Accelerating and improving survey implementation with mobile technology: Lessons from PMA2020 implementation in Lagos, Nigeria

Linnea Zimmerman[1],[a], Funmilola OlaOlorun[b], Scott Radloff[a]
[a] Johns Hopkins Bloomberg School of Public Health, USA
[b] College of Medicine, University of Ibadan, Nigeria

Abstract

Large-scale nationally representative surveys have traditionally been implemented using paper surveys, necessitating secondary steps of data entry and management after data collection. Errors occurring during data collection or entry may not be rapidly identified. The Performance Monitoring and Accountability (PMA2020[2]) project implementation in Lagos, Nigeria demonstrates four advantages to integrating mobile technology into survey implementation. First is the rapidity of data collection; data collection lasted six weeks from mapping/listing to final collection – and, since completed surveys are uploaded to a cloud-based server, identification of errors can occur in near real-time. Second, time-stamping and GPS marking allow for improved quality assurance. Third, the inclusion of GPS coordinates creates new opportunities to analyze relationships of distance with use of health services. Fourth, PMA2014/Lagos utilized a 10% resample of households to validate data collection allowing for rapid identification of questionable data and quality control.

Keywords: Survey implementation; Mobile technology; Family planning

Résumé

Traditionnellement, les enquêtes à grande échelle représentatives au niveau national sont effectuées en version papier, ce qui nécessite des étapes secondaires comme la saisie et la gestion de données après le recueil. Il se peut que les erreurs qui se présentent pendant la collecte ou la saisie de données ne soient pas identifiées de façon rapide. La mise en œuvre du projet « Performance Monitoring and Accountability » (PMA2020[1]) à Lagos au Nigeria démontre quatre avantages de l'intégration de la technologie mobile dans la réalisation des enquêtes. Le

[1] Corresponding author
Johns Hopkins Bloomberg School of Public Health, Baltimore, USA
615 N. Wolfe Street, W4511; Baltimore, MD 21205, USA
Tel: +1-410-955-2232; Email: lzimme12@jhu.edu
[2] www.pma2020.org

premier avantage est le temps de terrain rapide ; le recueil de données s'est effectué pendant six semaines, de la phase de cartographie et recensement des ménages jusqu'à la fin de la collecte. De plus, puisque les enquêtes complétées sont téléchargées au serveur basé dans les nuages, l'identification des erreurs peut se passer pratiquement en temps réel. Deuxièmement, l'horodatage et le géoréférencement en utilisant la fonction GPS permettent une assurance qualité améliorée. Troisièmement, l'inclusion des coordonnées GPS crée des nouvelles opportunités pour analyser l'association entre la distance et l'utilisation des services de santé. Quatrièmement, PMA2014 à Lagos réinterroge 10% des ménages pour la validation du recueil de données, ce qui permette l'identification rapide des données non valables et un contrôle qualité.

Mots clés: Mise en œuvre d'enquête ; Technologie mobile ; Planification familiale

Introduction

Mobile technologies are increasingly being utilized by the health sector in developed and developing countries for a variety of purposes, from educating clients on healthy behaviors to managing supply chains for essential medicines and commodities. Less widespread, however, is the use of mobile technology as a data collection tool. Some researchers have suggested that surveys are often the only way to collect reliable data in low- and middle-income countries where health data are not collected in a continuous systematic fashion due to administrative issues and manpower shortage. (Seebregts et al., 2009, Shirima et al., 2007) Incorporation of mobile technology into survey implementation appears to have multiple potential benefits.

