COLLECTING DATA DURING AN EPIDEMIC: A NOVEL MOBILE PHONE RESEARCH METHOD

ELISA M. MAFFIOLI* (iD)
Department of Health Management and Policy, School of Public Health, University of Michigan, Ann Arbor, MI, USA

Abstract: This study developed a data collection method, combining (i) random-digit dialling and interactive voice response to sample and screen respondents and (ii) computer-assisted telephone interviewing, to survey 2265 respondents during the 2014 Ebola epidemic. The response, cooperation, refusal and contact rates computed according to the American Association for Public Opinion Research were 51.97, 52.62, 41.85 and 98.77 per cent, for interactive voice response and 91.10, 91.65, 8.30 and 99.40 per cent for computer-assisted telephone interviewing. A comparison with Demographic and Health Surveys confirmed that the sample is not nationally representative. However, this method offers promise for data collection at a low cost ($24) and without any in-person interaction. © 2020 John Wiley & Sons, Ltd.

Keywords: data collection; mobile phones; random-digit dialling; interactive voice response; computer-assisted telephone interviewing

1 INTRODUCTION

More and more researchers collect survey data through the use of mobile phones (Toninelli, Pinter, & de Pedraza, 2015). Mobile phone technology has removed significant barriers inherent to traditional survey data collection: It helps in gathering high-frequency panel data (Dillon, 2012; Hoogeveen et al., 2014; Ballivian, Azevedo, & Durbin, 2015), it provides timely access and monitoring of data collected through face-to-face surveys (Tomlinson et al., 2009; Schuster & Brito, 2011; Hughes, Haddaway, & Zhou, 2016) and it allows for fast and low-cost data collection (Schuster & Brito, 2011; Leo et al., 2015; Mahfoud et al., 2015; Garlick, Orkin, & Quinn, 2019). Yet as mobile phones

*Correspondence to: Elisa M. Maffioli, Department of Health Management and Policy, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, USA.
E-mail: elisamaf@umich.edu

are widely used as a data collection tool in developed countries where access is common, this research method is getting more and more traction in developing settings, too (Gibson et al., 2017).

Given the rise in mobile phone penetration rates in developing economies (World Bank, 2016), mobile phone surveys are increasingly used to gather national statistics and to conduct monitoring, bio-surveillance and disaster management (Gallup, 2012; Bauer, Akakpo, Enlund, & Passeri, 2013; Twaweza East Africa, 2013; Hoogeveen et al., 2014; van der Windt & Humphreys, 2014; Garlick et al., 2019). Researchers use mobile phones to conduct different types of mobile phone interviews (Lau, Cronberg, Marks, & Amaya, 2019): interactive voice response (IVR) surveys, which rely on a pre-recorded voice recording to ask questions to respondents; computer-assisted telephone interviewing (CATI), which requires trained interviewers to make live calls to respondents following a script provided by a software application; and SMS surveys, which require respondents to type an answer and send back a text message (Dabalen et al., 2016).

Yet gathering data in developing settings—where there might be weak institutions, limited resources and infrastructure, cultural constraints and low literacy—can be even more challenging than in developed countries (Grosh & Glewwe, 2000; Ganesan, Prashant, & Jhunjhunwala, 2012; Dabalen et al., 2016). In developing economies, for example, there is often no access to an initial list of contacts or publicly available data sources, and researchers need to gather baseline data themselves to have a sampling frame. Times of emergency, such as conflicts, infectious disease epidemics or weather-related disasters, when it is hard to reach respondents in person for interviews, exacerbate these challenges.

This study tested the feasibility of a novel mobile phone data collection method to interview more than 2000 respondents during the 2014 Ebola epidemic in Liberia. The research method uses (i) random-digit dialling (RDD) and IVR surveys to sample and screen respondents and (ii) CATI to conduct (30- to 45-minute) interviews.

This method builds upon the RDD selection approach outlined by Leo et al. (2015), in which the sample was randomly selected through an online platform and the data collection was performed through an IVR survey.[1] The authors assessed whether mobile phone surveys were a feasible and cost-effective approach to collect data in four middle-income or low-income countries, focusing on whether the method could reach a nationally representative sample and how to improve its survey completion rates. Similarly, L'Engle et al. (2018) used RDD and IVR surveys to collect survey data in Ghana, assessing the response rate and representativeness of the obtained sample compared with face-to-face national surveys.

This method also builds upon previous studies, which used CATI to collect high-frequency data. Hoogeveen et al. (2014) provided examples of phone surveys at high frequency in Tanzania and South Sudan through a call centre. A similar approach was used by Dillon (2012) to elicit data regarding farmer expectations, production and income levels over time. Demombynes et al. (2013) also used a similar high-frequency survey approach, where the authors randomized the level of incentives and the phone equipment to increase response rates in South Sudan. Finally, Garlick et al. (2019) compared frequencies of in-person or CATI interviews to micro-enterprises, and they found no difference in data quality or response rates between high-frequency CATI interviews and low-frequency in-person interviews.

---

[1]See Waksberg (1978) and Massey et al. (1997) for the use of RDD in the developed world.

By combining these established methods (RDD, IVR, CATI) in a novel manner, this study developed a data collection procedure that does not require in-person contacts, thus allowing researchers to gather survey data in challenging settings. In fact, although limitations in the application of mobile phones as unique data collection devices (Kempf & Remington, 2007) remain, evidence regarding their use—as a method both to gather survey data and to select and screen an initial sample of respondents to interview—is still lacking. This research method aims at overcoming two specific challenges related to data collection in developing countries and at times of emergency.

First, researchers begin field research by developing a sampling frame. Usually, they seek access to an initial list of contacts, such as a list of respondents from past studies or a list of phone numbers contained within the datasets from collaborating institutions (The World Bank Group, 2014; The World Bank Group, 2015). Alternatively, researchers may develop a sampling frame from publicly available data sources, such as large-scale national household surveys or population censuses, which provide the advantage of being nationally representative; or they simply gather data themselves through a face-to-face baseline survey. However, getting a sampling frame is challenging when data do not exist or face-to-face data collection is not feasible.

Second, in a state of emergency, reaching survey respondents in person for interviews can be challenging, or the risks and costs associated with data collection in order to have a large enough sample may be insurmountable. For example, at the time of the Ebola epidemic—or currently during Covid-19—in-person contacts should be limited, if not avoided altogether. Similarly, in the aftermath of hurricanes, floods or earthquakes, travelling to remote areas might be impossible. As weather-related disasters or infectious diseases remain a worldwide threat, especially in developing countries (United Nations Office for Disaster Risk Reduction, UNISDR, 2015), in-field data collection may not be always feasible.

This study proposes to use mobile phone technology as the sole tool for all stages of data collection, in order to overcome both of the aforementioned challenges. The research method allows researchers to conduct the entire data collection while relying solely on mobile phones, precluding the need for prior data or fieldwork activities to have a sampling frame or to gather survey data. While the studies mentioned above required at least one in-person interaction in order to facilitate data collection through mobile phones, the proposed method does not necessitate any physical contact with the respondents. This is key when a sampling frame is not available or fieldwork activities are excessively demanding, such as in the case of emergencies. Furthermore, this method was implemented and tested in a developing country where this type of technology is most needed.

The paper is organized as follows: Section 2 describes the method used to collect the data and the analysis conducted; Section 3 provides a description of the results by estimating call outcomes and rates, the representativeness of the survey sample and the implementation costs of this research method; Section 4 discusses the lessons learned; Section 5 concludes the paper.

2 METHODS

The goal of the initial project was to study the political economy of the 2014 West African Ebola epidemic (Maffioli, 2020), by gathering survey data on individuals' level of trust

and perceived corruption toward several institutions, and their opinions on the government's actions during the response. However, successfully accomplishing this goal required surmounting significant survey data challenges. There were indeed no baseline data available to select respondents, and gathering face-to-face data was impossible owing to the high costs and risks at the time of the epidemic. Thus, the project relied solely on mobile phone technology for both stages of the data collection procedure: (i) sampling and screening of the respondents and (ii) data gathering.

This study tested the feasibility of this research method, by conducting more than 2200 interviews in Liberia between October 2015 and June 2016. Call outcomes and response, cooperation, refusal and contact rates were calculated according to the American Association for Public Opinion Research guidelines (AAPOR, 2016). The representativeness of the survey sample was explored using the nationally representative Demographic and Health Surveys (DHS) (Demographic Health Survey Liberia, 2013) as a benchmark. Costs were also computed in comparison with similar mobile phone approaches used in past research studies.

2.1   Data Collection

2.1.1   Sampling and screening of respondents (IVR)

Owing to the lack of access to a sample available in the pre-Ebola period, an online platform called VotoMobile was employed to draw an initial list of respondents through RDD.[3]

Given the known structure of the mobile phone numbers in Liberia, the platform created a list of randomly generated phone numbers that fit that structure through an algorithm. The platform was set up to select phone numbers from the two main Liberian phone companies at that time (Liberian Telecommunications Authority, 2012), LonestarCell/MTN with a share of 49.55 per cent and Cellcom with a share of 40.36 per cent.

