Using Mobile Phone Data Collection Tool, Surveda, for Noncommunicable Disease Surveillance in Five Low- and Middle-income Countries

Yang Song[1], Rachael Phadnis[1], Jennifer Favaloro[1], Juliette Lee[1], Charles Q. Lau[2], Manuel Moreira[3], Leenisha Marks[2], Matías García Isaía[3], Jason Kim[4], and Veronica Lea[1]*

[1] Global Noncommunicable Diseases Branch, Division of Global Health Protection, Center for Global Health, Centers for Disease Control and Prevention, Atlanta, GA
[2] RTI International, Research Triangle Park, NC
[3]Innovative Support to Emergencies Diseases and Disasters, Seattle, WA
[4]Invesco Ltd., Atlanta, GA

Abstract

Objectives: The Noncommunicable Disease (NCD) Mobile Phone Survey, a component of the Bloomberg Philanthropies Data for Health Initiative, determines the prevalence of NCDs and their associated risk factors and demonstrates the use of mobile phone administered surveys to supplement periodic national household surveys. The NCD Mobile Phone Survey uses Surveda to administer the survey; Surveda is an open-source, multi-modal software specifically developed for the project. The objective of the paper is to describe Surveda, review data collection methods used in participating countries and discuss how Surveda and similar approaches can improve public health surveillance.

Methods: Surveda features full-service survey design and implementation through a web application and collects data via Short Messaging Service (SMS), Interactive Voice Response (IVR) and/or mobile web. Surveda's survey design process employs five steps: creating a project, creating questionnaires, designing and starting a survey, monitoring survey progress, and exporting survey results.

Results: The NCD Mobile Phone Survey has been successfully conducted in five countries, Zambia (2017), Philippines (2018), Morocco (2019), Malawi (2019), and Sri Lanka (2019), with a total of 23,682 interviews completed.

Discussion: This approach to data collection demonstrates that mobile phone surveys can supplement face-to-face data collection methods. Furthermore, Surveda offers major advantages including automated mode-switch, question randomization and comparison features.

Conclusion: Accurate and timely survey data informs a country's abilities to make targeted policy decisions while prioritizing limited resources. The high acceptance of Surveda demonstrates that the use of mobile phones for surveillance can deliver accurate and timely data collection.

Keywords: mobile phone survey, noncommunicable diseases, LMICs

*Corresponding Author: vcl7@cdc.gov

DOI: 10.5210/ojphi.v12i2.10574

Introduction

Reducing the burden of non-communicable diseases (NCDs), including cardiovascular diseases, cancer, diabetes, and chronic lung diseases, is a major challenge in international development of the 21st century, and in 2011, the United Nations (UN) declared NCDs as such [1-3]. Furthermore, in 2012, the World Health Assembly endorsed the WHO Global Action Plan 2013-2020, which set for Member States the voluntary target of a 25% relative reduction in premature mortality from NCDs by 2025 [4]. Given the burden of NCDs, one of the UN Sustainable Development Goals also calls for a one third reduction in premature mortality from NCDs by 2030 [5]. In 2015, 70% of global deaths (39.5 million out of 56.4 million) were due to NCDs. The burden of these diseases rests disproportionately with low- and middle-income countries (LMICs), where more than 75% of NCD deaths occur [6]. NCDs are the leading causes of death in developed countries and will increasingly dominate the global pattern of death [7,8]. In addition, the estimated output loss, including loss of productivity and healthcare costs, attributed to NCDs is US$47 trillion if they are not addressed [4]. Despite growing evidence of the epidemiological and economic impact of NCDs, the global response to the problem remains inadequate in terms of financing and attention [9]. For example, while NCDs contribute to 50% of global disability adjusted years, the conditions only received 1% of total donor assistance for health in 2011 [10].

The mobile phone industry is making striking contributions to cross-sector innovations, including the health sector with NCD management and prevention, accomplished by mhealth methods such as health education, email and text reminders and data collection [11]. The existing mobile phone technology landscape can serve as a catalyst to scale up NCD data collection, dissemination and use. The systematic monitoring of risk factors is essential for a country's ability to prioritize essential resources and make sound policy decisions to address the growing NCD burden. In many LMICs, the systematic monitoring of risk factors is completed using household surveys, which means that data is collected via face-to-face interviews conducted in respondents' homes [12,13]. This method can be labor intensive, cost-prohibitive, and infrequent. Mobile phone surveys offer an opportunity to supplement traditional household health surveys given mobile phones can produce high-quality data for short and frequent surveys more cheaply [14]. According to the International Telecommunications Union (ITU), in 2018 there were more mobile phone subscriptions than people with 108 mobile-cellular subscriptions per 100 inhabitants globally and in LMICS there is one mobile-cellular subscription on average for every inhabitant [15]. As mobile phone access and ownership continue to increase globally, mobile phone surveys could produce timely, affordable and accurate data to monitor and address NCD trends.

The Bloomberg Philanthropies Data for Health Initiative, launched in 2015, aims to strengthen the collection and use of critical public health information through multiple components. The NCD Mobile Phone Survey component of the initiative determines the prevalence of common NCDs and their associated risk factors and demonstrates the feasibility of using mobile phone surveys as an interim data collection method to supplement periodic and often infrequent national NCD household surveys. The NCD Mobile Phone Survey is a representative survey of adults 18 years of age and older. The survey uses

standardized instruments and procedures, including a core questionnaire with optional questions and survey design using random digit dialing (RDD). The surveys are implemented by participating countries and ministries of health (MoHs) in collaboration with relevant ministries of information and technology, with technical assistance from Centers for Disease Control and Prevention (CDC), RTI International, and Innovative Support to Emergencies Diseases and Disasters (InSTEDD).