Use of mobile technology to collect survey data is an improvement on the traditional paper based method of data collection in a number of ways. First, interviewers may enter incorrect, inconsistent, or illegible information on paper forms, which may not be discovered until the data-cleaning phase, at which time it is too late to cross check information with the respondent. The scripting of software packages used on mobile devices, such as smartphones, allows for the introduction of constraints that will minimize, if not completely abolish, such errors. Further, many surveys contain skip patterns which can create another source of error if not followed precisely; software packages can build these skip patterns into its sequencing routine, preventing an interviewer from skipping a relevant question. Second, while it may be difficult or almost impossible to catch field workers who falsify data in paper based surveys, data fabrication can be caught almost as soon as it occurs when mobile technology is used to collect data (See example described by Tomlinson et al., 2009). (Tomlinson et al., 2009) Third, double data entry to minimize data entry errors, which has been the standard of paper-based surveys is expensive and time consuming, but is unnecessary when mobile devices are used to instantaneously enter and store data. Fourth, cleaning data using the paper-based platform is often a long

arduous process whereby queried responses are checked alongside the paper forms. With the use of mobile technology, data collection and data entry are merged into a single process, alleviating much of this burden. Finally, the storage of paper forms can take up non-trivial physical space and may require environmental and security controls which may be unavailable or expensive. In contrast, hundreds of forms can be saved as backups on a mobile device and in internal memory storage, and hundreds of thousands can be stored on secure servers in a remote location.

Although innovative, time-saving and efficient, the use of mobile phones in data collection carries with it challenges that must be overcome within the specific sociocultural context. Research suggests that respondents of mobile phone surveys are often afraid their pictures may be taken or voices recorded without their consent. (van Heerden et al., 2013) Privacy issues continue to be an ethical concern in mHealth research. (Shilton, 2012) Even as technology alleviates some challenges with survey research, it can introduce new ones. Smartphones require a level of technical competency and comfort that may not be common in settings where diffusion of smartphones is not widespread. Trainings for interviewers need to cover both survey content and introduction to smartphones, including basic functions, survey software, and basic troubleshooting. More complex errors can arise that may be beyond the scope of the interviewer to address and require advanced skills from supervisors and managerial staff. However, these challenges can be overcome through focused training and supervision.

A testimony to the superiority of mobile devices such as personal digital assistants (PDAs) and mobile phones for data collection is the observation that they are gradually displacing paper based methods.(Seebregts et al., 2009, Lane et al., 2006) A project in South Africa trained community health workers to conduct household surveys using mobile phones, (Tomlinson et al., 2009) demonstrating the feasibility of showing individuals with limited survey experience how to use mobile phones to collect health-related data. The evidence that minimally skilled individuals can be trained to conduct surveys using smartphone technology, leading to improvements in data quality and rapidly available data, is further strengthened by the PMA2020 project, discussed below. An introduction to the project is provided, followed by data gathered from PMA2020 implementation in Nigeria that demonstrate the advancements in survey implementation made possible by utilizing mobile technology.

PMA2020

The Performance Monitoring and Accountability (PMA2020[3]) project, in partnership with country research organizations, utilizes a network of resident enumerators (REs) who reside in or near a probability sample of-selected enumeration areas to conduct nationally representative surveys. REs are female community members, generally with little to no survey experience, who are trained to conduct surveys of households, individuals, and health facilities – and can be deployed to conduct multiple rounds of surveys.

[3] www.pma2020.org

Females are employed as the female-specific questionnaire contains questions on sexual history and contraceptive use that would be culturally inappropriate for males to ask. At the household level, they gather information on household assets and amenities, including water and sanitation facilities, as well as basic roster of household members. All females between the ages of 15-49 are interviewed to gather additional information with a focus on reproductive health, including contraceptive access, use, cost, and choice. Surveys are also conducted at selected Service Delivery Points to gather data on health service availability. In the first two years of country implementation, data are collected at six-month intervals and annually thereafter.

The utilization of a network of resident enumerators, combined with improvements made possible by mobile technology, generate rapid, high-quality data that can be used to improve reproductive health and water and sanitation programs. The integration of mobile technology into survey implementation engenders numerous other improvements over the traditional paper-based model.