The platform went through the randomly generated phone numbers. It was set up to attempt up to four calls to the same phone number: After the first attempt, the second call was placed after 5 minutes, while the third and fourth calls were placed after 8 hours each. The calls were made 7 days a week from 8:00 a.m. to 8:00 p.m. Once the phone number connected and a person picked up the call—implying that it was an existing Liberian phone number—a short pre-recorded survey (IVR) informed the respondent that she/he had been selected for an interview. The IVR message asked the respondent three questions to gather her/his residence location at the beginning of the epidemic: (i) whether she/he lived in Montserrado County; (ii) if not, in which other county did she/he live; and (iii) which district did she/he live in.

The aim was to gather a heterogeneous sample of individuals from all 15 counties in Liberia, to compare individuals affected by Ebola with those unaffected by it. More specifically, the first question was set up to limit respondents from Montserrado County, the most urbanized and populous region in Liberia and where most of the Ebola cases were concentrated: Selecting respondents from other counties would allow for a heterogeneous

---

[2]The author lacked access to pre-existing baseline data or publicly available nationally representative surveys for which respondents' contacts were available and for which a sample could be drawn.
[3]VotoMobile is now called Viamo. https://viamo.io/

sample to make meaningful comparisons for the analysis.[4] If the respondent answered all of the three IVR questions, then she/he would also be informed that someone would call back from the local non-governmental organization (NGO) and that, upon completion of the CATI survey, she/he would receive $1 of free airtime for her/his phone as a sign of appreciation. The IVR survey was conducted between 21 and 31 October 2015.

As no restrictions were placed on the selection process through IVR, any person answering the phone was considered eligible. Following AAPOR (2016) guidelines, *complete interviews* (*I*) were defined as answering the three location questions to gather information on both the county and the district where the respondent resided at the beginning of the epidemic. *Partial interviews* (*P*) were defined as answering the first two questions of the survey to gather the county, but not the district. Because the IVR confirms that a real person picked up the call only when the respondent answers the first question, *refusals* (*R*) were defined as not answering the first IVR question, that is, whether the respondent resided in Montserrado County. *Break-offs* (*R*) were defined in a similar way for two reasons: First, following the definition of break-offs in AAPOR (2016) guidelines, the IVR survey was so short that answering the first question meant answering almost 50 per cent of the survey; second, partial and complete interviews already took into account the completion of the second and third questions in the IVR message. In practice, I conservatively categorized as break-offs phone numbers, which were dialled, for which the phone rang but the respondent did not pick up the call. According to VotoMobile coding, this category does not allow to distinguish whether the user did or did not deliberately respond to the call.[5] In addition, *unknown eligibility* (*UH*) was classified as phone numbers which were always busy.

Finally, phone numbers that were dialled but could not be confirmed as known working numbers were classified as *not eligible*. A high number of ineligible calls are expected because of the automated nature of the RDD calling system.[6] This category includes (i) phone numbers that never responded because the call never rang on a person's phone owing to an error on the provider's end. In practice, these are non-existing phone numbers resulting from the random dialling[7]; (ii) phone numbers temporarily out of service; (iii) phone numbers unable to connect due to specific technological issues; and (iv) phone numbers for which the call connected at the network level, but a valid connection to an individual's mobile phone could not be confirmed. These numbers were categorized as phone numbers that connected, but there was no or invalid selection, and correspond to quick hang-ups.

---

[4]The survey data were merged with administrative data from the Ministry of Health on Ebola cases to determine which districts and counties were affected by the epidemic.

[5]This category corresponds to the code 'NO_ANSWER' recorded byVotoMobile. Please see https://freeswitch. org/confluence/display/FREESWITCH/Hangup+Cause+Code+Table

[6]Based on some basic calculations, VotoMobile expected that no more than 37 per cent of the calls made by the online platform could be real phone numbers. In Liberia, there exist about 2.6 million owners of SIM cards. Because the platform tries six digits for each of the seven Liberian prefixes, this would total up to $10^6 \times 7 = 7\,000\,000$ possible phone numbers; $2\,600\,000/7\,000\,000$ corresponds to about 37% potential real numbers that we would have been able to reach.

[7]This category corresponds to the code 'NO_USER_RESPONSE' recorded by VotoMobile. Please see https:// freeswitch.org/confluence/display/FREESWITCH/Hangup+Cause+Code+Table. This code is used when a called party does not respond to a call-establishment message with either an alerting or connection indication within the prescribed period of time allocated. However, this is due to an error from the provider's side because the call never got connected to the phone number of the called party. These phone numbers were randomly generated by the VotoMobile platform and then called, but they could not connect as they were non-existing phone numbers. Note also that L'Engle et al. (2018), who used the same platform (VotoMobile), similarly categorized this group of calls.

2.1.2   Data gathering (CATI)

Screened and selected respondents through the IVR survey were called back by real enumerators from a local NGO to conduct a 30- to 45-minute interview.[8] The sample for CATI was selected for the initial research project (Maffioli, 2020) among the phone numbers called in stage 1, for which the IVR survey was either complete (I) or partial (P).

The enumerators from the local NGO were instructed to call the full list of phone numbers multiple times to reach the respondents. They also had the flexibility to re-contact the respondents at their most preferred time and to call them back when the survey was interrupted for any reason. Owing to budget constraints, the main data collection by the local NGO proceeded in two rounds. Round 1 was conducted between December 2015 and February 2016, while round 2 was conducted in June 2016.

The CATI survey was performed by 18 enumerators, five of whom were female and who had received training in human subject research and a training for mobile phone data collection. In addition to respondent and household socio-demographics, the survey tool collected data on the following: (i) political outcomes, such as self-reported level of trust in governmental and non-governmental institutions and people, perceived corruption of similar institutions and past voting behaviour; and (ii) Ebola-related questions, such as self-reported Ebola incidence in the community, the level of information received, the experience with the response and perceptions about the government's performance (see Maffioli, 2020; Gonzalez & Maffioli, 2020, for the use of the survey sample for other research).[9]

The data were collected in Kobo Toobox through mobile phone devices and then exported automatically in Excel, and quality checks were performed by the researcher daily. Data cleaning and analysis were conducted using Stata software v15. Ethical approval was obtained from the University of Liberia and Duke University.

The CATI interviews for which respondents gave verbal consent and that were more than 80 per cent completed were classified as *complete* (I). Because all respondents finished the entire survey, there were no CATI interviews classified as *partial* (P). *Refusals* (R) were defined as not agreeing to participate in the survey, while *break-offs* (R) were defined as phone numbers that were dialled for which the phone rang but the respondent did not pick up the call. *Unknown eligibility* (UH) was classified as phone numbers not screened for eligibility, for example, in the case respondents reported having already been interviewed. Finally, *not eligible* phone numbers were classified as those numbers which (i) were ineligible because respondents were younger than 18 years: This restriction was placed on the CATI selection process because sensitive information was asked; (ii) were temporarily out of service; and (iii) never responded because the call never rang on a person's phone owing to an error on the provider's end; in practice, because the call did not connect at the time of the CATI—between 2 and 8 months after the IVR survey—it was impossible to know whether the number was still existing and valid.

It is important to notice that this last group of phone numbers was active during the IVR survey implemented in October 2015, but the phone numbers turned out to be non-existing or non-active at the time of CATI, and this is why they could be classified as *not eligible*

---

[8]The implementation partner was a local Liberian NGOcalled Parley, based in Gbarnga, Bong County, that had past experience with phone surveys, also during the Ebola epidemic.
[9]To avoid potential priming effects due to the order of the questions between political outcomes and experience with Ebola, half of the sample was randomized to receive the political outcomes section first, while the second half received the Ebola-related questions first. No statistically significant differences were found in the data collected between the two samples.

(4.31) according to the AAPOR (2016) guidelines. This classification assumes the phone numbers to be existing and valid at the time of CATI to be defined eligible. However, an alternative classification could assume that these phone numbers were instead *eligible*, but not interviewed (non-contacts 2.20) because they were working the first time they were contacted for the IVR, that is, between 2 and 8 months before CATI. An alternative estimation of call rates is performed under this assumption.

