Health surveys are important tools for studying the incidence and prevalence of diseases, health expenditures and use of services, effects of policy decisions, and association of important risk factors to various disease outcomes.

Conducting health surveys in developed countries can be difficult, but conducting health surveys in developing countries poses enormous challenges—for example, when a survey is carried out in a country trying to revive itself after a long civil war. The population may live mostly in rural and inaccessible areas, and a great deal of population migration might be taking place. It is difficult to compile a list of addresses to create a sampling frame (list of households or people in the population of interest) unless there has been a recent census, and so one has to choose an alternative way to get the information. Getting a representative sample is important, especially when one wants to obtain estimates of disease prevalence/incidence in a population, and a good sampling frame can help ensure this. There are several common sampling strategies (e.g., stratified, systematic, multi-stage cluster) to draw representative samples, but most of them are based on having access to a sampling frame.

In developed countries, several modes of data collection often are used, such as mailed paper surveys, telephone calls, or in-person interviews. Sometimes, a combination of two or more of these modes is employed to increase the response rate and efficiency while minimizing costs. The selection of the mode depends on the size and topic of the survey, available funds, and feasibility of obtaining a high response rate. In developing countries, one often has to conduct in-person interviews because regular telephone services might not be reliable and people might not be able to respond to mail surveys easily.

Manual collection of information for a sampling frame and conducting surveys that later have to be entered into a computer can be expensive and time consuming. However, an exciting technology application that cuts down the time and makes both these aspects (sample-generation and data collection) efficient and effective is global positioning systems (GPS)-enabled mobile devices. GPS-enabled mobile devices such as personal digital assistants (PDAs) have great potential for widespread use for surveys in developing countries. They are a great tool for building a sampling frame where none exists, selecting a sample, and collecting accurate data. They can be used in a variety of environments, but are especially useful in developing countries where a majority of the population lives in rural areas and/or there is no recent census information. There is no need for double data entry or the additional cost that comes with it since all the information is entered directly into the PDA.

Disease and Child Mortality in Sierra Leone

Sierra Leone (meaning “lion mountains”) is a country located on the west coast of Africa that experienced a civil war for more than 10 years—between 1991 and 2002—disrupting many services, including the health care infrastructure. It has an estimated population of about 5.9 million, with approximately 60% of the population living in rural areas. Sierra Leone now...
has some of the poorest people and worst health indicators in the world. Sixty-three percent of the population falls below the international poverty line of $1.25 per day. Fifteen percent is under five years of age. In 2009, the infant mortality rate was estimated at 123 per 1,000 live births, the second-worst in the world (the worst being Afghanistan), and the “under-five” (i.e., under five years old) mortality rate was estimated at 192 per 1,000 live births, ranking fifth for under-five mortality in the world. Sierra Leone is ranked the sixth-highest in the world for maternal mortality ratio (MMR), with estimated MMR of 970 per 100,000 live births.

Under-five mortality due to malaria, diarrhea, and pneumonia and fever in Sierra Leone has been high since the country came out of its civil war. Because health care in the rural areas was almost nonexistent, the Ministry of Health and Sanitation allowed nongovernmental organizations to deliver community case management (CCM) with community health workers to increase coverage of treatments for malaria, diarrhea, and pneumonia in an attempt to improve the health conditions of children. In 2006, this form of CCM was piloted in one district by the International Rescue Committee and then extended to two additional poor and marginalized areas with funding from the Canadian International Development Agency and technical assistance from United Nations Children’s Fund (UNICEF).

UNICEF wanted to evaluate the success of the CCM program by comparing the rates of treatment for the three diseases (malaria, pneumonia, and diarrhea) before and two years after launching the program. The main objectives were to compare the change from baseline to 2012 in coverage of treatment for the three diseases between areas that were selected to receive CCM and areas that were not. No baseline estimates were available. UNICEF decided to estimate these rates by conducting a survey before the launch of the program. A baseline survey was conducted in the survey areas in June/July 2010, with a follow-up survey to determine the success of the CCM program planned for June/July 2012.

Theresa Diaz, the principal investigator from UNICEF, collaborated with Adam Wolkon from the Centers for Disease Control and Prevention (CDC) and statisticians from Statistics Sierra Leone to conduct the survey. She contacted Statistics Without Borders (SWB), which is an outreach group of the American Statistical Association comprised entirely of volunteers for statistical assistance. SWB members provide pro bono statistical consulting and assistance to organizations and government agencies in support of these organizations’ not-for-profit efforts to assist with international health issues. Two from SWB provided sample size estimates and support on the design and with the preliminary analysis of the data.

