnaires and data entry-related expenses. In a very large-scale survey in Burkina-Faso, the cost of PDAs was estimated to be less than the cost of GPS units, computers for data entry and one million photocopies [1].

The objective of this paper is to document the experience of using handheld computers for direct data capture in a national survey to monitor progress toward universal salt iodization in Senegal, thus contributing to the debate of the relative merits of handheld computers used in surveys. More specifically, this paper outlines the lessons learned in terms of programming, training, field data capture with PDAs, data processing and management in a context where enumerators were accustomed to

paper-based surveys and where handheld computers were being introduced for the first time.

METHODS

Background on study context

Iodine deficiency disorders (IDD) are a problem of public health importance in Senegal. About 60% of households do not use adequately iodized salt and over 75% of children are iodine deficient [6]. The Government of Senegal, in collaboration with other development partners, including UNICEF, WFP and the Micronutrient Initiative (MI), has made efforts through its National Nutrition Program to address micronutrient deficiencies, including iodine deficiency. A national program to promote universal salt iodization (USI) was started in 1995 and scaled up in 2006 to build the capacity of small salt producers to supply adequately iodized salt. Working with both industrial and small-scale producers, the program's target is 90% of households in Senegal consuming iodized salt.

To assess progress, in collaboration with the Institute for Research and Training in Population, Development and Reproductive Health (IPDSR, University Cheikh Anta Diop), MI provided technical and assistance and financial support to the Government of Senegal to design a national IDD survey in 2009. The main objective of the survey was to evaluate the prevalence of IDD in Senegal after two years of an intensive program to support small-scale salt producers. In addition, the survey was designed to examine factors associated with IDD across the country, provide baseline information for future studies of the program's impact, evaluate the quality of the salt iodization program in Senegal, and develop recommendations for enhancing the effectiveness of the national salt iodization program.

Survey design

The survey drew a nationally representative sample. Two-stage cluster sampling was used to select 3,500 households. In the first stage, 240 clusters (defined as "districts de recensement" (DR)) were randomly selected using probability proportional to size based on the sampling frame from the most recent population census [7]. For each DR selected, a team of cartographers prepared a detailed list of all households. From this list, 15 households were randomly selected for inclusion in the survey as well as five replacement households. Among the 15 selected households in each DR, three were randomly assigned as "sub-sampled" households from which salt and urine samples were to be collected.

The questionnaire, initially designed in a paper format and later adapted for PDA, consisted of several modules that were administered to each household. The introductory module consisted of household informed consent, location identification and a time stamp for interview start. The second module was a listing of all household members, their characteristics and the results of goiter assessment for eligible members (all school-age children and women of reproductive age present in the household at the time of the interview). The third module described household characteristics and included the salt rapid test, salt and bouillon cube consumption and

knowledge of IDD. A fourth module assessed the reproductive history (childbearing history, miscarriages and stillbirths) for one eligible woman in each household and the last module recorded the GPS reading and the time stamp of interview completion. In sub-sampled households, samples of salt and urine were collected for laboratory analysis.

Ethical Concerns

The study protocol was approved by the Division of Research and Ethics Committee of the Ministry of Health in Senegal. Informed consent was obtained from the household head and each individual involved in the study, either directly for adults or by an adult guardian for the children.

The PDA component

Data were captured directly on PDAs at the time of the household interview. Technical support on the use of PDAs for data collection, management and analysis was provided by HealthBridge (Ottawa, Canada). The PDAs were ASUS 696 PDAs with built in GPS (ASUS Computer International, Fremont, CA) using extended life batteries and running Windows Mobile 6 (Microsoft Corporation, Redmond, WA). Waterproof OtterBox™ cases (Otter Products LLC., Fort Collins, CO) were used to protect the PDAs from dust, rain, and accidental drops. The PDA batteries were recharged from AC outlets or car chargers.