This paper will demonstrate four such advances using data collected from PMA2020 implementation in Lagos, Nigeria:

1. Increased speed of data collection, cleaning, and dissemination
2. Improved data quality with checks - made possible with smartphone capabilities
3. New research implications
4. Data verification with 10% rapid resample of selected households

PMA2014/Lagos[4]

Data

In 2014, PMA2020 was implemented in two Nigerian states, Kaduna and Lagos. All data used in this paper come from the first round of completed data collection in Lagos, Nigeria. The samples drawn for both PMA2014/Lagos and Kaduna were representative at the state level. The primary sampling unit within PMA2014/Lagos is a cluster, made up from between two to five enumeration areas (EAs), which are drawn from the sampling frame of the National Population Commission. Thirty-nine clusters were selected in Lagos to generate state-level estimates of contraceptive use. Within each cluster, a household listing was conducted to create a sampling frame of all households and from this, 35 households were randomly selected for interview. Of the 1,365 households selected for interview, 1,014 household interviews were completed, identifying 899 eligible women between ages 15-49. Of these, 807 female interviews were completed.

REs were recruited from within or near each of the identified clusters. Of the 39 REs engaged in data collection in Lagos state in 2014, only three had previous experience of paper based data collection, and none had been involved in electronic data collection. All were conversant with the use of smart

---

[4] A note on nomenclature: When referring to the overall project or standardized protocols, the project is referenced as PMA2020. When country or round specific information is referenced, it is referred to by date and geography (e.g. PMA2020 implementation in Lagos in 2014 is referred to as PMA2014/Lagos).

phones, although not everyone owned one. Six supervisors interfaced between REs and the state coordinator, each closely supervising 6-7 REs. Training was conducted using a cascade model, with the supervisors first attending a training of trainers' workshop, and thereafter facilitating a resident enumerator training. Each training workshop lasted 2 weeks and included didactic lectures, group discussions, role-play, group hands-on practice sessions, and field practice. Training included both survey content and mobile phone utilization. Focused attention was given to REs who had limited experience with smartphones. REs were assigned to supervisors on the first day of training, and thus started early to form a bond with their respective supervisors.

Advances using mobile technology

Rapid data collection

During data collection, survey forms are transmitted directly from the smartphone through the telecommunication provider and uploaded to a cloud server, meaning progress can be tracked from almost anywhere in the world with internet connectivity in near real time. Additionally, Open Data Kit (ODK), the software platform used to collect data on the smartphone, utilizes skip patterns and logical constraints to limit data entry errors and minimize cleaning time. (Hartung et al., 2010)
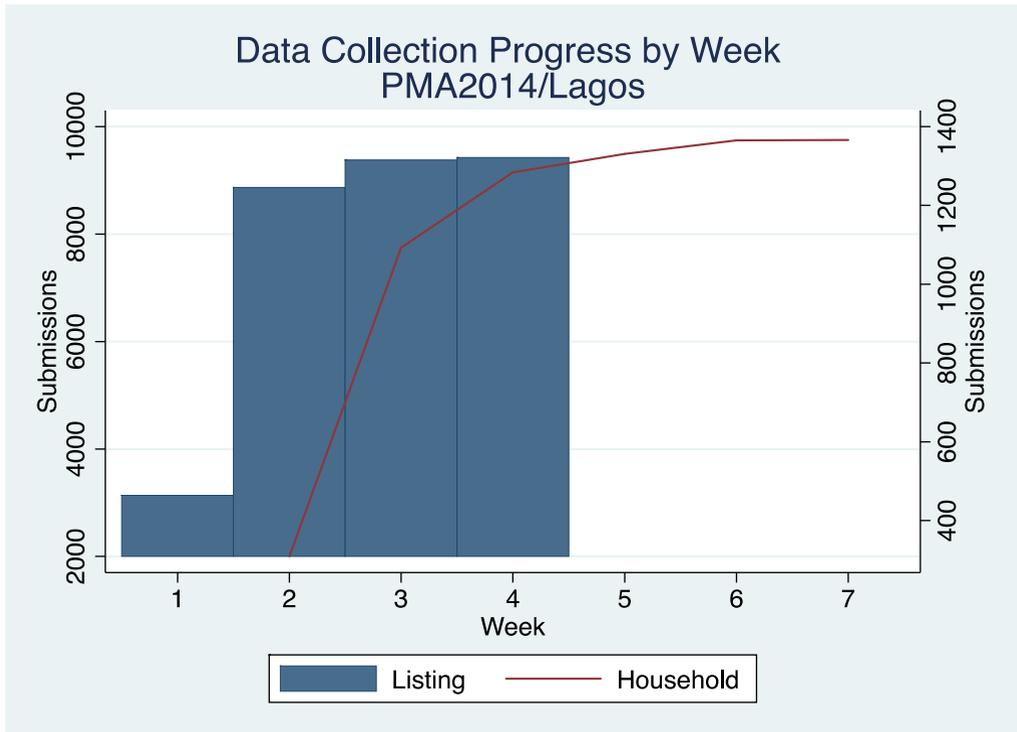
Data collection in all PMA2020 countries has two stages: 1) listing of all households and Service Delivery Points (SDPs) within the selected enumeration area and 2) administration of the household, female, and SDP surveys. Functionalities of the smartphone and the ODK program allow for automatic time-stamping of forms, making it possible to track progress from the beginning of data collection to completion.