For both IVR and CATI, response, cooperation, refusal and contact rates were computed according to the AAPOR (2016) guidelines, as follows:

Response rate 1:

$$\frac{I}{(I + P) + (R + NC + O) + (UH + UO)} \tag{1}$$

where *I*, *P*, *R* and *UH* were defined as above; *NC* was defined as non-contacts, including cases in which the number was confirmed as an eligible respondent, but the selected respondent was never available or only a telephone answering device was reached (see AAPOR, 2016, categories 2.21 or 2.22); *O* was defined as other cases, such as instances in which there was a respondent who did not refuse the interview, but no interview was obtainable, including death or inabilities (see AAPOR, 2016 categories 2.31–2.36); *UO* was defined as contacts who remained of unknown eligibility, such as failure to complete a needed screener, instances in which a person's eligibility status could not be confirmed or disconfirmed, and other miscellaneous cases in which the eligibility of the number was undetermined and which did not clearly fit into one of the other designations (see AAPOR, 2016, categories 3.2–3.9). Response rate 2 added partial interviews *P* to the numerator; response rate 3 multiplied (*UH*+*UO*) by *e*, defined as the proportion of all callers screened for eligibility who were eligible (see details below); response rate 4 added *P* to the numerator and multiplied (*UH*+*UO*) by *e*.

Cooperation rate 1:

$$\frac{I}{(I + P) + (R + O)} \tag{2}$$

where *I*, *P*, *R* and *O* were defined as above. Cooperation rate 2 added partial interviews *P* to the numerator; cooperation rate 3 did not consider other causes (*O*) among eligible respondents, but not interviewed (see AAPOR, 2016 categories 2.0–2.3); cooperation rate 4 added partial interviews *P* to the numerator and did not consider other causes (*O*).

Refusal rate 1:

$$\frac{R}{(I + P) + (R + NC + O) + (UH + UO)} \tag{3}$$

where *I*, *P*, *R*, *O*, *NC*, *UH* and *UO* were defined as above. Refusal rate 2 multiplied (*UH*+*UO*) by *e*, defined as the proportion of all callers screened for eligibility who were eligible; refusal rate 3 excluded (*UH*+*UO*).

Contact rate 1:

$$\frac{(I + P) + R + O}{(I + P) + (R + NC + O) + (UH + UO)} \tag{4}$$

where *I*, *P*, *R*, *O*, *NC*, *UH* and *UO* were defined as above. Contact rate 2 multiplied (*UH*

+*UO*) by *e*, defined as the proportion of all callers screened for eligibility who were eligible; refusal rate 3 did not consider (*UH+UO*).

It continues to be debated which the best method is to estimate *e*. Most AAPOR methodologies assume that all callers are either eligible (*e* = 100 per cent) or all ineligible (*e* = 0 per cent) to define a range of minimum or maximum response rate, respectively (Smith, 2009). However, assuming that *some* are eligible seems to be a more plausible assumption (Martsolf, Schofield, Johnson, & Scanlon, 2012). I followed Martsolf et al. (2012) and estimated *e* using the most conservative method (AAPOR4), which considered all unknown eligibility refusals (quick hang-ups) to be eligible non-interviews. In practice, *e* was calculated by taking the sum of cases that were considered eligible (*P,I, R*) and dividing by the sum of those cases (*P,I,R*) plus the known ineligible non household cases (*UH*).[10] However, robustness checks were also implemented assuming the extreme cases of *e* = 100 per cent or *e* = 0 per cent.

2.2   Sample Representativeness

To assess the representativeness of the survey sample, the study used data from DHS (Demographic Health Survey Liberia, 2013), most recently collected in 2013, which are a nationally representative sample of household face-to-face interviews; 4118 male and 9239 female respondents between 15 and 49 years old were used as a benchmark for the survey sample (2265 respondents), to compare respondent and household socio-demographic characteristics. A *z*-test for the difference in proportions of respondent and household characteristics was performed between the DHS sample and the (unweighted) survey sample. A *t*-test for the difference in means was implemented for one continuous variable (age), instead. The statistical significance of the differences in proportions and means was estimated at the 5 per cent level.

A second comparison was drawn between DHS and the survey sample, re-weighting the survey sample based on four selected socio-demographic characteristics, which were widely unbalanced across samples: whether the respondent was male, whether she/he had no or primary education, whether she/he owned a mobile phone and whether she/he lived in a rural area.[11] Sixteen strata were constructed, given by the combination of the four socio-demographic characteristics.[12] Conditional probabilities were derived in each stratum, and the survey sample weights were reconstructed as the proportion of respondents in DHS 2013 divided by the proportion of respondents in the survey sample within each stratum, in order to perfectly match the nationally representative distribution from DHS 2013. Similar *z*-test and *t*-test—for the difference in proportions and means, respectively—of respondent and household characteristics were performed between the DHS sample and the *weighted* survey sample.

---

[10]Download AAPOR Response Rate Calculator at https://www.aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Response-Rates-An-Overview.aspx where the formulas are fully described.
[11]See Valliant et al. (2013) and Himelein (2014) for more information on different weighting approaches.
[12]The analysis is limited to 16 strata to avoid having too few observations per stratum, given the total size of the sample of 2265 respondents.

2.3 Costs

The recording of the implementation costs during the study period allows for a comparative analysis between this mobile phone research method and other data collection methods used in past studies. First, a detailed summary of the costs associated with each stage of the data collection, that is, sample and screening of respondents (through IVR) and data gathering (through CATI), was presented. Second, the costs of this novel method were compared with those of data collection implemented through IVR and CATI in past studies (see Section 1 for these studies).

3 RESULTS

3.1 Data Collection

3.1.1 Sampling and screening of respondents (IVR)
Table 1 describes the classification of the mobile phone numbers used, as well as the response, cooperation, refusal and contact rates for the IVR survey in stage 1, which are computed according to the AAPOR (2016) guidelines.

The platform placed 214 823 calls, and the numbers were called on average 3.55 times. The average duration of the IVR survey was of only 1.11 minutes (min 0.41–max 18.91), because the majority of eligible respondents (71 per cent) who were from Montserrado County answered only two questions; 12 761 respondents completed the IVR survey, while 1216 answered only the first two questions out of the three in the survey. Break-offs and refusals, which were similarly defined as answering the first IVR question, that is, whether the respondent resided in Montserrado County or not, were 10 276; 302 phone numbers were always busy and thus classified as of unknown eligibility. The majority of the numbers dialled were classified as ineligible because the phone numbers were not valid as they did not connect with an eligible respondent owing to an error on the provider's end (107 967), and this was due to the automated nature of the RDD calling system; they were temporarily out of service (52 249); they had specific technological issues with connection (31); or they connected, but there was no or invalid selection, such as quick hang-ups (30 021). The proportion of all callers screened for eligibility who were eligible (e) was estimated at 11.30 per cent. However, Table A1 describes alternative response, cooperation, refusal and contact rates, assuming $e = 100$ per cent or $e = 0$ per cent (Smith, 2009).

AAPOR response rates were around 52 and 57 per cent: Response rates 1 and 3 were 51.97 and 52.54 per cent, respectively, while response rates 2 and 4 were slightly higher at 56.92 and 57.55 per cent, respectively. Cooperation rates 1 and 3 were similar at 52.62 per cent, while cooperation rates 2 and 4 were at 57.63 per cent. Refusal rates were at about 42 per cent (41.85, 42.31 and 42.37 per cent, respectively, for refusal rates 1, 2 and 3). Finally, contact rates 1, 2 and 3 were estimated at 98.77, 99.86 and 100 per cent, respectively.

Table A1 confirms similar rates under alternative assumptions of the proportion of all callers screened for eligibility who were eligible (e): Response rates 3 and 4 were still around 52 per cent (52.62 and 51.97 per cent) and 57 per cent (57.63 and 56.92 per cent), assuming $e = 100$ per cent or $e = 0$ per cent, respectively. Estimates were also similar for refusal rate 3 at around 42 per cent (42.37 and 41.85 per cent), and for contact rate 2 at more than 98 per cent (100 and 98.77 per cent). This is because the proportion of *unknown*

Table 1.   Call outcomes and rates for IVR—sampling and screening 0.9

Interview (Category 1)	
Complete (1.1)	12761
Partial (1.2)	1216
Eligible, non-interview (Category 2)	
Break-off/refusals (2.1)	10276
Unknown eligibility, non-interview (Category 3)	
Always busy (3.12)	302
Not eligible (Category 4)	
Unknown if number is valid, call did not connect (4.31)	107967
Temporarily out of service (4.33)	52249
Technological issues (4.4)	31
Other (call connected but no/invalid selection) (4.9)	30021
Total phone numbers used	214823
I = Complete interviews (1.1)	12761
Partial interviews (1.2)	1216
R = Refusal and break-off (2.1)	10276
NC = Non-contact (2.2)	0
O = Other (2.0, 2.3)	0
Calculating e:	
UH = Unknown household (3.1)	302
UO = Unknown other (3.2–3.9)	0
e	11.30%
Response rates	
Response rate 1: $I/(I + P) + (R + NC + O) + (UH + UO)$	51.97%
Response rate 2: $(I + P)/(I + P) + (R + NC + O) + (UH + UO)$	56.92%
Response rate 3: $I/(I + P) + (R + NC + O) + e\,(UH + UO)$	52.54%
Response rate 4: $(I + P)/(I + P) + (R + NC + O) + e\,(UH + UO)$	57.55%
Cooperation rates	
Cooperation rate 1 [and 3]: $I/(I + P) + R + O$	52.62%
Cooperation rate 2 [and 4]: $(I + P)/(I + P) + R + O$	57.63%
Refusal rates	
Refusal rate 1: $R/(I + P) + (R + NC + O) + (UH + UO)$	41.85%
Refusal rate 2: $R/(I + P) + (R + NC + O) + e\,(UH + UO)$	42.31%
Refusal rate 3: $R/(I + P) + (R + NC + O)$	42.37%
Contact rates	
Contact rate 1: $(I + P) + R + O/(I + P) + (R + O + NC) + (UH + UO)$	98.77%
Contact rate 2: $(I + P) + R + O/(I + P) + (R + O + NC) + e\,(UH + UO)$	99.86%
Contact rate 3: $(I + P) + R + O/(I + P) + (R + O + NC)$	100.00%