Visual CE 11.0 software (Syware Inc., Cambridge, MA) was used to program the PDAs with the questionnaire and to handle data input and uploading into a Microsoft Access database. There was embedded logic in the PDA programming to enhance the quality and consistency of the data collected. For example, if the respondent refused to participate in the survey, the questionnaire was closed and the interview ended. During the registration of all household members, skip patterns were programmed to facilitate the efficient collection of data that were appropriate for the age of each member. For example, questions regarding occupation were not asked for children under 13 years of age. In all modules, enumerators were required to enter a response to a question before proceeding to the subsequent question.

The PDA prompted the enumerator to enter data related to specific target groups, based on the information entered during the household member listing at the beginning of the interview. For all women 15-49 and children 6-12 years of age, the PDA was programmed to prompt the enumerator to perform a clinical assessment for goitre and enter the result during the household member listing process.

The random selection of one child and one woman from the household to provide a urine sample was performed by the PDA. During the registration of all household members, each woman 15-49 y and child 6-12 y was assigned a random number between zero and one by the PDA (stored on the PDA but invisible to the enumerator). The woman and child with the highest number were then selected as the members of the household to provide the urine sample and their identification codes displayed on the screen for recording on the urine sample containers. This enabled an

unbiased sub-sample of women and children to be selected at the household level. The randomly sampled woman also was asked the reproductive history module questions.

Survey Team Members and Training

District health team members, the majority of whom were public health nurses involved in nutrition-related activities, were appointed by the Ministry of Health to be enumerators within regions where their respective districts are located. Five senior team members who had previous experience with conducting household surveys in their respective zones were chosen as supervisors.

Survey training was done in two stages. An initial three-day training session was held for the survey coordination team and the five zone supervisors. This first training session was used to introduce the group to the PDAs and finalize the survey questionnaire. The training covered all key aspects of the survey, including operation of the PDA, interviewing techniques, clinical assessment of goitre, collection of salt and urine samples, practice sessions, and field testing. Team supervisors received specific training in PDA maintenance, battery charging, troubleshooting, and data backup. Supervisors were instructed how to save the data on each PDA to SD memory cards every day to avoid any potential data loss. As a result, a complete record of all data collected to date was saved as a separate file at the end of each day. In case of hardware or software failure, the backup files enabled one to go back in time and “recreate” the status of each PDA on any given day.

A second stage of training was conducted for the enumerators. As with the supervisors, one day was spent on introducing the teams to the PDAs and providing training in their operation, while also reviewing the questionnaire. Another day was spent on training in clinical assessment of goitre, collection of salt and urine samples, practice sessions and field-testing.

Data management

On the last day of survey data collection, data were downloaded from each PDA to a Microsoft Access database. The cleaning process involved removing incomplete interviews and any data entered outside of the dates of the survey. The data were then imported for further analysis into SPSS.

Laboratory results for urinary and salt iodine concentration were provided in Excel spreadsheets two months following the completion of the survey. Matching these results to the households and individual members was done manually using the identification code produced by the PDA for each sample taken.

RESULTS

Twenty-five PDAs were programmed and used by five teams across 13 regions for entering data from 3,768 households over a period of three weeks (October 15 to November 5, 2009). Health staff learned how to use the PDAs within the two days of training provided. There were no problems with theft or loss of equipment. The duration of the household interview (from start of interview to end of GPS recording,

as recorded by the PDA) was available in 3766 cases (99.9%); mean duration of interviews was 19 minutes; median duration was 16.1 minutes. Average length of interview did not vary by location (for example, urban/rural areas). However, as expected, it did vary substantially by size of household, ranging from 12.5 minutes in households with 2-4 members to over 25 minutes in households with more than 10 members (see Figure 1).

Data were available for 3,768 households, 1,506 from the goitre endemic regions and 2,262 from the other regions. A total of 31,375 household members were enumerated, with an average of 8.3 members per household nationally (urban 7.9, rural 8.7) similar to the Senegal DHS 2005 (mean 8.7, urban 7.8, rural 9.5) [8]. A total of 15 households (0.4%) refused to participate in the survey. Anecdotal reports from the enumerators suggested that although some respondents asked questions about the PDAs, such as whether they were equipped with a camera (they were not), non-participation in the survey was not directly related to the use of PDAs for collecting the data. The PDA programming ensured that responses were entered for all questions. As a result, there were virtually no instances of missing data. However, one error was made in the programming of a skip pattern which resulted in the questionnaire prompting the enumerator to assess goitre only in children 6-12 years who reported attending school, even though the survey protocol required that all children 6-12 years be assessed for goitre.