Training of data collectors in Lagos took place from September 8 to September 18, 2014. At the listing phase, information on the number of households within residential structures in each enumeration area was collected in order to sample households. The first listing form was submitted to a secure cloud-based server on September 21, 2014 and the final form was submitted by phone on October 7, 2014. The first household questionnaire was started on September 24, 2014 and the final household questionnaire was submitted on November 1, 2014.

Figure 1 below shows the number of listing forms and household forms submitted by week, beginning with the first week of listing (September 14-20) through to the final week of data collection (October 26-November 1).

Figure 1: Cumulative progress of data collection of household listing and household surveys by week in PMA2014/Lagos

Data collection, consisting of listing 9,426 households and completing 1,013 household interviews, 807 female interviews, and 94 SDP interviews, was completed within seven weeks, with the majority of work completed within four weeks.

Additionally, the inclusion of logical constraints and pre-programmed skip patterns reduced data cleaning time, meaning that data were available for preliminary analyses within two weeks of the completion of data collection.

A direct comparison with large scale, paper-based surveys such as the DHS or MICS is not possible as the amount of time allotted to data entry and editing is not publicly available. However, with thousands of forms to enter, compare and resolve, this is not a trivial task and can easily take weeks.

Quality control measures

Quality control measures are critical to any survey in order to guarantee the highest quality data possible. Traditional paper based surveys rely heavily on supervisory personnel to ensure that surveys take place within selected households and are conducted correctly. While smartphones can never completely replace this supervisor structure,
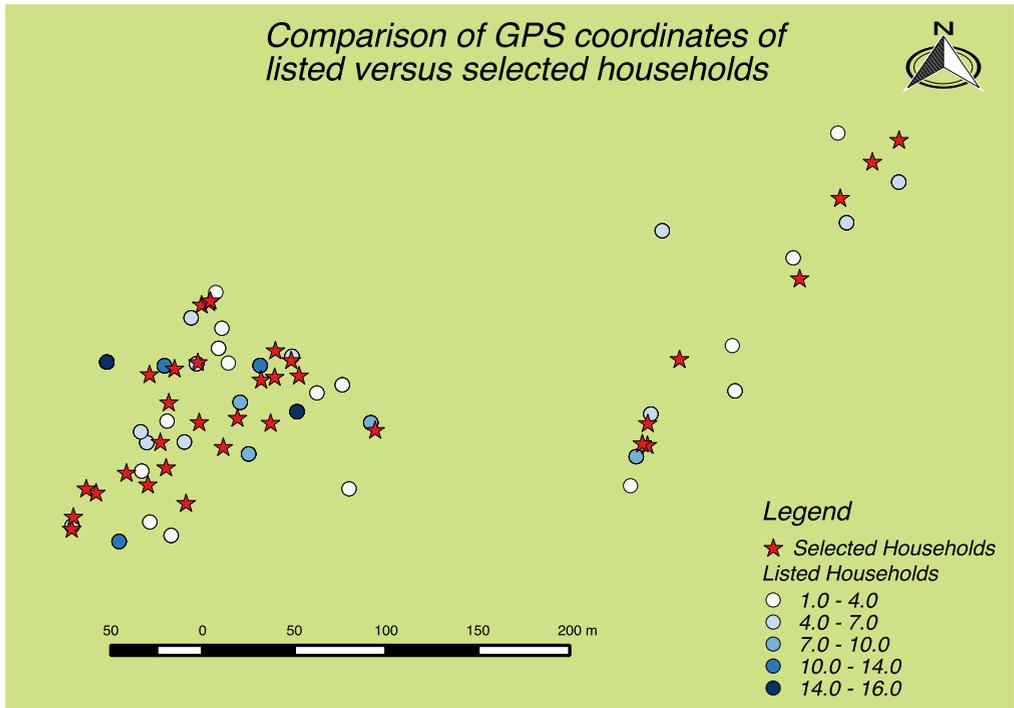
certain functionalities contribute additional means with which to monitor data. Time-stamping, mentioned previously, allows for the monitoring of how long each survey takes to complete and can identify any REs whose response patterns differ markedly from expectations.