Notes: This table illustrates call outcomes and rates for the interactive voice response (IVR) survey for sampling and screening, constructed using American Association for Public Opinion Research standards (AAPOR, 2016).

household (*UH*) is small and the proportion of *other* (*O*) is null in the sample. In fact, only 302 phone numbers were of unknown eligibility and counted toward the part of the denominator (*UH+UO*), which was multiplied by *e* in the AAPOR call rates.

Overall, these estimates suggest that a short IVR is a feasible and successful procedure to sample and screen respondents. In fact, in line with another study using the same method (L'Engle et al., 2018), these results established the feasibility of RDD and IVR surveys in another developing country, such as Liberia, and at the time of an epidemic.

3.1.2   Data gathering (CATI)

Table 2 describes a similar exercise for the survey sample interviewed through CATI in stage 2.

Table 2.   Call outcomes and rates for CATI—data gathering

	Round 1	Round 2	Total
Interview (Category 1)			
Complete (1.1)	1957	314	2271
Partial (1.2)	0	0	0
Eligible, non-interview (Category 2)			
Refusals (2.1)	79	34	113
Break-off (2.1)	25	69	94
Unknown eligibility, non-interview (Category 3)			
No screener completed (3.21)	15	0	0
Not eligible (Category 4)			
Less than 18 years old	49	1	50
Call did not connect (4.31)	194	408	602
Temporarily out of service (4.33)	0	634	634
Technological issues (4.4)	0	0	0
Total phone numbers used	2319	1460	3779
I = Complete interviews (1.1)	1957	314	2271
Partial interviews (1.2)	0	0	0
R = Refusal and break-off (2.1)	104	103	207
NC = Non-contact (2.2)	0	0	0
O = Other (2.0, 2.3)	0	0	0
Calculating e:			
UH = Unknown household (3.1)	0	0	0
UO = Unknown other (3.2–3.9)	15	0	15
e	89%	29%	66%
Response rates			
Response rate 1 [and 2]: $(I + P)/(I + P) + (R + NC + O) + (UH + OU)$	94.27%	75.30%	91.10%
Response rate 3 [and 4]: $(I + P)/(I + P) + (R + NC + O) + e\,(UH + OU)$	94.34%	75.30%	91.28%
Cooperation rates			
Cooperation rate 1 [and 3]: $I/(I + P) + R + O$	94.95%	75.30%	91.65%
Cooperation rate 2 [and 4]: $(I + P)/(I + P) + R + O$	94.95%	75.30%	91.65%
Refusal rates			
Refusal rate 1: $R/(I + P) + (R + NC + O) + (UH + UO)$	5.01%	24.70%	8.30%
Refusal rate 2: $R/(I + P) + (R + NC + O) + e\,(UH + UO)$	5.01%	24.70%	8.32%
Refusal rate 3: $R/(I + P) + (R + NC + O)$	5.05%	24.70%	8.35%
Contact rates			
Contact rate 1: $(I + P) + R + O/(I + P) + (R + O + NC) + (UH + UO)$	99.28%	100.00%	99.40%
Contact rate 1: $(I + P) + R + O/(I + P) + (R + O + NC) + e\,(UH + UO)$	99.35%	100.00%	99.60%
Contact rate 3: $(I + P) + R + O/(I + P)\,(R + O + NC)$	100.00%	100.00%	100.00%

Notes: This table illustrates call outcomes and rates for computer-assisted telephone interviewing (CATI) for data gathering, constructed using American Association for Public Opinion Research standards (AAPOR, 2016).

The local NGO, which conducted the CATI survey, was provided with a total of 3779 phone numbers (Table 2) selected for the initial research project (Maffioli, 2020).[13] Out of the 3779 phone numbers, the enumerators called back 2319 respondents in round 1 and 1460 respondents in round 2. In round 1, 1957 respondents completed the interview, while in round 2, owing to the high marginal costs of interviewing additional respondents,

---

[13]The sample for the initial project was supposed to exclude respondents from Montserrado County. The local NGO was then provided with 2733 phone numbers from complete responses and 1276 phone numbers from incomplete responses, for those individuals who reported in the IVR survey (stage 1) that they did not live in Montserrado County at the beginning of the Ebola epidemic. After the respondents were interviewed through CATI (stage 2), 14% of them confirmed that they lived in Montserrado County at the end of 2013, despite having answered differently in the IVR survey. These individuals were kept in the final sample.

the NGO stopped at 314 individuals interviewed. In fact, because phone number prefixes are associated with different phone companies and each phone company allows taking advantage of different call, text or data promotions, it is common for Liberians to switch between phone companies and thus frequently change phone numbers. Between 2 and 8 months after the selection and screening process (stage 1), the NGO found that, of all the numbers provided for round 2 (1460 phone numbers), 43 per cent (634) of the numbers were permanently switched off and 28 per cent (408) were not ringing.

The final sample that was eligible for the initial research project (Maffioli, 2020) and consented to be interviewed through CATI included 2271 individuals (1957 from round 1 and 314 from round 2, Table 2) across the entirety of Liberia: These respondents were the ones completing the interview. In addition, 113 respondents refused to be interviewed; 94 were phone numbers that were dialled, for which the phone rang but the respondent did not pick up the call, and they were defined as break-offs; 15 phone numbers were defined as ineligible as no screening was completed: These respondents reported that they have been already interviewed. Finally, phone numbers were classified as non-eligible: (i) if respondents were younger than 18 years (50); (ii) phone numbers never responded because the call never rang on a person's phone (602): these phone numbers were categorized as unknown if the number is valid, because the call did not connect; and (iii) were temporarily out of service (634). However, Table A2 describes alternative response, cooperation, refusal and contact rates, defining the 602 phone numbers as non-contacts (2.20) because these phone numbers were valid and contact was established during the IVR survey.

Response rate 1 (and 2), cooperation rate 1 (and 3), refusal rate 1 and contact rate 1 were 91.10, 91.65, 8.30 and 99.40 per cent, respectively (Table 2). The implementation of the CATI survey was more successful in round 1 compared with round 2, suggesting that a longer waiting time (in round 2) between stage 1 and stage 2 led to a higher number of non-eligible phone numbers as well as a higher proportion of refusals and break-offs. Call rates under the alternative classification of 602 phone numbers as non-contacts are slightly different: Response rate 1 (and 2) is lower at 73.38 per cent and contact rate 1 is lower at 80.06 per cent. However, refusal rate 1 is lower at 6.69 per cent, while cooperation rate 1 (and 3) is the same at 91.65 per cent. This is expected since non-contacts (2.2) count toward the denominator of the AAPOR call rates.

Overall, these estimates suggest that a 30- to 45-minute CATI survey with 100 questions was feasible in a developing country and at the time of an epidemic. They also shed light on how sensitive information on individuals' experience with Ebola or their political views could be asked through CATI without compromising the response rate.

3.2   Sample Representativeness

To understand the representativeness of the survey sample, the analysis is conducted on 2265 respondents out of the initial 2271, because for six of them, their reported location did not match up with the list of villages provided by the Liberian Institute of Statistics and Geo-Information Services.