The NCD Mobile Phone Survey data collection utilizes Surveda, an open-source, multi-modal mobile phone survey tool specifically developed for the project. It is a flexible platform that can deploy mixed modal surveys (including Short Messaging Service [SMS], Interactive Voice Response [IVR] and mobile web), perform comparisons of questionnaires and mode sequences, uphold the highest data privacy standards, simplify the user experience, enable local and cloud hosting capacities, and contribute to the open-source community. Surveda is designed to be easily implemented in any data hosting environment and to quickly and efficiently deploy mobile phone surveys at scale.

Since 2015, Surveda has been used to collect NCD data in 7 countries, and data for 5 countries (Zambia [2017], the Philippines [2018], Morocco [2019], Malawi [2019], and Sri Lanka [2019]) are presented here, including the number of mobile phone number dialed, the number of surveys completed, response rates, and key demographics of respondents. This paper also describes in detail the survey technology tool, Surveda, reviews methods used for the NCD Mobile Phone Survey data collection in participating countries, and discusses how Surveda and similar approaches can improve public health surveillance.

Methods

Survey Implementation

The NCD Mobile Phone Survey is a cross-sectional survey with the target population being adults 18 years of age or older with a mobile phone number within the project site's mobile phone number series. The core questionnaire includes demographics, questions to assess how many mobile phones the respondent uses and whether anyone else uses the mobile phone they are using to respond, and questions on NCDs and related behavioral risk factors. The questionnaires are adapted to each country's context before implementation.

During data collection, a respondent is contacted via their mobile phone through a call or SMS and consent (opt in or out) is obtained before they are screened to determine eligibility. Once eligibility (age 18 years and above) is determined, the respondent is asked to answer questions on NCDs and associated key behavioral risk factors, including tobacco and alcohol use, diet (fruits, vegetables, salt), diabetes, and hypertension. A specific number of recontacts are set for each survey if the respondent does not answer or finish the survey on the first contact. Once the recontacts are exhausted, no additional attempts were made to contact the respondent.

The NCD Mobile Phone Survey used a two-phase sample design where in phase 1, a random sample of mobile phone numbers were selected using simple random sampling, and in phase 2, from the sample of mobile phone numbers, each mobile phone number was allocated to the age and sex strata to which it is a member. Once the sample size for an individual stratum was met, any respondents meeting the criteria for a filled stratum were thanked for their participation and the interview was terminated. Data regarding their age and sex was retained for sampling weight adjustments. This survey design yielded nationally

representative prevalence estimates of NCDs and key behavioral risk factors for males and females over the age of 18 years. Ethical clearance for survey implementation was obtained in all five countries.

Mobile Phone Data Collection Tool Features

Surveda was designed to help MoHs conduct surveillance of NCDs in line with the objectives of the NCD Mobile Phone Survey but can be used for any type of survey. It features full-service survey design and implementation through a web application. Currently, Surveda can collect data via SMS, IVR, mobile web, and mixed modes. The mode(s) of the surveys, the way or method which the surveys are sent out by, in individual countries are determined by literacy rate, smartphone penetration, and overall technology usage. In an SMS survey, respondents receive and answer questions via text messaging. In an IVR survey, respondents listen to prerecorded questions and select responses by pressing numbers on the mobile phone keypad. In a mobile web survey, respondents receive a text containing a link that opens the survey within a webpage. While the former two modes can be used across all types of mobile phones, the latter, mobile web, can only be used on web browser enabled phones, such as a smartphone.

The Surveda tool (GNU General Public License 3) is a web application that is built on top of existing open-source tools - specifically InSTEDD's *Verboice* and *Nuntium*, and leverages popular open-source web technologies including ReactJS, Phoenix Web Framework, and MySQL databases. Surveda's unique functions and features are designed to uphold the highest standards of data quality according to best practices:

- *Mixed mode survey deployment*- surveys can be deployed in a combination of two modes, one primary mode and one fallback mode, configured by the user. If a participant does not respond to the primary mode of contact, Surveda triggers the fallback mode to deploy the survey. This functionality leverages the strengths of the modes to reach participants [16].
- *Comparisons*- mode sequence and questionnaire can be assessed by running two surveys in parallel and evaluating data quality (e.g., response rate) of the two.
- *Languages*- currently more than 400 languages are supported by the tool for international use.
- *Quota targets*- the number of completed interviews for specific categories (i.e. strata) can be set in Surveda, such as age, gender, etc.
- *Scheduling*- the user can determine the days of the week, time zone, hours of the day, and days to block out such as national holidays.
- *Call-back sequence*- the number of times a phone number is contacted via each mode that is configured and the time interval between each contact can be set. Users can include the number of times that the primary mode is tried as well as the number of times a secondary mode will be tried (if a secondary mode is used).
- *Question randomization*- Surveda includes a function to randomize the sequence of questions each participant is sent. This feature aims to reduce question item non-response due to survey fatigue. Users create sections within which a question or questions are contained and then these are randomly shuffled during survey execution. Users can also specify sections that will not be randomized.