As a standard quality control check for all PMA2014/Lagos survey submissions, time from the start to the end of the interview was recorded. Any interview that took less than five minutes but was marked as complete was flagged and communicated to supervisors for follow-up. While supervisors confirmed with REs, response patterns were checked to see if there were viable patterns for these shorter interviews. As data is automatically uploaded to the cloud server, these suspicious interviews can be identified in almost real time. In PMA2014/Lagos, automatic downloading and verification of data happened at 12-hour intervals.

A second innovation for quality assurance demonstrated using PMA2014/Lagos data is the use of GPS. GPS coordinates are taken at both the listing stage when the sampling frame is constructed and during the household and SDP surveys. These coordinates were cross-checked to ensure that the correct households were interviewed and that there was geographic distribution of survey GPS points. Figure 2 shows the geographic dispersal of all residential structures identified in two EAs within Lagos (blue dots) and the GPS coordinate of household surveys that were completed (red stars). Each residential structure may have more than one household and more than one household may be selected from within one residential structure. The geographic disbursement of the household form submissions in PMA2014/Lagos demonstrates that REs moved around the area and conducted interviews in households that were selected from the sampling frame.

Figure 2: Comparison of GPS coordinates of listed and selected households in two PMA2014/Lagos enumeration areas
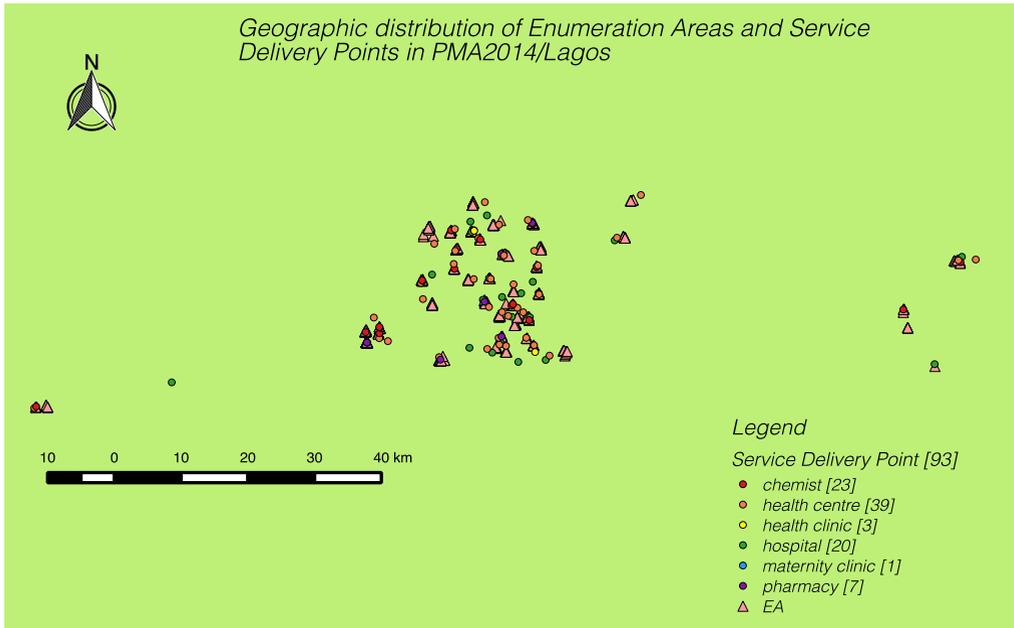