GPS-Enabled Personal Digital Assistants

To conduct surveys in developing countries where a sampling frame is not available, either new sampling frames can be constructed on paper, or quicker but biased methods not requiring frame construction can be performed (e.g., the “EPI spin the bottle” method, in which one spins a bottle, walks in the direction the bottle points, and chooses households systematically (maybe every third or fourth house)). Constructing sampling frames on paper are time consuming, and the EPI method does not generally permit all households to have prespecified, or even known, probabilities of selection. Being able to estimate the probability of selection accurately is important to produce population-based unbiased estimates.

The GPS-enabled PDA is a powerful tool designed for rapid use and is superior to the traditional listing methods. Typical use of this
technology, an example being the freely available software CDC GPS Sample (based on a presentation by Adam Wolkon at UNICEF, NY), involves five steps: creating a sampling frame, merging data between multiple mappers, performing the random sampling, navigating to the selected households, and conducting the survey.

To create a sampling frame of houses in a neighborhood, one walks around the area with the PDA in hand and records GPS points by mapping the location (latitude, longitude) of the houses. If it is a big area, multiple people can map at the same time. The sampling frame is then created by merging the data from the different PDAs via an infrared or Bluetooth connection. As the sampling frame is constructed, the program generates a random ID as each household is being mapped. Later, when a random sample is desired, any number of procedures can be used to select a random sample based on the generated random household IDs.

The GPS navigates the interviewer back to the sampled households with maps, location information, distance, and compass direction pointer. The household selection directly inserts the selected household into the survey module. The survey is directly programmed into the PDA and enables the interviewer to conduct interviews more efficiently, since they do not have to worry about the skip patterns or consistency or range checks. The PDA also can incorporate visual aids (e.g., pictures of medicines), which are helpful in improving accuracy in responses to some questions (though due to the small screen, this is usually done on laminated cards).

The data are stored on the persistent memory and backed up to the solid-state memory card, so the loss of data is minimized, even when the PDAs break or lose power. The privacy of the respondents can be protected and confidentiality of the data maintained by restricting access to the data and nonessential programs using a password-protected program. Added advantages of the PDA are that interviewers are happy not to have to carry huge loads of paper surveys around and that the data can be instantly downloaded to an Access database for immediate analysis.

With all these advantages in mind, the GPS PDA technology is not a “one-size-fits-all” solution. They need to be programmed for different environments and uses. Use of PDAs cannot overcome a poorly designed survey instrument. All the rigor that goes into planning a paper survey also should go into planning a survey to be programmed into a PDA.

The Sample

Sierra Leone is divided into four provinces, each consisting of 14 districts that have their own elected local government. Selection of districts to receive CCM was based on a composite score that included wealth status, immunization and stunting rates, and availability of health services. Of the seven districts, two (from different parts of the country) were selected to receive CCM and two nearby districts were selected as comparison districts. A representative sample of households (HHs) was drawn from each of these districts.

Each district is comprised of many enumeration areas (EAs), which can best be understood as census tracts. Based on the 2004 census, a complete set of EAs was recorded, and, for our purposes, they could be assumed to capture the entire Sierra Leone population. To obtain data on households within the four districts, we sampled 50 EAs with probability proportional to size of the EA population and then sampled 30 households completely at random within each EA. Enumeration areas consisting of fewer than 55 households in the 2004 census were combined to enable sampling from at least 50 HHs.

The Survey Instrument

The survey consisted of three questionnaires: one for the household, one for women of child-bearing age (women’s survey), and a third for women with children under the age of five (child survey). The questionnaire asked about the medications
given to children for the treatment of the three conditions—malaria, diarrhea, and pneumonia and fever. We did not incorporate the visual aids into the PDA, but instead had interviewers carry the pictures with them.