Figure 1 summarizes the available functions of Surveda. Users can design or upload questionnaires, upload sample of mobile phone numbers, configure channels and modes, set schedules, define call-back

protocol and timelines, and deploy surveys. Surveda also employs survey management features including real-time survey monitoring and survey progress visuals as well as data downloads in several formats. Respondents receive and respond to surveys using their mobile phones and their data are sent to the tool and stored securely. For the NCD Mobile Phone Survey, the survey design inputs are defined within an implementation plan by the MoHs in collaboration with all partners.

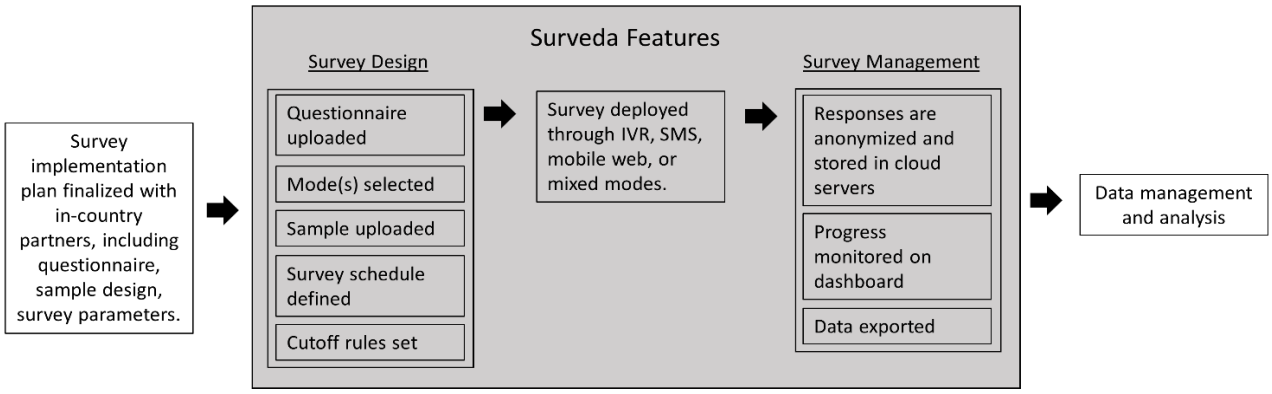

Figure 1. Functions of Surveda

Survey Design using Surveda

Surveda's survey design process employs five main steps: creating a project, creating questionnaires, designing and starting a survey, monitoring survey progress, and exporting survey results. These steps are sequential, as portrayed in Figure 2.

Figure 2. Surveda survey design process

The first step in this process is to *create a new project*. A project is the top level of organization used to store and execute surveys. In a single project, a user may have multiple surveys—for example, separate surveys on diet, physical activity, and alcohol use. For the NCD Mobile Phone Survey, the in-country ministries of health took the lead in creating a project, which stores all the components of the survey including questionnaire, channels, and data.

The next step in the process is to *create a questionnaire*. The questionnaire contains the questions, translations, responses, skip logic, and question randomization set up. For survey deployment, users must create questionnaires tailored to the mode selected (i.e., SMS, IVR, mobile web) for the survey.

Next, using the interactive, step-by-step tool, users can *design the survey* by selecting a created questionnaire, uploading the sample of mobile phone numbers, selecting the survey mode, setting days and times for sending invitations to potential respondents, setting the number of re-contacts, and time between re-contacts for both the primary and secondary modes.

To use the Surveda tool at scale, typically it is necessary to set up agreements with Mobile Network Operators (MNOs) or aggregators to establish channels that allow Surveda to exchange messages and calls with respondents. Channels established through MNOs or aggregators would typically result in the ability to dial 20 or more calls concurrently.

Users can monitor the survey progress through Surveda's dashboard, which displays the real-time status of all calls sent through the established channels. In addition, the dashboard displays the number of completed interviews, partial interviews, and other types of non-response (e.g., call failed, refusal). This allows survey managers to gauge the overall performance of the survey, estimate how long the survey will take to achieve the desired sample size, adjust course if necessary, and report on survey progress. In addition, survey data can be exported even as the survey is still in progress.

Surveda can *export survey data* for multiple purposes. A menu of four data files can be exported including the survey results from completed surveys, the disposition history or call status for each survey respondent, phone numbers for respondents who completed the survey for incentive distribution, and the interactions records between Surveda and the respondent. Currently, all exports are available in a comma delimited (CSV) format.

Data Hosting

For this project, there are two types of data hosting: permanent data hosting and temporary data hosting. Permanent data hosting refers to the long-term storage of completed survey results, and temporary data hosting refers to storage of the data within Surveda while the survey is running. The temporal data hosting is cloud based using Amazon Web Services (AWS).

From a technical standpoint, the project protocol recommends that permanent data is hosted on a 3rd party cloud provider environment (e.g., AWS, Microsoft Azure). However, if the implementing agency has an existing system for storing and managing these data and the in-country policies requires a data hosting solution other than cloud hosting, an alternative solution may be created. Ultimately, the implementing agency determines where the survey data are stored permanently as they own the data. Privacy under any of the hosting options is upheld to the latest security requirements and maintenance protocols. Under all the options, subscriber mobile phone numbers were only visible during sample upload; all data are unlinked, and de-identified in the tool as well as in data exports.

Analysis

Data on call flow including the number of mobile phone numbers dialed to the number of surveys completed are presented in the Results section below. The American Association for Public Opinion Research (AAPOR)-defined response rate, cooperation rate, refusal rate, and contact rate were calculated according to definitions and formulas in Appendix I. The number and proportion of respondents who completed the survey in primary and secondary modes are presented in the Results section. The age, sex,

and education of respondents who completed the survey in primary and secondary modes are also presented. In addition, we used the chi-square test to test if the aforementioned characteristics of respondents who completed the survey were associated with switching from the primary mode to the secondary mode.