New research implications

In addition to improvements in quality control, the ability of the smartphone to record GPS coordinates also has research implications. Linkages can be made between households and service delivery points leading to a more precise understanding of the relationship between distance, access, and utilization of health services. In PMA2014/Lagos, GPS coordinates were obtained both at the household and at selected SDPs. Figure 3 below shows the distribution of SDPs by type and the geographic distribution of the household sample in PMA2014/Lagos. To protect the anonymity of our respondents, we have removed any identifying geographic layers, but the map below shows the scale and dispersion of EAs in the PMA2014/Lagos sample.

Figure 3: Geographic distribution of EAs and SDPs in PMA2014/Lagos sample

Utilizing GPS coordinates can generate more complex maps that show prevalence rates, average distance, or distribution of commodities and stock-outs, but GPS coordinates need not be limited to maps and figures. Distances between geographic points can also be calculated and used in analyses.

Table 1 below shows the average minimum distance that a woman in PMA2014/Lagos travels to reach service delivery points.

Table 1: Average minimum distance in kilometers to service delivery points

	All women n=782	95% CI
Any SDP	0.56	(0.51, 0.61)
Any SDP with 1+ FP method	0.61	(0.56, 0.66)
Managing Authority		
Public	1.01	(0.93, 1.09)
Private	1.51	(1.31, 1.71)
Facility Type[5]		
Chemist	2.62	(2.38, 2.85)
Health center	1.11	(0.99, 1.23)
Health clinic	12.12	(11.07, 13.17)
Hospital	3.78	(3.49, 4.07)
Pharmacy	7.83	(6.85, 8.80)

On average, women live within one kilometer of a service delivery point that offers at least one method of family planning, however distance to each type of facility can vary greatly. As the map above demonstrates, EAs that lie on the outskirts of the survey area are served primarily by hospitals, pharmacists and chemists, with health clinics clustered primarily in the center of the surveyed area. As such, the average distance to a health center in Lagos state is high relative to the distance to other delivery points. Average distance to facilities may be associated with utilization of facilities and interventions that address a host of reproductive health issues, from contraceptive use and method choice, to skilled birth attendance, neonatal care, and anti-retroviral therapy (ART) compliance. The utilization of smartphones in the PMA2020 model make it possible to obtain nationally representative data that link households and service delivery points, contributing critical information on service availability and barriers to care.

10% resample of selected households

As a final quality control check, PMA2014/Lagos instituted re-interviewing a 10% sample of selected households by supervisors. In PMA2014/Lagos, three out of the 35 households originally selected and interviewed by the RE were randomly selected from each cluster. In these households, supervisors conducted a limited re-interview to confirm the accuracy of basic information. While this is not in itself a new innovation, the ability of the smartphone to capture both photographic and GPS information provide an additional layer of quality checking. During the RE's originalinterview, a photograph of the residential structure was taken, along

---

[5] Includes both public and private facilities

with GPS coordinates. When the supervisor in Lagos conducted the re-interview, they also recorded GPS coordinates and a photograph. This allows for three levels of data verification; photographic, GPS coordinates, and verbal confirmation with the household residents. If questions arose whether the correct household was surveyed, comparison of GPS coordinates and photographs, all aggregated to one database, allowed data managers to monitor the consistency of information quickly.