The comparison in proportions of respondent and household characteristics between DHS and the survey sample yielded statistically significant differences at 5 per cent level, with the exception of one variable (whether the household has pigs). Table 3 describes a survey sample biased toward male, educated respondents from urban areas and with access to mobile phones (column 2 versus 1). Survey respondents were also on average wealthier

Table 3. Socio-demographic characteristics of national sample from DHS 2013 and survey sample 2016

	(1)	(2)	(3)	(4)	(5)
	DHS 2013	Survey Sample (unweighted)		Survey Sample (weighted)	
	[Obs = 13 357]	[Obs = 2265]		[Obs = 2265]	
	Mean	Mean	(1)–(2)	Mean	(1)–(4)
Resp male*	31	65	−34	31	0
Resp educ none or primary*	66	17	49	66	0
Has mobile phone*	60	96	−36	60	0
Rural*	60	30	30	60	0
Resp age	29.3	32.6	−3.3	34.3	−5
Married	62	49	13	50	12
Resp Christian	84	86	−2	80	4
Resp not working	36	25	11	13	23
Has electricity	6	36	−30	30	−24
Has radio	59	80	−21	74	−15
Has bank account	14	22	−8	7.9	6.1
Has refrigerator	3.8	6.4	−2.6	2.4	1.4
Has vehicle	13	17	−4	6.9	6.1
Improved toilet (WHO)	37	87	−50	81	−44
Improved wall material (WHO)	27	68	−41	57	−30
Improved roof material (WHO)	68	97	−29	96	−28
Has chickens	42	48	−6	56	−14
Has goats/sheep	12	8.3	3.7	10	2
Has pigs	3.2	2.7	0.5	1.7	1.5
Has cows	0.82	0.44	0.38	0.14	0.68
Resp hh size	6.62	6.88	−0.26	7.18	−0.56
Montserrado	16	14	2	15	1

Notes: This table illustrates the comparisons of proportions and means in socio-demographic characteristics between the national sample from Demographic Health Survey (DHS, 2013) and survey sample (2016). Column 1 reports summary statistics from DHS 2013; column 2 reports summary statistics from the survey sample (unweighted); column 4 reports summary statistics from the survey sample, weighted by selected socio-demographics characteristics (*). Columns 3 and 5 report differences in the proportions and means between the samples.

as defined by several measures of asset ownership and improved sources of toilet, wall and roof material.[14]

The fact that the survey sample is not representative of the national Liberian population is not surprising, because both stages of the data collection were conducted through mobile phones, and individuals needed to have access to a mobile phone at the time of the call: Male, more educated and wealthier individuals from urban areas are more likely to own mobile phones (Demographic Health Survey Liberia, 2013). Furthermore, stage 1 was set up to limit respondents from Montserrado County, the most economically developed

[14]Indicators for improved versus unimproved assets were reconstructed following WHO standards. These indicators were then used to implement a principal component analysis to construct a wealth index, in a similar way to the DHS wealth index. See https://dhsprogram.com/topics/wealth-index/Wealth-Index-Construction.cfm

and urban county. In addition, a selection based on the county and district where the respondent resided at the beginning of the epidemic was imposed to define the final sample frame (3779 phone numbers, Table 2), which the local NGO called back in stage 2. Table 3 confirms that the final sample of 2265 respondents is very different from a nationally representative survey.

Table 3 column 4 shows the mean estimates from the CATI survey weighted by four selected characteristics (whether the respondent is a male, whether she/he has no or primary education, whether she/he owns a mobile phone and whether she/he lives in rural areas). By construction, the weighted survey sample is identical to DHS 2013 in the four dimensions selected. After weighting, for 10 out of the 18 variables considered, the difference in proportions or means between the weighted survey sample and the DHS sample is reduced (Table 3, columns 3 versus 5). However, even comparing the DHS sample with the weighted survey sample, the proportions or means reported in Table 3 remain statistically significantly different from each other at 5 per cent level, suggesting that weighting did not entirely solve the bias.[15][16]

It is important to highlight that the IVR survey was not set up to target a nationally representative sample, by imposing quotas of respondents with certain geographical or socio-demographic characteristics. Thus, it should not be surprising that the survey sample of mobile phone owners is not representative of the country's population.

3.3   Costs

Table 4 presents a summary of the costs. The costs for the sampling and screening of respondents through IVR in stage 1 include the fixed initial cost of consulting for the use and maintenance of the platform, piloting costs and airtime. The cost per each picked-up call in stage 1 was only $0.10. The cost of the IVR survey was higher ($1.49) for each survey, considering both complete and partial, and even higher ($1.63), considering only complete surveys.

Regarding the CATI survey used to gather data in stage 2, the costs depend on the country in which researchers work as well as constraints due to the emergency situation, lack of electricity to charge phones or lack of money to buy airtime. In this study, the costs included both common data collection costs and additional costs due to the high risk of the epidemic: enumerators' monthly salaries, human resources costs for survey programming, testing and revisions; other data cleaning costs; internet, fuel for electricity generators, mobile phone airtime for enumerators and survey respondents (gift of $1 airtime); vehicle maintenance to bring enumerators to the office during the Ebola epidemic; and security and Ebola safety measures. Working with the local NGO partner resulted in a cost of about $13.49 per respondent they tried to reach and $22.45 per complete survey. Completing a

---

[15]It is worthy to highlight, however, that despite the variables compared in Table 3 being constructed in similar ways, the questions in the DHS and in the survey were asked differently. As a result, the construction of similar but not identical variables such as occupation or improved assets might contribute to explaining some of these differences.

[16]The survey in the initial research project was conducted to collect political outcomes. Another potential comparison would be to test the difference between political outcomes in survey sample and Afrobarometer data (2015). Unfortunately, trust and perceived corruption questions were asked so differently across surveys that it is hard to homogenize the outcomes. Future studies should take this into account, and they should consider asking questions in a format similar to existing national representative surveys to be able to weight the sample on the basis of socio-demographic characteristics and compare the outcomes of interest.

Table 4. Costs, by stage and survey type

	Type unit	No. units	Cost/unit	Total cost
1. RDD and IVR—sampling and screening				
Fix cost				
Platform access/support	Unit	1	$500.00	$500.00
Pilot				
Consulting services	Days	5	$400.00	$2000.00
Airtime	Pilot	1	$500.00	$500.00
Data collection				
Complete surveys	Interview	4000	$3.15	$12 600.00
Call back additional 3 times[a]				$3995.00
Consulting services	Days	3	$400.00	$1200.00
Total costs				$20 795.00
Cost/call	Phone number	214 823		$0.10
Cost/complete and partial survey	Phone number	13 977		$1.49
Cost/complete survey	Phone number	12 761		$1.63
2. CATI—data gathering				
Total costs				$50 980.73
Round 1 + 2				
Cost/call		3779		$13.49
Cost/complete survey		2271		$22.45
Round 1				
Cost/call		2319		$21.98
Cost/complete survey		1957		$26.05
Round 2				
Cost/call		1460		$34.92
Cost/complete survey		314		$162.36

Notes: This table illustrates the costs for stage 1: the sampling and screening process through the random-digit dialling (RDD) and interactive voice response (IVR); and stage 2: data gathering through computer-assisted telephone interviewing (CATI), and by round of data collection.
[a]The cost of $3395 to call back each phone number up to three additional times was estimated assuming that 150 000 calls are made, and (i) 15% of them would be picked up a first time; thus, calling back 85% of phone numbers a second time would cost $1275 (127 500 calls × $0.01 per call); (ii) 12.5% of them would be picked up a second time; thus, calling back 87.5% of phone numbers a third time would cost $1116 (111 562 calls × $0.01 per call); (iii) 10% of them would be picked up a third time; thus, calling back 90% of phone numbers a fourth time would cost $1004 (100 406 calls × $0.01 per call). The assumptions were made by VotoMobile on the basis of their past experience. The total costs sum-up to $3995 (1275 + 1116 + 1004 = $3995).

CATI survey 8 months after the IVR survey (round 2) costs up to six times more than gathering the data between 2 and 4 months after (round 1) ($162.36 versus $26.05). In summary, the total cost of this novel method per complete survey (including both stage 1, sampling and screening; and stage 2, data gathering) is around $24.

Several studies estimated the costs of collecting data through different methods. More expensive data collection methods are face-to-face surveys, which cost at least $25 but can reach values as high as $150, depending on the complexity of the survey and the distances that have to be covered to find respondents. For example, Lietz et al. (2015) estimated a cost of $25 per survey in Burkina Faso; Mahfoud et al. (2015), $36 per survey in Lebanon; Ballivian et al. (2015), $40 per survey in Peru and Honduras; Hoogeveen et al. (2014) and Dillon (2012), between $50–150 and $97 per survey, respectively, in Tanzania; and Dabalen et al. (2016), $150 per survey in Malawi.