Results

Country Specific Implementation

The NCD Mobile Phone Survey has been successfully implemented in five countries: Zambia (2017), the Philippines (2018), Morocco (2019), Malawi (2019), and Sri Lanka (2019). The surveys were programmed in 15 languages across the five countries. Furthermore, surveys were programmed in various non-Roman scripts such as Arabic and Hindi. Surveys were deployed via SMS, IVR and mobile web. To date approximately 112 Ministry of Health staff, across seven countries, have been trained on using Surveda. On average, data collection lasted 56.4 days in each country, with 84.0 interviews completed per day. In three countries (Zambia, Philippines and Morocco), due to unavailability of a list of active mobile phone numbers, implicit sampling frames that contained all possible mobile phone numbers were used. This type of sampling frame will contain a large proportion of non-active mobile phone numbers. Explicit sampling frames were used in Malawi and Sri Lanka with known active subscribers. An incentive of $1 USD was provided to participants who completed the survey in all countries except for Sri Lanka. Table 1 below summarizes the call flow for each country.

Table 1. Mobile phone numbers dialed and interviews obtained

Response Level	Zambia	Philippines	Morocco	Malawi	Sri Lanka	Total
Mobile phone numbers dialed	339,073	977,957	326,359	238,570	220,101	2,102,060
Respondent answers call or text	32,386	11,208	29,581	29,608	50,819	153,602
Respondent consented	25,045	8,523	16,667	20,889	18,464	89,588
Respondent screened	13,010	7,555	9,030	13,886	15,031	58,512
< 18 (Ineligible)	1,816	1,463	69	2,504	1,417	7,269
>= 18, stratum full (Rejected)	4,688	2,302	5,057	5,074	8,080	25,201
>= 18, stratum not full (Eligible)	6,506	3,790	3,904	6,308	5,534	26,042
Respondent completed interview	6,056	3,673	3,515	5,814	4,624	23,682

Out of 2,102,060 mobile phone numbers dialed across five countries, 153,602 (7.3%) resulted in contact with potential respondents, of which 89,588 (58.3%) consented to participate. Of those who consented, 58,512 (65.3%) completed the screening questions about age and sex. Ultimately, out of the 26,042 respondents who were eligible to participate, 23,682 (90.9%) provided an interview. These include both fully completed and partial interviews, with partial interview defined as having completed the demographic module and one NCD question. More information on the sampling strategy for the survey can be found on www.ncdmobile.org.

Table 2. AAPOR rates

Country	Response Rate 2 (RR2)	Cooperation Rate 2 (COOP2)	Refusal Rate 1 (REF1)	Contact Rate 1 (CON1)
Zambia	2.0%	55.7%	1.6%	3.5%
Philippines	0.5%	84.8%	0.1%	0.6%
Morocco	1.2%	84.6%	0.2%	1.4%
Malawi	3.3%	63.9%	1.9%	5.2%
Sri Lanka	2.3%	31.8%	4.9%	7.2%

Table 2 shows the response rates, cooperation rates, refusal rates, and contact rates in each country. Response rate ranged from 0.5% in the Philippines to 3.3% in Malawi. Cooperation rate was lowest in Sri Lanka at 31.8% and highest in the Philippines at 84.8%. Refusal rate stayed below 5% for all countries, ranging from 0.1% in the Philippines to 4.9% in Sri Lanka. The highest contact rate was found in Sri Lanka at 7.2% and lowest was 0.6% in the Philippines.

Mixed Mode Implementation

Table 3. Completed interviews by mode

Mode	Zambia		Philippines		Morocco[s]		Malawi		Total
	N	Col %	N	Col %	N	Col %	N	Col %	
Primary	3,467	57%	2,346	64%	2,908	83%	3,148	54%	11,869
Secondary	2,589	43%	1,327	36%	607	17%	2,666	46%	7,189

[s]Only one out of the three MNO's subscribers reached during primary mode were contacted using secondary mode.
Note: Sri Lanka is not included in the table as the survey was conducted in only one mode.

As shown in Table 3, across all four countries, most interviews were completed in the primary mode; however, an additional 7,189 interviews were completed due to the availability of a secondary mode, with 37.7% of all interviews completed in second mode. Table 4 shows completed interviews by primary and secondary mode, age, gender, and educational status. With the availability of a secondary mode, 199 to 888 additional interviews in the oldest (45+) age group were completed, which is consistently the hardest group to contact via mobile phone in LMICS [17]. Regarding gender, there was a range of 297 to 1,277 additional females who completed the survey due to the addition of a secondary mode. Lastly, the secondary mode allowed additional completed interviews in the lowest education range, which can often be under-represented in mobile phone surveys.