Discussion

We have outlined above four ways in which the integration of mobile technology can improve large-scale survey implementation. However, this integration is not without its challenges and any program considering using mobile technology on a wide scale should consider the benefits and drawbacks carefully.

As discussed in detail above, mobile technology brings with it numerous improvements. In PMA2014/Lagos, a survey of over 1,000 households, 800 women and 94 SDPs was fielded within six weeks, from the sampling frame construction through re-interviews. Within two weeks of completion, data were ready for preliminary analyses. Much of the rapidity in turnaround time was due to the elimination of data entry and editing through automatic aggregation and scripting. For surveys with complex skip patterns, such as PMA2014/Lagos, automatic scripting can greatly reduce the potential for skip or logic errors, substantially shortening the amount of time spent in the data editing and cleaning process. However, the technical expertise that is needed to write these programs and time necessary to develop and test the forms on the smartphone may not always be available. In PMA2014/Lagos, the initial scripting was dependent on technical assistance from Johns Hopkins and took several iterations to test and modify, while a paper-based survey, once finalized, can be easily photocopied. Once written, automatic scripts are simple to edit, but the initial time and cost investment can be significant. For less complex surveys with simple skip patterns, paper based surveys may be a preferable option. Similarly, while instantaneous aggregation on servers automatically generates databases and eliminates the need for double data entry, for smaller surveys, the time and cost of manual entry may not be overly burdensome.

Quality control can be greatly improved by using built-in capacities of smartphone, such as GPS recording and time-stamping, but these capacities alone cannot ensure high quality. Personal supervision is still a critical piece of any survey and mobile technology does not eliminate the need for human supervision. As with paper-based surveys, supervisors must ensure timely completion of interviews, communicate errors, and provide support to interviewers. Integration of mobile technology requires additional training in logistics and IT support, however. In PMA2014/Lagos, supervisors addressed challenges with mobile connectivity, lost or stolen phones, and questions from REs that arose simply from being unfamiliar with new technology. While PMA2014/Lagos was able to find supervisory staff with sufficient skills to overcome these challenges, it may not always be feasible to find personnel who have the survey

experience and technical skills to adequately supervise staff.

The integration of GPS coordinates directly into the form and database is one aspect of integrating mobile technology that cannot be replicated using paper forms. While GPS coordinates can be taken with a GPS reader during a paper-based interview, they cannot be stored automatically on the paper form; they must be either transcribed- leaving room for error - or built into a separate dataset and linked using unique identifiers, which again may leave room for errors. As we have demonstrated above, GPS coordinates can be used for both quality control (e.g. mapping the geographic spread of interviews) and as a research application (e.g. generating the ability to calculate distances to health facilities or analyze clustering of responses). GPS coordinates have the potential to be extremely valuable and the automatic integration of coordinates into interviews leaves little opportunity for lost or mismatched data.

New or remaining challenges

Some additional challenges unique to the use of mobile technology exist and some challenges will remain no matter what platform is used for survey implementation. While surveys were conducted offline in PMA2014/Lagos, internet access, provided either through wifi or the mobile network, was required to download blank forms from the server and upload completed forms to the server. Unreliable data connection was often a challenge when attempting to send forms and required multiple attempts to send to the server. In some instances, forms were lost in transit as a result of poor connection,

but were retrievable from the internal storage of the phone and/or from external storage. It is also possible for the mobile devices to be misplaced or stolen, which happened on at least one occasion, or for the internal or external memory to become corrupt. (Tomlinson et al., 2009) Automated internal and external backups of the phone reduce potential data loss and allow for retrieval of previously backed-up information.