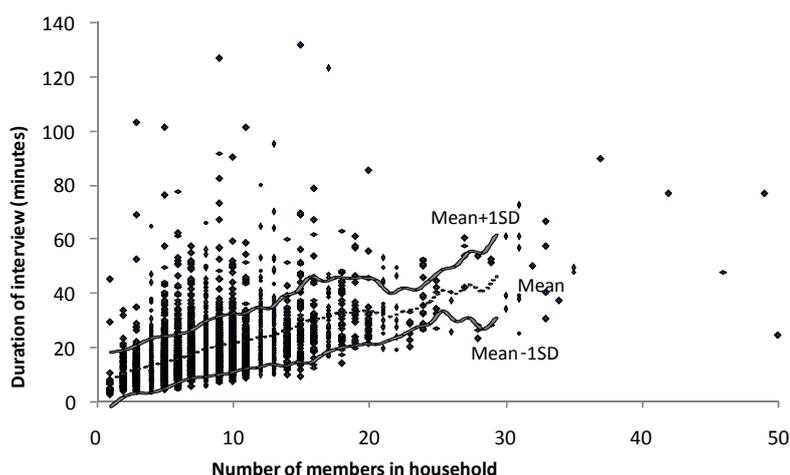


Figure 1: Scatterplot of duration of interview against household size

Usefulness of GPS data

Location readings using GPS functionality of the PDAs were taken for 82% of households surveyed. The missing GPS data arose in cases where interviewers were unable to obtain an adequate satellite position fix when capturing the GPS data, probably because they were indoors, under a large tree or in some way did not have a clear line to the satellites. The data points were plotted using Google Earth (Google Inc., Mountainview, CA), providing a visual representation of the geographical dispersion of the households surveyed (Figure 2A). A visual overview of three clusters surveyed in an urban setting (city of Tambacounda), each with 15 data points, is shown in Figure 2B. Analyzing the distribution of results for key indicators by location is planned. Finally, GPS data was also used for survey data quality control, since the geographical coordinates can be used to confirm the physical location of each enumerator at the time of collecting the data so that in combination with the date/time stamp for each interview, the daily journey of each enumerator can be reviewed, as shown in Figure 2C.