Table 4. Completed interviews by mode, age, gender and educational status

Mode		Zambia		Philippines		Morocco		Malawi	
	Age	n	%	n	%	n	%	n	%
	18-29	1720	50	1151	49	1277	44	1651	52
	30-44	1188	34	780	33	938	32	1116	35
Primary	45+	559	16	415	18	693	24	381	12
	18-29	1256	49	443	33	195	32	1473	55
	30-44	852	33	537	40	199	33	888	33
Secondary	45+	481	19	347	26	213	35	305	11
P-value*		0.04124		< 0.00001		< 0.00001		0.1016	
Gender									
	Female	1436	41	1258	54	1350	46	1057	34
Primary	Male	2031	59	1088	46	1558	54	2091	66
	Female	1277	49	642	48	297	49	1013	38
Secondary	Male	1312	51	685	52	310	51	1653	62
P-value*		< 0.00001		0.002524		0.2799		0.0005019	
Education									
	1[a]	409	12	68	3	406	14	169	5
	2[b]	430	12	41	2	420	15	886	29
Primary	3[c]	2602	76	2219	95	2017	71	2044	66
	1[a]	205	8	64	5	119	20	211	8
	2[b]	275	11	30	2	100	17	893	34
Secondary	3[c]	2093	81	1200	93	379	63	1523	58
P-value*		< 0.00001		0.003584		0.0004241		< 0.00001	

*The chi-square test was used.
[a]This refers to no education across all four counties.
[b]This refers to primary education in Zambia and Morocco; elementary education in the Philippines; and Standard 1-8 in Malawi.
[c]This refers to secondary and more than secondary education in Zambia; secondary, post secondary and college in the Philippines; high school and college in Morocco; and secondary and tertiary in Malawi.

Data Hosting

In the five countries that have completed the survey, various temporary data hosting solutions were employed (see Table 5 below).

Table 5. Temporary data hosting solutions

Country	Mode	Data Hosting		
		Surveda	Verboice	Nuntium
Zambia	IVR, SMS	Local data center	MNO, local data center	MNO, local data center
Philippines	SMS, mobile web	AWS	N/A	AWS
Morocco	IVR, SMS	AWS, ONA	MNO	AWS
Malawi	IVR, SMS	AWS, ONA	AWS	AWS
Sri Lanka	IVR	AWS	MNO	N/A

In the countries where the project was implemented, different data hosting solutions were used including: local hosting with MNO, AWS, and ONA. Hosting solutions were determined based on local laws and regulations with the hosting environment as well as capacities to provide best practices for data hosting. For example, in Zambia, it was required that data from governmental surveillance efforts be hosted at the national data hosting center, Zambia National Data Center. All other countries used AWS, of which a subset of countries used ONA as a final data destination. ONA is a data hosting platform that serves as a data destination, which has been in use by some of our in-country governmental partners.

Discussion

This approach to data collection demonstrates that mobile phone surveys can be used to supplement more traditional methods of data collection. Below are lessons learned based on conducting the NCD Mobile Phone Survey in five countries.

Country Implementation

The high contact rates in Sri Lanka and Malawi are likely due to the usage of a list of active mobile phone numbers rather than using lists of all possible mobile phone numbers generated by known prefixes within each country. It is interesting that even though the contact rate is high in Sri Lanka, its cooperation rate is the lowest among all countries. This could be attributed to Sri Lanka being the only country that did not distribute incentives to participants who completed surveys. This finding suggests that incentive-based surveys are more successful than non-incentive based surveys, as shown by previous studies [18,19]. Another finding of note is that while the Philippines had the highest contact rate and lowest refusal rate among all countries; it also had the lowest response rate. One explanation is that the response rate is calculated using all the mobile phone numbers dialed as the denominator and a large percentage of the numbers that were dialed in the Philippines were likely not active numbers. However, once someone is contacted at an active mobile phone number, they are likely to cooperate and complete the survey.

This elucidates the importance of using an explicit sampling frame versus an implicit frame which contains vast amounts of inactive numbers, given Sri Lanka and Malawi had the highest response and contact rates. We must dial more numbers to reach a working number when inactive numbers are not excluded from the frame. Including these sampling units in the calculation that are incapable of responding lower the

response rate. Additionally, explicit frames screened for active mobile phone numbers generally increase efficiency in survey implementation by reducing data collection time.

Mixed Modes

Previous literature suggests that mixing modes can increase response rates and improve the quality of mobile phone survey data [15]. Our results from the implementation of mobile phone surveys in Zambia, the Philippines, Morocco and Malawi support the use of mixed modes for survey deployment as it yielded a larger number of completed interviews and increased response rates, thus potentially reduced bias and non-representative samples [15]. Using a secondary mode in addition to a primary mode also showed that the secondary mode was able to capture different populations based on age, gender and education in the majority of the countries that implemented the NCD Mobile Phone Survey. For example, in the Philippines, more women completed the survey in the second mode than men, whereas in the first mode more men completed the survey. It is also evident in Morocco that the second mode was able to capture more people of lower education status, who are under-represented in the first mode. However, it should be noted that in Morocco, only a fraction of respondents were recontacted using the secondary mode given agreement was only reached with one out of the three MNOs available for primary mode. This could have impacted the percentage and characteristics of respondents who completed the survey using the secondary mode.

Each mode of survey delivery has its advantages and disadvantages that could attract different populations. For example, mobile web surveys can only be executed on web browser enabled mobile phones, which can be more expensive to own. SMS surveys require the respondent to be literate, which IVR surveys do not. Therefore, delivering surveys using mixed modes should be able to capture a more representative sample, which can be seen from our survey findings and previous literature [16]. Based on a landscaping of the current data collection software available, Surveda is the first tool to employ and automate the use of a primary and fallback mode using SMS and IVR.