Sufficient community sensitization to the conduct of mobile phone-assisted surveys and reassurance of participants regarding the confidentiality of their responses pose challenges not unlike those in standard paper surveys. Many potential respondents in Lagos were not sure whether or not to take our resident enumerators seriously, at first. REs were mistaken for "tax collectors"; "telecom marketers"; "political campaigners" etc. in some areas. Allowing REs to work in pairs rather than singly will add credibility to their work and as the survey is repeated over the next four years, community awareness is expected to increase.

Geographic and weather challenges are also present and will continue, regardless of the method of data collection. Data for this Lagos survey were collected during the rainy season and there was intermittent flooding in some of the enumeration areas. This lengthened data collection due to problems with accessibility and increased transportation costs for some REs.

Future applications/innovations

While PMA2020 is currently focused on family planning and water and sanitation, it need not be limited to these subject areas. Additional modules

on a range of topics can be designed, programmed, and implemented with ease. Once trained, REs can be deployed to gather national or sub-national estimates on a range of health and demographic matters. As smartphones become increasingly familiar, surveys can also be adapted to incorporate ACASI-style questions to gather sensitive information.

Researchers and health programs can use information from mobile surveys, including GPS readings, to tag prevalence of diseases and health-related events to localities and demographic characteristics of individuals in order to be able to design targeted interventions for communities. Such information can guide programmatic interventions, allocation of scarce resources and future research directions.

Innovations in mobile technology continue to advance at rapid speed and with it come opportunities to improve data collection for health systems. Development of devices that can diagnose and monitor a host of health conditions is underway, pushing health data wirelessly to providers and programs. Surveys have the potential to be pushed out from a central server to respondents themselves, making individual data collection possible on a range of conditions, such as home measured blood-pressure, blood sugar or weight. Such developments will take personalized medicine a step further and allow regular communication between clients and caregivers, while simultaneously improving surveillance systems and the ability to estimate the burden of disease at the national, and even global, level.

Authors' Contribution

L.Z. conducted analyses, generated figures, and wrote the paper. F.O. wrote the paper and interpreted findings. S.R. gave technical advice and edited the paper. All authors discussed the results and commented on the manuscript at all stages.

References

Hartung, C., Lerer, A., Anokwa, Y., Tseng, C., Brunette, W. & Borriello, G. 2010. Open data kit: tools to build information services for developing regions. ICTD '10: Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development, 18. doi:10.1145/2369220.2369236

Lane, S., Heddle, N., Arnold, E. & Walker, I. 2006. A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods for data collection. BMC Med Inform Decis Mak, 6. doi:10.1186/1472-6947-6-23

Seebregts, C., Zwarenstein, M., Mathews, C., Fairall, L., Flisher, A., Seebregts, C., Mukoma, W. & Klepp, K. 2009. Handheld computers for survey and trial data collection in resource-poor settings: development and evaluation of PDACT, a Palm Pilot interviewing system. Int J Med Inform, 78: 721-31.

Shilton, K. 2012. Participatory Personal Data: An Emerging Research Challenge for the Information Sciences. J Am Soc Inf Sci Technol, 63: 1905-1915.

Shirima, K., Mukasa, O., Schellenberg, J., Manzi, F., John, D., Mushi, A., Mrisho, M., Tanner, M., Mshinda, H. & Schellenberg, D. 2007. 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, 4. doi:10.1186/1742-7622-4-5

Tacit Dynamics 2014. Højbjerg, Denmark. http://www.tacit.dk/foldersync.

Tomlinson, M., Solomon, W., Singh, Y., Doherty, T., Chopra, M., Ijumba, P., Tsai, A. & Jackson, D. 2009. The use of mobile phones as a data collection tool: a report from a household survey in South Africa. BMC Med Inform Decis Mak, 9. doi: 10.1186/1472-6947-9-51

Van Heerden, A., Norris, S., Tollman, S., Richter, L. & Rotheram-Borus, M. 2013. Collecting maternal health information from HIV-positive pregnant women using mobile phone-assisted face-to-face interviews in Southern Africa. J Med Internet Res, 15: e116. doi: 10.2196/jmir.2207