On the other hand, costs for IVR and CATI surveys have been estimated to be much lower than those of face-to-face surveys in a similar setting (Schuster & Brito, 2011;

Mahfoud et al., 2015; Garlick et al., 2019; Lau et al., 2019). CATI interviews cost between $4.10 and $7.30 per survey in Tanzania (Hoogeveen et al., 2014; Dillon, 2012), $5.80–8.80 per survey in Malawi (Dabalen et al., 2016) and between $4.44 and $22.20 in Lebanon (Mahfoud et al., 2015). Similarly, lower costs have been estimated for IVR surveys, such as $17 in Ballivian et al. (2015), $4.95 in L'Engle et al. (2018) and about $2 in Leo et al. (2015), depending on the length of the survey and the criteria applied to select the sample. Compared with face-to-face data collection, this two-stage mobile phone research method is then advantageous by eliminating many of the implementation costs associated with in-field sampling and screening (in stage 1), and face-to-face surveys (in stage 2), such as personnel, logistics and distribution of phones. Compared with a single IVR or CATI survey, this method might be more expensive. However, both IVR and CATI surveys require a list of phone numbers to start with. If this initial list of respondents is not available and selecting a sample through in-person interviews is too costly or risky in challenging settings or during emergencies, then this method might still be a cost-effective solution. In fact, adding the costs of a baseline face-to-face data collection to gather the initial list of respondents to interview (stage 1) to the costs of IVR or CATI surveys (stage 2) to gather data, the combined costs would be as high as $160, compared with $24 for the method proposed in this study. It is also important to notice that the costs at each stage ($1.63 per IVR survey; $22.45 per CATI survey, Table 4) are similar or lower than the estimated costs reported in other studies.

This does not indicate that this method is superior to others; rather it provides evidence that this two-stage mobile phone research method can be affordably implemented in challenging settings where in-person data baseline collection is prohibitively costly or dangerous.

4   DISCUSSION

The two-stage data collection method proposed in this study uniquely samples and screens respondents and gathers data without any in-person interaction with respondents, in a developing country and at the time of an epidemic. The method allowed researchers to conduct more than 2200 interviews in Liberia during the 2014 Ebola epidemic, with an average estimated cost of $24 per survey and with call rates comparable with those of similar survey research methods.

By combining established data collection methods (RDD, IVR and CATI) in a novel manner, the main strength is that, relying solely on mobile phone technology, this precludes the need for prior data or fieldwork activities to have a sampling frame, and it allows researchers to gather survey data in challenging settings. The sampling and screening of respondents through IVR (stage 1) could be useful in any country with some phone access (Liberia has on average 65 per cent of phone coverage, Table A3), and in total absence of any initial list of respondents. The data collection through CATI (stage 2) documents how sensitive information can be asked in a 30- to 45-minute phone interview without compromising the response rate. Altogether, the survey data highlight how this innovative data collection method was feasible and cost-effective in a developing country and at the time of an epidemic.

There are several meaningful lessons learned, which are important to discuss for the use of this method in future research.

First, the study found that the group of respondents was not representative of the general population of the country. In fact, in Table 3, the survey sample was compared with a nationally representative survey (Demographic Health Survey Liberia, 2013), and it was shown to be biased. Even after re-weighting on a few selected socio-demographic variables, most of the differences remained statistically significant. This is not surprising because the selection and screening were conducted through mobile phones and respondents needed to have access to a mobile phone in order to pick up the call. This is also in line with other research in developed (Lee, Brick, Brown, & Grant, 2010) and developing countries (Leo et al., 2015; Lau et al., 2019) where respondents interviewed through telephone surveys are different from the entire population, even after controlling for demographic characteristics. This method, then, does not appear to be the best approach if researchers are interested in a nationally representative sample, unless fixed quotas are used in the IVR survey to select respondents based on pre-defined geographical and socio-demographic characteristics, in order to reproduce a nationally representative sample.

Reaching sample representativeness is achievable, but it would come at the expense of higher implementation costs in the IVR message, because several screening questions would need to be added to the survey tool. As the number of questions asked through the IVR survey increases along with sample specificity (e.g. half women and half men, or a fixed number of respondents per geographical area), so do the difficulty and costs associated with finding the targeted nationally representative sample.[17]

The principal remaining advantage of gathering data using this innovative method (but from a non-nationally representative sample) is to collect high-quality data at time of emergency when in-person interactions are impossible and no publicly available data exist to select a nationally representative sample. Several research questions are and could be answered by focusing on a specific sample of respondents selected with a small set of IVR quotas.

Second, another important lesson learned from this study to consider for future research relates to the attrition that researchers could face from the initial list of phone numbers generated by the RDD to the completion of the CATI survey.

World Bank researchers who performed phone surveys during the Ebola epidemic in Liberia reported that only 30 per cent of the initial sample completed the survey (16 per cent of the original sample, The World Bank Group, 2014). In Sierra Leone, the response rate was also lower than expected, given the nature of the survey and the difficult conditions under which it was conducted: About 69 per cent of the sample respondents with phone numbers completed the survey (45 per cent of the original sample, The World Bank Group, 2015). Other surveys, performed through face-to-face interviews in Montserrado County, reached 95 per cent of the respondents (Blair, Morse, & Tsai, 2016). Follow-up phone surveys reached about 80 per cent of the original sample. The initial in-person interaction between the field enumerators and respondents during the baseline survey seemed to have been the main factor determining the lower attrition rate. Because this project was conducted starting in late 2015, when the epidemic was not at the peak and life was going back to normal, the response rate was expected to be similar or higher than it

---

[17]In this study, no location-specific quotas were established. Alternatively, one question in the IVR survey was intended to limit respondents from one county (Montserrado), where more than 30 per cent of the population live (Table A3). Although a fixed quota of respondents per county could has been used, the costs were too high given the budget available. Researchers should evaluate the costs of using quotas depending on the sample they want to select for their research.

was in the World Bank surveys and were indeed computed at 51.97 per cent for the IVR survey (Table 2), and 91.10 per cent for the CATI survey (Table 3).

Similarly, participation for confidentiality reasons was a related concern. First, although individuals never provided their phone numbers to enumerators, there was concern that respondents, once called back from the local NGO, would ask enumerators where they accessed their phone number from and refuse to participate. In Liberia, this was not a problem because people are used to receiving calls for advertisements or polls: The refusal rate of the CATI survey was in fact 8.30 per cent across the two rounds of data collection (Table 3).

Second, because of the complete lack of in-person interaction at both stages of the data collection, there was a concern that respondents would not feel at ease to provide opinions to a stranger.[18] There was also uncertainty that, owing to the personal nature of the questions asked in this study (e.g. about their experience with Ebola or their political views), respondents would be reluctant or they would refuse to stay on the phone for a long time (the survey lasted 30–45 minutes with 100 questions). Rather, only 113 respondents refused to participate (Table 2), and once enumerators established a first phone contact and call at an appropriate time, individuals completed the CATI survey.

Third, respondents were provided with a $1 airtime incentive that was directly transferred upon the completion of the survey.[19] Respondents were informed at the time of the sampling, through the IVR survey, about this direct benefit. Recent phone surveys in developing countries found small differences in attrition when incentives were randomly varied (Gallup, 2012; Demombynes et al., 2013; Hoogeveen et al., 2014; Leo et al., 2015); therefore, the incentive may have contributed to a lower non-response rate.

Finally, this method does not solve the problem of people sharing a phone number. For both stage 1 and stage 2, whoever answered the call was interviewed. It is then possible that whoever picked up the call during stage 1 would not be the same person in stage 2. Even if that was the case, stage 1 only collected the geographical location of the respondent, while the survey data were collected in stage 2. During the CATI surveys, the respondent was asked about her/his location again. If different from what was reported during the IVR, the respondent was confronted about it and was asked to confirm the correct location at the beginning of the epidemic. In about 16 per cent of the cases, individuals reported a different location at the two stages. Qualitatively, the majority of respondents said that they had problems with the speed of the IVR message and how they inputted their answers into the keyboard. For the analysis, the location data collected through CATI interviews were trusted more than the data collected through IVR. However, this problem should be taken into account and addressed in future uses of this method.

A last important lesson learned relates to the time waited between the sampling and screening through IVR (stage 1) and the data collection through CATI (stage 2). For budgetary reasons, a mobile phone data collection (round 2) was added at a later date, and this caused between 2 and 8 months of delay between stage 1 and stage 2. In round 2, the local NGO found that 43 per cent (634) of the numbers provided (1475) were

---

[18]There is a debate on whether face-to-face or mobile phone surveys allow respondents more confidentiality in their responses. Those in favour of mobile phone surveys argue that because mobile phone surveys do not allow enumerators to read the body language and facial expressions of the respondent, they might ensure more confidentiality. Respondents might feel more at ease providing opinions to a stranger by phone, rather than through personal interactions in front of an enumerator.

[19]See Brick et al. (2007) and Oldendick and Lambries (2013) for evidence on the effect of incentives on response rates, and see Singer and Ye (2013) for a review.

permanently switched off and 28 per cent (408) were not ringing (Table 3). While Liberians have the habit of owning multiple SIM cards and switch between them on the basis of cost advantages for airtime, text messages and internet data, it is not clear whether this would happen in other countries. Still, it is advised that researchers limit the time window between the two stages of data collection as much as possible to reduce potential additional problems of not finding (at least temporarily) working phone numbers.[20] Despite this caveat, respondents were willing to answer long (30–45 minutes, 100 questions) CATI surveys with low refusal and break-off rates (Table 3). This is a good indication that CATI surveys are feasible in challenging settings such as at the time of an epidemic in a developing country.