Sustainability

Surveda, as a data collection tool, is sustainable due to the reusability of the infrastructure, acceptance by countries, and ease of use, as assessed by in-country partners during user feedback sessions. Sustainability is also one of the objectives of the project given there has been a proliferation of short-lived digital health tools in recent years [20]. Based on engagement with the countries during protocol implementation, it is evident that in-country partners are interested in this innovative approach to collect site-representative NCDs-related information and beyond. To that end, the MNO channel infrastructure is configured for the MoHs to reuse for the duration of the contracts. This includes maintenance of hardware, rack space, VPN, SMPP, and technical configurations. The same strategy is utilized for data hosting so that configurations for server, Surveda software, and VPN remain and future data collection efforts can be easily accomplished. Three of the five countries that have completed the NCD Mobile Phone Survey are planning to implement a second round of the survey, utilizing MNO channels established during implementation.

The mobile phone data collection method, facilitated by Surveda, successfully simplified and performed procedures that are traditionally employed by household surveys. For example, staffing for the mobile phone survey was minimal, requiring 5-10 core staff members. Given that surveys are deployed via SMS and IVR, no interviewers nor field staff supervisors were hired. Once the channels to deploy SMS or IVR

are established, MoHs are able to reuse the channels for repeat NCD surveillance or other public health data collection efforts. Country implementing teams trained on Surveda are equipped to design future surveys using the process outlined above to deploy and monitor the surveys, all in one tool. Throughout the implementation of the NCD Mobile Phone Survey, activities were undertaken with the aim of enabling MoH staff's subsequent use of the survey methods and tools. For example, step-by-step hands-on training on Surveda were provided to the core implementing in-country teams to build capacity for using the tool.

In-country partners understand the flexibility of the Surveda tool for other use cases, such as community- and facility-level surveillance. For example, Zambia MoH has expressed interest to use Surveda for a schistosomiasis survey and the Municipal Corporation of Greater Mumbai is considering employing the tool for a livability index. Ecuador and Sri Lanka's MoHs have recently conducted COVID-19 mobile phone surveys using Surveda and existing infrastructure set up for their NCD Mobile Phone Surveys. These surveys demonstrate the speed and utility of gathering population-level data to inform effective and rapid decision making during a pandemic. It is promising that several countries' core implementing teams have used the tools and methods of the NCD Mobile Phone Survey for other topics of interest outside of NCDs to better the health of their countries' population.

Open-Source

There has been increasing demand to better coordinate existing digital health initiatives for potential collaborations [21]. Surveda is an existing global technology platform that can strengthen the value and impact of digital health investments, improve coordination and facilitate scale. One important feature of Surveda that offers potential for collaborations and enables its sustainability is that it is an open-source tool, which means that Surveda is publicly accessible to individuals and organizations interested in creating their own surveys. The source code for Surveda is freely available online and can be modified as needed. Open-source software produces cost savings and provides researchers the opportunity to develop software for their own specific needs, then share it with others doing similar work.

Data Hosting

Using cloud hosting services was proven to be the fastest method to begin deploying surveys. It enabled the technical team to standardize the system's setup procedures across different country implementations, making it easier and cheaper to maintain over the project's lifecycle. Different hosting providers offered a wide variety of systems and protocols to setup and access the servers they provide, with different sets of rules and conditions that increase the complexity of software deployment and maintenance. Therefore, cloud hosting reduced risk, improved the sustainability and security of implementations, and is the most cost efficient of all the hosting options used for this project. The use of cloud computing has been garnering attention for its aforementioned benefits and is increasingly being implemented [22].

Mobile Network Connections and Scalability

To send out surveys on a national scale, setting up agreements and establishing connections with the MNOs was a necessity. This process was often challenging and is a limitation, which contributed to prolonging the timeline of survey implementation. In the planning stages of survey implementation, several options were usually considered in each country: MNOs, local aggregators, and global aggregators. During the beginning stages of implementation, an extensive landscaping process was

conducted to determine the available options, given the diverse technology and regulatory environments that varied from country to country. Therefore, there was no one-size-fits-all standard solution available.

For the NCD Mobile Phone Survey, local aggregators were typically preferred due to their existing connections to all in-country MNOs. In some cases, however, they did not always meet the project's technical requirements, so MNOs and local aggregators were both used depending on the country. Working with global aggregators (e.g. Twilio) was another option but they have different levels of reliability in different countries and are potentially cost-prohibitive at large scales. Working with local MNOs provides more reliable and cost-effective connections but requires extensive one-time efforts in setting up and testing the connectivity, while dealing with a great variety of regulations, contractual agreements, technical solutions, response times and bureaucracy. The ITU has explored the importance of these partnerships in the context of information and communication technology for development, specifically for meeting the Sustainable Development Goals and provide useful guidance on establishing these partnerships [23].

One of the lessons learned is that for IVR it is usually enough to work with a single MNO in the country, meanwhile SMS tends to require working with each MNO (or an aggregator) to setup short codes and reverse billing, which prevents respondents from being charged for text messages while responding to the survey. Mobile web could also be a simple approach, if without IP whitelisting. This means respondents will be charged for data usage. However, testing confirms that data is minimal to complete the entire survey. To a smaller-scaled data collection effort, these may not be obstacles.

Limitations

The results of the mobile phone surveys are representative of mobile phone owners rather than the general public. However, as mentioned, mobile phone ownership is generally high in the countries where the survey was implemented. An additional limitation is that the information collected is based on self-report and may be subject to bias. Given the anonymous nature of the data collection, however, the risk of social desirability bias is reduced. Lastly, sending out surveys on a national scale involves setting up agreements and establishing connections with MNOS, which is often challenging.