Investigators from UNICEF and CDC developed the questionnaires and trained native Sierra Leoneans to be interviewers. Statistics Sierra Leone provided assistance with some of the training and all the field work. The 50 interviewers had previous experience with national household surveys and participated in a one-week training that included going through each question to ensure they understood what was being asked, practicing interviewing with each other, learning techniques for approaching the household, and administering consent. They also participated in two field practices with the PDA, one in the surrounding urban community, after which problems with the PDA were fixed, and one in a farther off community. The survey was written in English (it is the only language with script in Sierra Leone) and verbally translated by interviewers into the language preferred by the respondent (generally Krio, Timne, Mende, or Limba) using standardized key words obtained from a qualitative study and the interviewers themselves.

We pilot-tested the survey in 10 households to gauge the accuracy and validity of the survey questionnaires so any emergent problems with the questions and layout could be solved before programming into the PDAs for the main survey. The pilot test enabled us to modify the questionnaires to address errors in skip patterns, rewording and reordering of some questions, and the inclusion of alternatives in the response category that fit into the Sierra Leonean context. For example, in the household questionnaire, the alternative, “an animal-drawn cart” was replaced with “ormolankay,” a cart drawn by human beings, since animal-drawn carts are not used in Sierra Leone.

**Data Collection via GPS-Enabled PDA**

To construct a sampling frame within EAs, the surveyors went to the sampled EAs and walked around with the GPS-enabled PDAs to map every HH in the region. The EAs generally cover a fairly small geographic area and a small number of households. We had some EAs that had less than the required sample size of 50 HHs, so we combined them with other similar EAs to create a large enough area to sample the required number of HHs. The CDC GPS Sample allowed the collection of information on the name of the township/village, an ID number for the household, and a text field to store any additional information that might help in identifying the particular households later. It also stored the coordinates (longitude, latitude) of the mapped households, which helped in the navigation later. When more than one interviewer mapped the same area, the information collected was transmitted to each interviewer’s PDA to create a complete listing of the households. This created our sampling frame within an EA.

This sampling frame was then used to sample the households for the actual survey. A random number was generated as each household was being mapped, which was then used as the ID number for that household. These were then sorted in an ascending manner and the required numbers of households were sampled from the total number of households. The mapping of the households was done separately in each EA, and 30 households were drawn completely at random within each of the 50 EAs per district we had identified.

The interviewers were given the PDAs that had the sampled households and the surveys already programmed in them. The GPS guided the interviewers back to the sampled households based on the coordinates and by displaying a dot showing the location of the house on a compass on the PDA.

Figure 2. The PDA screen showing the list of households sampled and a map of the layout of the households with the red arrow indicating the sampled household.
The selected household was directly inserted into the household module. The interviewer then administered the household survey, which collected baseline information to assess the eligibility for the other two surveys. If eligible, the interviewer was prompted to continue with the next survey.

We downloaded the data from the PDAs into an Access database for analysis. We managed to conduct a preliminary analysis in a week's time, which would not have been possible if the survey had been conducted using paper questionnaires employing double data entry to get all the data onto the computer for analysis.

In spite of pilot testing the instrument, we did encounter some problems in the field. The program on the PDA created duplicate records when an interview was not completed at one sitting and the interviewer later went back to the same household. We found that some of the records were incomplete because interviewers had stopped the questionnaire at some point and either did not return to complete the interview or did return to complete the interview but started all over again creating a new record. We had not programmed all the consistency checks for the variables. For example, persons could enter zero for number of persons living in the household when they may have actually meant 10. Where possible (e.g., duplicate records), we cleaned data by using the record with the complete information or using some kind of cross-checks of questions to find the answers for the unanswered questions.

The original surveys were conducted in June/July of 2010. A follow-up survey is being planned, and we intend to address the problems encountered in the earlier survey as well as incorporate visual aids in the PDA.

**Final Thoughts**

Several technologies are available for data collection: paper, smart paper, scannable forms, PDAs (smart phones, data-only devices), tablets, netbooks, notebooks, etc. PDAs and/or tablets are the most convenient since they are easy to program (even when collecting complex data), can be used interactively with a respondent, do not require additional data entry, can transmit data on the fly, and are hand-held and not cumbersome to carry around. In addition to being well-suited as a survey tool for developing countries, PDAs with cellular data connectivity or smartphones provide an ideal platform for mobile health applications that allow remotely located patients to enter information or upload pictures that can immediately be relayed to a remote health professional. These applications are expected to improve health care access to the immobile, or rural, populations while also cutting health care costs.

**Further Reading**

CDC GPS Sample: https://sites.google.com/a/wolkon.com/gps-sample.