5   CONCLUSION

This study proposes and describes a novel two-stage data collection method which combines (i) RDD of phone numbers and an IVR survey to conduct sampling and screening and (ii) data collection through CATI. This procedure was used to conduct more than 2200 interviews in the country of Liberia, at the time of the 2014 Ebola epidemic. Following the AAPOR (2016) guidelines, response, cooperation, refusal and contact rates were computed at 51.97, 52.62, 41.85 and 98.77 per cent, respectively, for the IVR survey. The CATI survey was more successful: Response, cooperation, refusal and contact rates were 91.10, 91.65, 8.30 and 99.40 per cent, respectively.

Unsurprisingly, because both stages of the data collection method were conducted by mobile phones and no quotes were set up to reach a nationally representative sample, male, educated respondents from urban areas and with access to mobile phones were more likely to be interviewed. A comparison between the survey sample and the nationally representative sample from the DHS indeed confirmed statistically significant differences in socio-demographic characteristics. The re-weighting of the sample on a few selected covariates reduced the gaps, but the differences remained statistically significant. Yet in an analysis of the costs compared with other past research approaches used, the study found that this method offers promise for data collection in developing countries at a low cost ($24 per survey), especially in challenging settings.

The results on call rates, sample representativeness and costs are in line with other studies that use RDD, IVR or CATI to collect data in developing countries. First, the closest studies to mine in terms of the methods used to gather data in other developing countries (L'Engle et al., 2018; Lau et al., 2019) found worse call rates. Second, similar to what I found, some research—in both developed (Lee et al., 2010) and developing countries (Leo et al., 2015; Lau et al., 2019)—found that respondents interviewed through CATI were different from the entire population, even after controlling for demographic characteristics. Third, the costs of this innovative method were lower than those of face-to-face surveys (Dillon, 2012; Hoogeveen et al., 2014; Ballivian et al., 2015; Lietz et al., 2015; Mahfoud et al., 2015; Dabalen et al., 2016) and within the range of studies using IVR or CATI (Schuster & Brito, 2011; Dillon, 2012; Ballivian et al., 2015; Leo

---

[20]The local NGO, Parley, was instructed to call back the phone numbers not ringing or switched off a few times, but the callbacks were concentrated over a few days. Potentially, more respondents could have been reached with multiple calls at later dates or longer intervals if the phones were temporarily not working.

et al., 2015; Mahfoud et al., 2015; Dabalen et al., 2016; L'Engle et al., 2018; Garlick et al., 2019; Lau et al., 2019), suggesting that this method is feasible and affordable.

I suggest that researchers weigh the advantages and limitations of this approach before implementing it in any specific country or context. Still, the proposed method remains the first to only rely on mobile phone technology for all stages of data collection. As the utilization of mobile phones for data collection and research is increasing in developing countries, this innovative two-stage procedure improves the ability of researchers to gather information in emergencies in developing economies, making it a unique and feasible approach to implement in challenging settings.

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

Data available on request from the author.

REFERENCES

AAPOR. 2016. AAPOR 1. The American Association for Public Opinion Research. Standard definitions: final dispositions of case codes and outcome rates for surveys 9th edition. Technical report

Ballivian A, Azevedo JP, Durbin W. 2015. Using mobile phones for high-frequency data collection. In *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies*, Toninelli D, Pinter R, de Pedraza P (eds). Ubiquity Press: London; 21–39. http://10.0.20.214/bar. c.license:CC-BY4.0.

Bauer J-M, Akakpo K, Enlund M, Passeri S. 2013. A new tool in the toolbox: using mobile text for food security surveys in a conflict setting. http://www.odihpn.org/the-humanitarian-space/news/ announcements/blog-articles/a-new-tool-in-the-toolbox-using-mobile-text-for-food-security-surveys-in-a-conflict-setting

Blair R, Morse B, Tsai L. 2016. Public health and public trust: survey evidence from the Ebola virus disease epidemic in Liberia. *Social Science and Medicine* **172**: 89–97.

Brick JM, Brick PD, Dipko S, Presser S, Tucker C, Yuan Y. 2007. Cell phone survey feasibility in the U.S.: sampling and calling cell numbers versus landline numbers. *Public Opinion Quarterly* **71** (1): 23–29.

Dabalen A, Etang A, Hoogeveen J, Mushi E, Schipper Y, von Engelhard J. 2016. *Mobile Phone Panel Surveys in Developing Countries: A Practical Guide for Microdata Collection*. World Bank: Washington, D.C.

Demographic Health Survey Liberia. 2013. Technical report.

Demombynes G, Gubbins P, Romeo A. 2013. Challenges and opportunities of mobile phone-based data collection: evidence from South Sudan. The World Bank, Africa region, poverty reduction and economic management unit, policy research working paper 6321

Dillon B. 2012. Field report: using mobile phones to collect panel data in developing countries. *Journal of International Development* **24**: 518–527.

Gallup. 2012. The World Bank listening to LAC (L2L) pilot: final report. Technical report

Ganesan M, Prashant S, Jhunjhunwala A. 2012. A review on challenges in implementing mobile phone based data collection in developing countries. *Journal of Health Informatics in Developing Countries* **6**(1): 366–374.

Garlick R, Orkin K, Quinn S. 2019. Call me maybe: experimental evidence on using mobile phones to survey African microenterprises. Forthcoming at the *World Bank Economic Review*

Gibson DG, Pereira A, Farrenkopf BA, Labrique AB, Pariyo GW, Hyder AA. 2017. Mobile phone surveys for collecting population-level estimates in low- and middle-income countries: a literature review. *Journal of Medical Internet Research* **19**(5): e139.

Gonzalez R, Maffioli EM. 2020. Is the phone mightier than the virus? Cell phone access and epidemic containment efforts. Working paper. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3548926

Grosh M, Glewwe P. 2000. *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study.* World Bank: Washington, D.C.

Himelein K. 2014. Weight calculations for panel surveys with subsampling and split-off tracking. *Statistics and Public Policy* **1**(1): 40–45.

Hoogeveen J, Croke K, Dabalen A, Demombynes G, Giugale M. 2014. Collecting high frequency panel data in Africa using mobile phone interviews. *Canadian Journal of Development Studies* **35**(1): 186–207.

Hughes SM, Haddaway S, Zhou H. 2016. Comparing smartphones to tablets for face-to-face interviewing in Kenya. *Survey Methods: Insights from the Field* **12**.

Kempf AM, Remington PL. 2007. New challenges for telephone survey research in the twenty-first century. *Annual Review of Public Health* **28**: 113–126.

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**: 305–318.

Lee S, Brick JM, Brown ER, Grant D. 2010. Growing cell-phone population and noncoverage bias in traditional random digit dial telephone health surveys. *Health Services Research* **45**(4): 1121–1139.

L'Engle K, Sefa E, Adimazoya EA, Yartey E, Lenzi R, Tarpo C, Heward-Mills NL, Lew K, Ampeh Y. 2018. Survey research with a random digit dial national mobile phone sample in Ghana: methods and sample quality. *PLoS ONE* **13**(1): e0190902.

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

Liberian Telecommunications Authority. 2012. 2012 annual report. Technical report.

Lietz H, Lingani M, Sie A, Sauerborn R, Souares A, Tozan Y. 2015. Measuring population health: costs of alternative survey approaches in the Nouna health and demographic surveillance system in rural Burkina Faso. *Global Health Action* **8**: 28330.

Maffioli EM. 2020. The political economy of health epidemics: evidence from the Ebola outbreak. Working paper. Available at SSRN: https://ssrn.com/abstract=3383187

Mahfoud Z, Ghandour L, Ghandour B, Mokdad AH, Sibai AM. 2015. Cell phone and face-to-face interview responses in population-based surveys: how do they compare? *Field Methods* **27**(1): 39–54.

Martsolf GR, Schofield RE, Johnson DR, Scanlon DP. 2012. Editors and researchers beware: calculating response rates in random digit dial health surveys. *Health Research and Educational Trust* **48**(2pt1): 665–676.

Massey JT, O'Connor D, Krotki K. 1997. Response rates in random digit dialing telephone surveys. Proceedings of the Section on Survey Research Methods, American Statistical Association; 707–712.

Oldendick RW, Lambries DN. 2013. Incentives for cell phone only users: what difference do they make? *Survey Practice* **4**(1):

Schuster C, Brito CP. 2011. Cutting costs, boosting quality and collecting data real-time: lessons from a cell phone-based beneficiary survey to strengthen Guatemala's conditional cash transfer program. Technical report. Available at http://siteresources.worldbank.org/INTLAC/Resources/257803-1269390034020/EnBreve_166_Web.pdf

Singer E, Ye C. 2013. The use and effects of incentives in surveys. *The Annals of the American Academy of Political and Social Science* **645**(1): 112–141.