Conclusion

The Bloomberg Philanthropies Data for Health Initiative successfully completed the first large scale NCD Mobile Phone Survey in five LMICs. The innovative mobile phone data collection methodology yielded a large sample in a relatively short period without the logistical efforts a traditional door-to-door survey would require. The rapid availability of data allows for the potential for speedy dissemination and use of the results. Accurate and timely data is essential for a country's ability to make policy decisions while prioritizing limited resources. The successful use of this technology demonstrates that the use of mobile phones for surveillance can deliver accurate and timely data collection.

Worldwide, utilization of mobile phones for data collection and research is increasing, including for applications such as public opinion surveys, health interventions, and citizen feedback [15]. Although mobile phone surveys are unlikely to fully replace door-to-door demographic and health surveys, monitoring and evaluation staff should consider the utility of Surveda, and mobile phone data collection as a supplemental method to provide timely data. These data add to the limited, but growing research base

documenting that mobile phone survey research in LMICs is feasible, fast, and potentially cost-effective for collecting data on NCDs and related risk factors and other priority topics critical for population health. Surveda can be especially useful during situations where face-to-face data collection is not feasible. For example, during the COVID-19 pandemic, using Surveda, Ecuador and Sri Lanka governments were able to quickly set up and conduct mobile phone surveys on COVID-19 topics to inform program and policy implementation.

With the growth of the digital community, Surveda has the potential to become an integrated component of a government-led harmonized approach to monitoring and achieving the Sustainable Development Goals and can even one day contribute to the non-health sectors as an agnostic data collection tool.

Acknowledgements

The authors wish to thank the governments of Zambia, Morocco, Philippines, Malawi, Sri Lanka, Ecuador and Mumbai for implementing the NCD Mobile Phone Survey.

Financial Disclosure

Funding for the Data for Health initiative NCD component and NCD Mobile Phone Survey is provided by the Bloomberg Philanthropies and the Government of Australia, Foreign Affairs.

References

1. Hosseinpoor AR, Bergen N, Mendis S, Harper S, Verdes E, et al. 2012. Socioeconomic inequality in the prevalence of noncommunicable diseases in low- and middle-income countries: results from the World Health Survey. *BMC Public Health*. 12, p. 474. [cited December 2019]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22726343.

2. Magnusson RS. 2010. Global health governance and the challenge of chronic, non-communicable disease. *J Law Med Ethics*. 38(3), p. 490-507. [cited November 2019]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20880237.

3. United Nations General Assembly. 2011. A/RES/66/2, Political declaration of the high-level meeting of the general assembly on the prevention and control of non-communicable diseases. [cited January 2019]. Available from: https://undocs.org/en/A/RES/66/2.

4. World Health Organization. 2013. Global Action Plan for the prevention and control of noncommunicable diseases 2013-2020. [cited April 2020]. Available from:

   https://apps.who.int/iris/bitstream/handle/10665/94384/9789241506236_eng.pdf;jsessionid=BC16DFA386FFE3DF57D7E59F1529DA73?sequence=1.

5. United Nations General Assembly. 2015. A/RES/70/1, Transforming our world: the 2030 agenda for sustainable development. [cited April 2020]. Available from:

   https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E.

6. World Health Organization. 2020. Noncommunicable Diseases Progress Monitor 2020. [cited April 2020]. Available from: https://www.who.int/publications-detail/ncd-progress-monitor-2020.

7. Beaglehole R, Yach D. 2003. Globalisation and the prevention and control of non-communicable disease: the neglected chronic diseases of adults. *Lancet*. 362(9387), p. 903-8. [cited November 2019]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/13678979.

8. Murray CJ, Lopez AD. 1997. Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet*. 349(9064), p. 1498-504. [cited November 2019]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/9167458.

9. Yach D, Hawkes C, Gould CL, Hofman KJ. 2004. The global burden of chronic diseases: overcoming impediments to prevention and control. *JAMA*. 291(21), p. 2616-22. [cited October 2019]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15173153.

10. Nugent R. 2015. Policy Brief: Bilateral and multilaterial financing for NCDs. WHO global coordination mechanism on the prevention and control of noncommunicable diseases. [cited September 2020]. Available from: https://www.who.int/nmh/ncd-coordination-mechanism/Policybrief5.2docx.pdf?ua=1#:~:text=The%20World%20Bank%20provides%2030,by%20each%20major%20donor%20source.&text=Bilateral%20donors%20are%20the%20dominant,providing%2052%25%20of%20overall%20DAH.

11. Peiris D, Praveen D, Johnson C, Mogulluru K. 2014. Use of mHealth systems and tools for non-communicable diseases in low- and middle-income countries: a systematic review. *J Cardiovasc Transl Res*. 7(8), p. 677-91. [cited April 2020]. Available from:

    https://www.ncbi.nlm.nih.gov/pubmed/25209729.

12. Labrique A, Blynn E, Ahmed S, Gibson D, Pariyo G, et al. 2017. Health Surveys Using Mobile Phones in Developing Countries: Automated Active Strata Monitoring and Other Statistical Considerations for Improving Precision and Reducing Biases. *J Med Internet Res*. 19(5), p. e121. [cited October 2019]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28476726.

13. Mahfoud Z, Ghandour L, Ghandour B, Mokdad AH, Sibai AM. 2014. Cell Phone and Face-to-face Interview Responses in Population-based Surveys. *Field Methods*. 27(1), p. 39-54. [cited November 2019]. Available from: https://journals.sagepub.com/doi/full/10.1177/1525822X14540084.