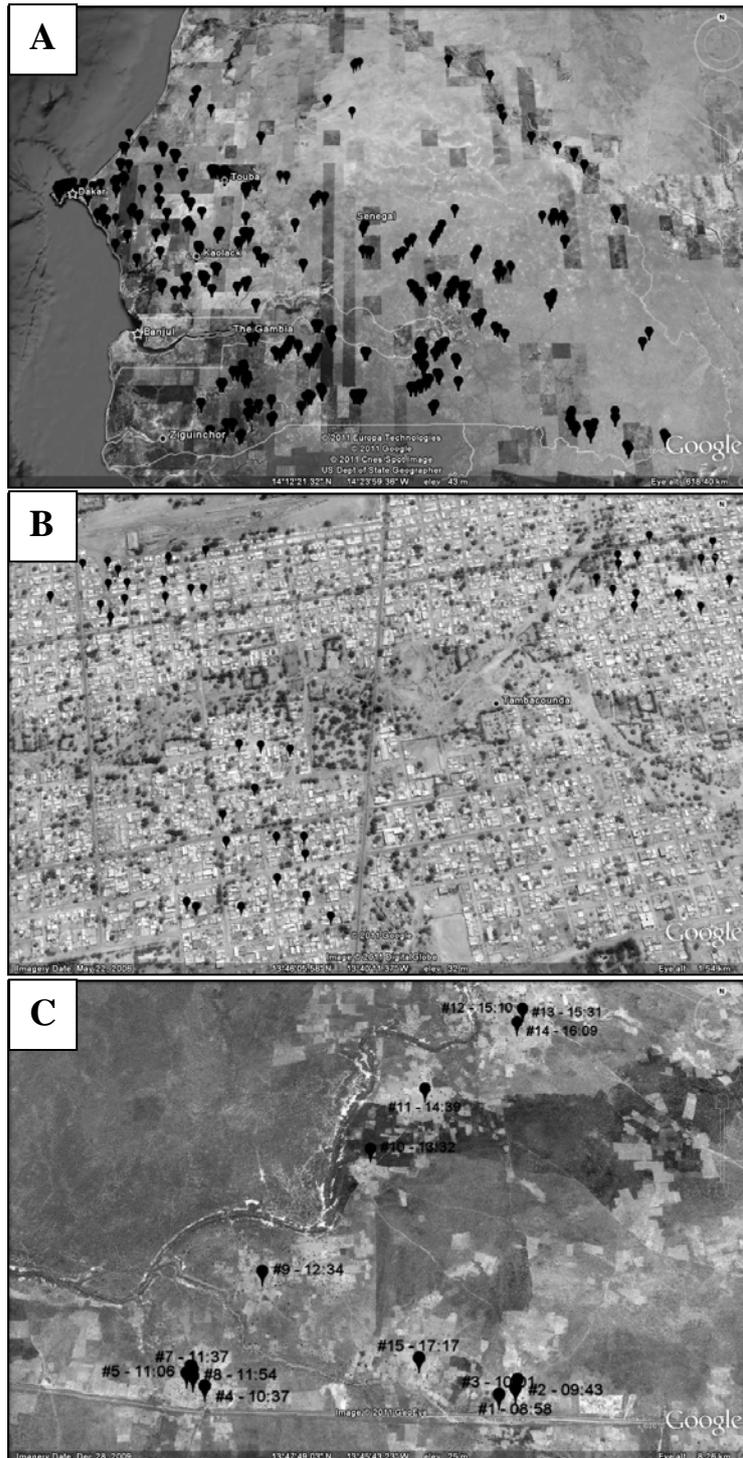


Figure 2: Map of surveyed households using GPS coordinates in Google Earth: A) national distribution; B) three clusters in an urban area; C) households surveyed by one enumerator on 26-Oct-09 with start time of interview for each household

Data processing and data matching

The use of PDAs reduced data processing time. Data were stored and downloaded to a central database, with the full dataset and preliminary results available to stakeholders within one week following the survey completion.

Data from the laboratory analysis were available after two months. The laboratory sample identification codes were based on location data (cluster name and household number) entered by the enumerator. When this data was entered incorrectly during the course of the survey, it resulted in some instances of duplicate identification codes. Therefore, a few laboratory samples were labelled with non-unique identification codes. Some errors also occurred in the manual transcription of these identification codes, either by enumerators when copying the code from the PDA to the sample container or by laboratory staff when copying the code from the sample container to the Excel spreadsheet of results. Despite these errors, 98% (1419/1453) of urinary iodine sample results analyzed by the lab were matched to the appropriate household member (711 children, 708 women) and 96% (705/737) of salt sample results analyzed by the lab were matched to the appropriate household. Approximately 95% of households with UI data for women/children also had salt iodine concentration results, enabling a strong basis for correlating iodine nutrition with household salt iodine levels.

DISCUSSION

Conducting a national survey in Senegal using handheld computers was successful and provided high quality data on key indicators for IDD control. The use of handheld computers contributed to high quality data collection due to automatic skipping of non-applicable questions, automatic checks of entered data, and selection of individuals eligible for goitre assessment. The handheld computers also enabled consistent and non-biased intra-household random sub-sampling of one woman and one school-age child for biochemical sample collection.

One strength of the survey was the random selection of a person in the household by the PDA, to provide biological samples. However, the occasional absence of the selected person at the time of the interview was a limitation. In such cases, the enumerator made an effort to return later to the household to collect the sample and assess the person for goitre, recording the results on a piece of paper which was then given to the supervisor. When this was not possible or the selected member refused to provide the sample, the enumerator selected another eligible and available member as a replacement. This refusal to provide a sample was observed only for a few individuals (13 women and 4 children) and would not have significantly biased the results. To avoid this problem in the future, it would be necessary to collect information on whether the individual is present at the time of the interview during the household member listing process.