Smith TW. 2009. A revised review of methods to estimate the status of cases with unknown eligibility.

The World Bank Group. 2014. The socio-economic impacts of Ebola in Liberia: results from a high frequency cell phone survey.

The World Bank Group. 2015. The socio-economic impacts of Ebola in Sierra Leone: results from a high frequency cell phone survey, round 1.

Tomlinson M, Solomon W, Singh Y, Doherty T, Chopra M, Ijumba P, Tsai AC, Jackson D. 2009. The use of mobile phones as a data collection tool: a report from a household survey in South Africa. *BMC Medical Informatics and Decision Making* **9**(51): 1–8.

Toninelli D, Pinter R, de Pedraza P. 2015. *Mobile research methods: Opportunities and challenges of mobile research methodologies*. Ubiquity Press: London.

Twaweza East Africa. 2013. Sauti za Wananchi: collecting national data using mobile phones. Twaweza, Dar es Salaam, Tanzania

United Nations Office for Disaster Risk Reduction, UNISDR. 2015. The human cost of weather-related disasters: 1995–2015. Technical report

Valliant R, Dever JA, Kreuter F. 2013. Practical tools for designing and weighting survey samples.

Waksberg J. 1978. Sampling methods for random digit dialing. *Journal of the American Statistical Association* **73**(361): 40–46.

van der Windt P, Humphreys M. 2014. Crowdseeding conflict data. Working paper

World Bank. 2016. World development report 2016: digital dividends. Technical report. http://www.worldbank.org/en/publication/wdr2016

APPENDIX A.

Table A1.   Robustness: call outcomes and rates for IVR—sampling and screening

Interview (Category 1)		
Complete (1.1)	12 761	12 761
Partial (1.2)	1216	1216
Eligible, non-interview (Category 2)		
Break-off/refusals (2.1)	10 276	10 276
Unknown eligibility, non-interview (Category 3)		
Always busy (3.12)	302	302
Not eligible (Category 4)		
Unknown if number is valid, call did not connect (4.31)	107 967	107 967
Temporarily out of service (4.33)	52 249	52 249
Technological issues (4.4)	31	31
Other (call connected but no/invalid selection) (4.9)	30 021	30 021
Total phone numbers used	214 823	214 823
I = Complete interviews (1.1)	12 761	12 761
Partial interviews (1.2)	1216	1216
R = Refusal and break-off (2.1)	10 276	10 276
NC = Non-contact (2.2)	0	0
O = Other (2.0, 2.3)	0	0
Calculating e:		
UH = Unknown household (3.1)	302	302
UO = Unknown other (3.2–3.9)	0	0
e	0.00%	100.00%
Response rates		
Response rate 1: $I/(I + P) + (R + NC + O) + (UH + OU)$	51.97%	51.97%
Response rate 2: $(I + P)/(I + P) + (R + NC + O) + (UH + OU)$	56.92%	56.92%
Response rate 3: $I/(I + P) + (R + NC + O) + e\ (UH + OU)$	52.62%	51.97%
Response rate 4: $(I + P)/(I + P) + (R + NC + O) + e\ (UH + OU)$	57.63%	56.92%
Cooperation rates		
Cooperation rate 1 [and 3]: $I/(I + P) + R + O$	52.62%	52.62%
Cooperation rate 2 [and 4]: $(I + P)/(I + P) + R + O$	57.63%	57.63%
Refusal rates		
Refusal rate 1: $R/(I + P) + (R + NC + O) + (UH + UO)$	41.85%	41.85%
Refusal rate 2: $R/(I + P) + (R + NC + O) + e\ (UH + UO)$	42.37%	41.85%
Refusal rate 3: $R/(I + P) + (R + NC + O)$	42.37%	42.37%
Contact rates		
Contact rate 1: $(I + P) + R + O/(I + P) + (R + O + NC) + (UH + UO)$	98.77%	98.77%
Contact rate 2: $(I + P) + R + O/(I + P) + (R + O + NC) + e\ (UH + UO)$	100.00%	98.77%
Contact rate 3: $(I + P) + R + O/(I + P) + (R + O + NC)$	100.00%	100.00%

Notes: This table illustrates call outcomes and rates for the interactive voice response (IVR) survey for sampling and screening, constructed using American Association for Public Opinion Research standards (AAPOR, 2016). The table shows alternative rates, assuming e equals to 0% or 100%.

Table A2.   Robustness: call outcomes and rates for CATI—data gathering

	Round 1	Round 2	Total
Interview (*Category 1*)			
Complete (1.1)	1957	314	2271
Partial (1.2)	0	0	0
Eligible, non-interview (*Category 2*)			
Refusals (2.1)	79	34	113
Break-off (2.1)	25	69	94
Non-contacts (2.20)	194	408	602
Unknown eligibility, non-interview (*Category 3*)			
No screener completed (3.21)	15	0	0
Not eligible (*Category 4*)			
Less than 18 years old	49	1	50
Temporarily out of service (4.33)	0	634	634
Technological issues (4.4)	0	0	0
Total phone numbers used	2319	1460	3779
I = Complete interviews (1.1)	1957	314	2271
Partial interviews (1.2)	0	0	0
R = Refusal and break-off (2.1)	104	103	207
NC = Non-contact (2.2)	194	408	602
O = Other (2.0, 2.3)	0	0	0
Calculating e:			
UH = Unknown household (3.1)	0	0	0
UO = Unknown other (3.2–3.9)	15	0	15
e	98%	57%	82%
Response rates			
Response rate 1 [and 2]: $(I + P)/(I + P) + (R + NC + O) + (UH + OU)$	86.21%	38.06%	73.38%
Response rate 3 [and 4]: $(I + P)/(I + P) + (R + NC + O) + e (UH + OU)$	86.22%	38.06%	73.44%
Cooperation rates			
Cooperation rate 1 [and 3]: $I/(I + P) + R + O$	94.95%	75.30%	91.65%
Cooperation rate 2 [and 4]: $(I + P)/(I + P) + R + O$	94.95%	75.30%	91.65%
Refusal rates			
Refusal rate 1: $R/(I + P) + (R + NC + O) + (UH + UO)$	4.58%	12.48%	6.69%
Refusal rate 2: $R/(I + P) + (R + NC + O) + e (UH + UO)$	4.58%	12.48%	6.69%
Refusal rate 3: $R/(I + P) + (R + NC + O)$	4.61%	12.48%	6.72%
Contact rates			
Contact rate 1: $(I + P) + R + O/(I + P) + (R + O + NC) + (UH + UO)$	90.79%	50.55%	80.06%
Contact rate 1: $(I + P) + R + O/(I + P) + (R + O + NC) + e (UH + UO)$	90.81%	50.55%	80.14%
Contact rate 3: $(I + P) + R + O/(I + P) (R + O + NC)$	91.40%	50.55%	80.45%

Notes: This table illustrates call outcomes and rates for computer-assisted telephone interviewing (CATI) for data gathering, constructed using American Association for Public Opinion Research standards (AAPOR, 2016). The table shows alternative rates, defining 602 phone numbers as eligible non-contacts (2.20) instead of non-eligible (4.31).

Table A3.   Summary statistics, by county 0.8

County	Pop (2014)	Pop (%) (2014)	Pop density (pop/sq mile) (2014)	Rural (%) (2012)	Phone coverage (%) (2012)	No. resp (2015)	No. resp (%) (2015)
Bomi	94418	2.41	127	79.94	91.37	75	3.30
Bong	401500	10.23	119	70.30	86.22	431	18.98
Gbarpolu	93598	2.38	24	88.71	40.00	20	0.88
Grand Bassa	251938	6.42	84	74.11	55.90	169	7.44
Grand Cape Mount	142304	3.62	77	91.96	72.29	59	2.60
Grand Gedeh	140594	3.58	34	61.83	58.13	60	2.64
Grand Kru	65004	1.66	43	93.26	17.98	15	0.66
Lofa	300747	7.66	78	70.65	84.13	235	10.35
Margibi	235625	6.00	227	60.50	90.69	316	13.91
Maryland	152582	3.89	172	60.27	69.93	13	0.57
Montserrado	1255 152	31.97	1729	8.08	89.89	320	14.09
Nimba	522155	13.30	117	76.09	78.70	499	21.97
River Gee	74966	1.91	34	74.02	53.63	7	0.31
Rivercess	80264	2.04	41	96.52	47.16	14	0.62
Sinoe	114927	2.93	30	83.61	32.06	38	1.67
Total/ average	3925 773	100	105	51.22	64.54	2271	100

Notes: This table illustrates summary statistics by county, comparing available data sources provided by LISGIS (Liberian Institute of Statistics and Geo-Information Services) and the survey sample.