14. Ballivian A, Azevedo JP, Durbin W. 2015. Using Mobile Phones for High-Frequency Data Collection. In: Tonineeli, D, Pinter, R & de Pedraza, P (eds.) *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies.*, p. 21-39. [cited April 2020]. Available from: https://ubiquitypress.com/site/chapters/10.5334/bar.c/download/293/.

15. International Telecommunication Union. Statistics. Mobile-Cellular Subscriptions. 2020; Available from: https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx.

16. Lau CQ, Johnson E, Amaya A, LeBaron P, Sanders H. 2018. High stakes, low resources: what mode (s) should youth employment training programs use to track alumni? Evidence from South Africa.

Journal of International Development. 30.7, p. 1166-1185. [cited January 2020]. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/jid.3359.

17. Lau CQ, Cronberg A, Marks L, Amaya A. 2019. In search of the optimal mode for mobile phone surveys in developing countries. A comparison of IVR, SMS, and CATI in NIgeria. *Survey Research Methods*. 13(3), p. 305-318. [cited August 2019]. Available from: https://ojs.ub.uni-konstanz.de/srm/article/view/7375/6821.

18. Gibson DG, Wosu AC, Pariyo GW, Ahmed S, Ali J, et al. 2019. Effect of airtime incentives on response and cooperation rates in non-communicable disease interactive voice response surveys: randomised controlled trials in Bangladesh and Uganda. *BMJ Glob Health*. 4(5), e001604. PubMed https://doi.org/10.1136/bmjgh-2019-001604

19. Leo B, Morello R, Mellon J, Peixoto T, Davenport ST. 2015. Do mobile phone surveys work in poor countries? Center for Global Development Working Paper, (398).

20. World Health Organization. 2019. WHO guideline: recommendations on digital interventions for health system strengthening. World Health Organization. [cited September 2020]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK541905/.

21. World Health Organization. Digital Health Atlas. [cited September 2020]. Available from: https://digitalhealthatlas.org/en/-/.

22. Maaref S. (2012). Cloud computing in Africa situation and perspectives. Telecommun. Dev. Sect.-ITU, 70. [cited September 2020]. Available from: http://www.itu.int/ITU-D/treg/publications/Cloud_Computing_Afrique-e.pdf.

23. International Telecommunication Union. 2018. M-Powering for Development: Turning Opportunities into Reality, 2018. [cited September 2020]. Available from: https://www.itu.int/en/ITU-D/Initiatives/m-Powering/Documents/mPowering_Report_2018_210918.pdf.

Appendix I

Response Rate 2 (RR2) = (I + P) / (I + P + R + NC + O + UH + UO)

Cooperation Rate 2 (COOP2) = (I + P) / (I + P + R + O)

Refusal Rate 1 (REF1) = R / (I + P + R + NC + O + UH + UO)

Contact Rate 1 (CON1) = (I + P + R + O) / (I + P + R + O + NC + UH + UO)

Reference: https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/Standard-Definitions2015_8thEd.pdf

Final Disposition Codes for RDD Telephone Surveys	Code	Conversion for Mobile Phones	Corresponding Surveda Disposition State	Surveda Definition
1. Interview	1			
Complete (I)	1.1		Completed	Answered all survey questions
Partial (P)	1.2	Demographic questions completed plus one NCD Question	Partial and Interim Partial	Answered at least once NCD question but did not finish the survey
2. Eligible, Non-Interview	2			
Refusal and break-off (R)*	2.1	Demographic questions completed and respondent eligible	Breakoff OR Started	Answered age and sex questions but did not answer any NCD questions
Non-contact (NC)	2.2	This code has no applicable definition in the MPS.		
Other (O)	2.3	This code has no applicable definition in the MPS.		
3. Unknown Eligibility, Non-Interview	3			
Unknown if housing unit	3.1	This code has no applicable definition in the MPS.		
Not attempted or worked	3.11	Call failed		
Always busy	3.12	Phone busy or network busy/down		
No answer (UH)	3.13	Subscriber status unknown	Unresponsive and Failed	No answer, possibly nonworking/non-active number
Telephone answering device (don't know if housing unit)	3.14	Voicemail		
Telecommunication technological barriers, e.g., call-blocking	3.15	Call blocking		
Technical phone problems	3.16	Bad audio quality (i.e., static, poor reception), Unable to connect because of network issues, Breakoff by respondent due to technical difficulties before Demographic questions began		
Ambiguous operator's message	3.161	Ambiguous error or ISDN code		
Other (UO)	3.9	**Breakoff** before Demographic questions were complete/eligibility determined, Pressed 3 to **refuse** the interview, Unable to understand language of interview, Immediate hang up, Temporarily out of service, or Part-time fax/data line, Out of coverage area	Refused OR Breakoff OR Started	Refused consent OR Answered some questions but stopped before age or sex
4. Not Eligible	4			
Fax/data line	4.2	Dedicated fax/data line		
Nonworking/disconnected number	4.3			
Nonworking number	4.31			
Disconnected number	4.32			
Temporarily out of service	4.33			
Pagers	4.44			
Nonresidence	4.5			
Business, government office, or other organization	4.51			
Institution	4.52			
Group quarters	4.53			
Person not household resident	4.54			
No eligible respondent	4.7	Less than 18 years	Ineligible	Under age 18
Quota filled	4.8	Respondent is 18 years or older but the stratum sample size has been attained. These respondents are rejected.	Rejected	Answered age and sex questions but quotas were full
Other	4.9	Phone or SIM (subscriber identity module) card not used		