One of the attractive features of an electronic questionnaire is that changes can be made at the last minute, following pre-testing and just prior to implementation. In this survey, last-minute changes were made to enhance the clarity of questions, simplify

data entry and modify an unnecessary skip pattern. However, in our experience, the flexibility of making these last-minute changes to the questionnaire was associated with a cost, as one error in the programming for recording the results of goitre assessment of school-age children was only observed after the majority of data had been collected. The level of association between school attendance and goitre is not known for this context, but there is likely a degree of bias in the goitre results as a consequence of this error. In light of this experience, it is recommended that no structural changes be made to the device-based questionnaire following the pre-testing process, unless there is adequate time for thorough re-testing of the revised version. Closer monitoring of data collected during the course of the survey may also have caught this error earlier and enabled correction of the programming. Alternatively, for optimal quality control, a subsample could have been validated using a paper questionnaire to detect such errors.

Adapting a paper questionnaire to a handheld computer-based questionnaire

There are conceptual and practical considerations when adapting a paper-based survey to a handheld computer-based survey. With a multiple-page paper-based survey, identification codes are manually assigned to each household and individual and in some cases, these must be recorded in multiple places in the questionnaire in order to enable matching of child characteristics (such as age or sex) with other variables, such as child's goitre status. If the pages become physically separated (for example, the staple comes out) the identification of the respondent is lost unless the time-consuming step of accurately copying identifying information to every page has been carried out. With a handheld computer-based survey, identifying information can be collected once and then automatically linked to all the data collected from a given respondent. In this survey, where clinical examinations for goitre were carried out on multiple individuals within each household, the goitre assessment was done on each individual in sequence within the middle of each interview, rather than measuring all goiter at a single time at the end of the survey. This is a subtle difference from paper-based surveys, but it has implications for interview logistics that are sometimes not fully appreciated during the design of the survey protocol. Other methods for dealing with this limitation have been documented [5].

Another challenge with the use of handheld computers for enumerators who are used to paper-based surveys is that there is no written record of the data that has been collected. For example, it was not possible for the enumerator to verify the household number that was used in the previous interview. In the Senegal experience, this resulted in occasional duplicate household numbers entered by enumerators. Another example is that the household member listing was done without the ability to look back and verify information about members who had already been entered into the handheld device. To improve the accuracy of data collection, enumerators were rigorously trained in a standardized order of members to be entered (household head, first spouse, first spouse's oldest child, and so on). In addition, after each member's information was added, the PDA was programmed to provide the number of members already entered so that enumerators could verify their progress and determine whether additional members were required to be entered. Given the close correspondence in

average household size between the Senegal Demographic and Health Survey 2005 and this survey, it is likely that the enumeration of household members was done accurately.

Survey Problems

Despite the advantages of the handheld computer-based questionnaire, problems with household selection processes, identification coding and standardized interview methods were still observed. There were two types of household replacement procedures: one to replace a household that was selected to participate in the survey but for which data were not collected (for example, refused to participate or no one available to respond) and another one to replace a sub-sampled household because it did not meet the eligibility criteria. Despite specific training and supervisor guidance for enumerators in these procedures, errors in household selection were observed. Associated with this were numerous cases of improperly entered household identification codes, resulting in duplicate ID numbers for different households. Lessons from other national surveys using PDAs equipped with GPS units to select a probability-based sample would help to address the concern with household selection [2]. Automatic computer-assigned identification codes were useful in differentiating between households with duplicate enumerator-assigned codes. More extensive training of enumerators on standardized interviewing techniques is also recommended, due to concerns with enumerator administration of the questionnaires at the household level, as expressed by coordination team members during supervision visits.

Cost of handheld computer-based surveys

Some researchers have reported a lower financial cost for handheld computer-based surveys, especially when the equipment is going to be used for multiple surveys [5]. In this experience, while costs for printing questionnaires and manual data entry were saved, significant costs were incurred for technical support by an international consultant. Building local capacity for handheld computer-based surveys would help to reduce this cost and provide opportunities for continued use of these devices during follow-up evaluations for the same program as well as application of this technology in other sectors. The development and free availability of software and online training for programming and using handheld devices in surveys is expected to help achieve this goal.

A challenge for surveyors in using handheld computers is keeping abreast of mobile technology. The ASUS PDAs used in this research are still fully functional and their size, computing power, battery life and durability suit the purposes of many field surveys. However, the units are approximately five years old and are now perceived as “old technology” and out of date. Indeed, it would be difficult to find this PDA or any other brand, for sale new, and new adoptees of handheld computers for field surveys will have to find a suitable, current technology. While the broad lessons about handheld computers that are discussed in this paper are valid and would apply to other technologies too, many new lessons will have to be learned about the specifics of conducting a survey using new technologies (what software to use, how to back up data during the course of data collection, how to integrate GPS data into the survey,

how to ensure sufficient battery life, what to do in the rain, and so on). With the inevitable rapid pace of change in technology, users of these devices for field survey purposes will need to adapt to each new generation of equipment.

CONCLUSION

Handheld computers are an excellent tool for enhancing capacity to collect complex household survey data and make it available for analysis in a timely manner. Quality training and adequate pre-testing of questionnaires are still essential. Increased use of handheld computers and other mobile technologies in health program design, monitoring and evaluation is recommended, along with efforts to build local technical capacity.

ACKNOWLEDGEMENTS

This study was financially supported by the Micronutrient Initiative, with the financial support of the Government of Canada through the Canadian International Development Agency (CIDA). The authors would like to thank all those who helped to make the survey possible, especially Dr. Mouhamadou Guélaye Sall (Director, IPDSR) and Dr. Boubacar Camara (Research Manager, IPDSR) for coordinating the survey, the Heads of the Regional Directorates of Health for their logistical support and the members of selected households who participated in the survey.

REFERENCES

1. **Byass P, Hounton S, Ouédraogo M, Somé H, Diallo I, Fottrell E, Emmelin A and N Meda** Direct data capture using hand-held computers in rural Burkina Faso: experiences, benefits and lessons learnt. *Trop Med Int Health*. 2008; **13 Suppl 1**: 25-30.
2. **Vanden Eng JL, Wolkon A, Frolov AS, Terlouw DJ, Eliades MJ, Morgah K, Takpa V, Dare A, Sodahlon YK, Doumanou Y, Hawley WA and AW Hightower** Use of handheld computers with global positioning systems for probability sampling and data entry in household surveys. *Am J Trop Med Hyg*. 2007; **77**: 393-399.
3. **Blaya JA, Fraser HS and B Holt** E-health technologies show promise in developing countries. *Health Aff (Millwood)*. 2010; **29**: 244-51.
4. **Gallihier JM, Stewart TV, Pathak PK, Werner JJ, Dickinson LM and JM Hickner** Data collection outcomes comparing paper forms with PDA forms in an office-based patient survey. *Ann Fam Med*. 2008; **6**: 154-160.
5. **Shirima K, Mukasa O, Armstrong Schellenberg J, Manzi F, John D, Mushi A, Mrisho M, Tanner M, Mshinda H and D Schellenberg** The use of personal digital assistants for data entry at the point of collection in a large household survey in southern Tanzania. *Emerg Themes Epidemiol*. 2007; **4**: 5.
6. **Micronutrient Initiative** Investing in the future: a united call to action on vitamin and mineral deficiencies. Micronutrient Initiative: Ottawa, Canada, 2009.
7. **Agence Nationale de la Statistique et de la Démographie (ANSD) Sénégal**: résultats du troisième recensement général de la population et de l'habitat (RGPH) – 2002. ANSD: Dakar, Senegal, 2006.
8. **Ndiaye S and M Ayad** Enquête Démographique et de Santé au Sénégal 2005. Centre de Recherche pour le Développement Humain [Sénégal] et ORC Macro: Calverton, Maryland, USA, 2006.