Interactive voice response surveys (IVRs) as a method for increasing representativeness of rural respondents in mobile phone surveys: findings from Malawi.

Malebogo Tlhajoane[1,2], Funny Muthema[3], Michael Chasukwa[3], Kelly McCain[1,4], Shammi Luhar[1], Julio Romero Prieto[1], Jacob Saikolo[3], Cremildo Manhica[5], Sarah Walters[1], Georges Reniers[1], Boniface Dulani[3]

1. Department for Population Studies, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK.
2. Department of Clinical Research, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, UK.
3. Institute of Public Opinion and Research, Zomba, Malawi
4. Department for Infectious Disease Epidemiology, Imperial College London, London, UK
5. National Institute of Health, Maputo Mozambique

§ Corresponding author: malebogo.tlhajoane@lshtm.ac.uk

Abstract

Background: Although random digit dialing (RDD) is a straightforward method for constituting a sample in a mobile phone survey (MPS), it may be resource intensive to derive a nationally (or regionally) representative sample due to gender and urban/rural differences in mobile phone ownership and usage. In this analysis, we evaluate the use of a fully automated interactive voice response (IVR) survey as a screening tool for identifying rural respondents, women in particular, in a mortality MPS in Malawi.

Methods: An IVR survey was conducted among 25,524 unique mobile phone numbers to determine (i) language of correspondence, (ii) sex, and (iii) place of residence. IVR outcomes, durations and costs were determined after 4 call attempts were made for each number, and evaluated amongst all numbers dialled.

Results: Approximately half of all phone numbers were answered, and among those 55.7% engaged with the IVR survey by answering the language of correspondence question. Where information on sex was requested, a small proportion (27.6%) were found to be female. Overall, 33.6% of respondents who engaged with the survey were found to live in a rural area, 49.9% of whom went on to complete the mortality MPS at a cost of US$8.91 per IVR number identified, and US$17.9 per mortality survey completed.

Conclusions: IVR surveys can be used to improve the representation of rural and female respondents in MPSs at an acceptable cost. Modifications to the IVR survey process (e.g. survey timing and number of call attempts) should be explored further to increase engagement rates.

Funding: Bill and Melinda Gates Foundation (INV-023211).

Extended Abstract

Introduction

The increasing coverage of mobile phones in low- and middle-income countries has provided new opportunities for rapid survey data collection [1-3]. In recent times, health research and surveillance using mobile phone surveys has become more common, as evidenced by the COVID-19 pandemic where mobility and face-to-face data collection were limited in most areas [4-6].

Various methods can be used to constitute a sampling frame for a mobile phone survey. These include: (i) sampling individuals from a secondary data source, i.e. a face-to-face survey that had been previously conducted where mobile phone numbers had been collected, (ii) obtaining phone numbers directly from telephone operators or other collaborators who may be able to identify numbers fitting certain criteria (i.e. by geographic location of use), (iii) through random digit dialing (RDD) where phone numbers are randomly generated based on local numbering structures [7, 8].

While RDD provides an easy way to develop a sampling frame for a mobile phone survey, it is often necessary to set quotas for different population strata to ensure that the resulting sample is balanced in terms of a number of background characteristics [9]. Even then, filling strata for population subgroups with low mobile phone ownership can be challenging and resource intensive if this is to be done by CATI enumerators.

In this study we evaluate the utility and costs of interactive voice response surveys (IVRs) as a method to identify hard to reach populations and increase the representativeness of a sample generated for a national mortality mobile phone survey using RDD in Malawi, by targeting rural respondents and rural women in particular.

Methods

Study Population and Sampling Criteria

This study was nested in the Rapid Mortality Mobile Phone Surveys (RaMMPS) project in Malawi [10]. Malawi provides an interesting case study for mobile phone surveys because, (i) mobile phone ownership is relatively low compared to other countries in Sub-Saharan Africa, and (ii) the population is predominantly rural [11, 12].

Target quota for the RaMMPS survey were created following the 2018 national census distribution of the population by sex, age, region and urban/rural place of residence. Quota were set with the goal of reaching a total sample size of 20,000 participants, organised into four trimesters, with a target of 5,000 completed CATI interviews each. Women of reproductive age (18-49 years) were oversampled by a third, to ensure sufficient power for estimating under-five mortality. RaMMPS interviews were conducted through Computer-Assisted Telephone Interviews (CATI) Data collection for trimester 1 began on the 24th of January 2022. Following evaluation of CATI completion rates after the first two trimesters (table 1), and in order to increase representation of rural participants within the RaMMPS sample – particularly women living in rural areas – we devised an IVR survey to identify rural participants prior to the CATI process.

	Male		Female		Total
Age Group	Urban	Rural	Urban	Rural	
18 – 49	97.5%	108.3%	92.9%	43.6%	70.2%
50 – 64	92.9%	54.5%	82.1%	36.3%	52.4%
Total	96.8%	100.5%	92.1%	43.0%	68.4%

Table 1: Progress towards filling quota targets by age, sex and place of residence for central, northern and southern regions of Malawi from the 24th of January, to the 13th of September 2022.

IVR Methodology

A script was created for the IVR survey constituting three questions: (i) language screening, (ii) gender, and (iii) urban/rural place of residence (table 2).

Audio recordings were created in each of the four main languages spoken in Malawi – Chichewa Chisena, Chiyao, Chitumbuka – and the IVR created through the use of the engageSPARK web-based platform [13]. Calls were charged at US$0.59 per minute, and mobile phone numbers were acquired and validated by a mobile sampling organization known as Sample Solutions. Sample Solutions identifies operational mobile phone numbers by drawing a simple random sample based on the numbering structure of the mobile phone providers in Malawi and verifies activity against the Home Location Register; a database that contains data regarding authorized subscribers. An estimated 80% of the numbers provided by Sample Solutions are operational.

An IVR survey was conducted among 25,524 mobile phone numbers. Calls were placed using two strategies: firstly, the IVR was conducted using all three questions included in table 2. This was fielded to 600 mobile numbers. The aim of this approach was specifically to reach rural women. Secondly, after evaluating results from the first batch of the IVR survey, the question on sex was removed with an aim to identify all those living in rural areas, thereby fulfilling quota for rural men, while also allowing for referrals of rural women among men identified by IVR.

IVR calls were made during weekdays between the hours of 09:00 and 16:00 local time. Where a call was unanswered, repeated attempts were made 15 minutes, 1 hour and 24 hours after each successive call attempt. Calls were placed using a local number in order to increase engagement rates.

Analyses

Outcomes of the IVR were recorded once all four attempts were completed for each number, and calculated as the proportion of unique numbers that had been answered, or were rural respondents (total yield), among all of the numbers dialled. The costs and duration of each call attempt made were combined for each number dialled to estimate a total cost and total duration. Total costs of the IVR were estimated against the number of rural respondents identified within the IVR. IVR data were merged with data from the RaMMPS CATI interviews to estimate call attempts and the call outcome status for mobile phone numbers pre-screened using the IVR method, and those that had undergone RDD without IVR pre-screening.

Table 2: Interactive Voice Response Survey Script

Question Number	**Category**	**Voice Prompt**	**Response Categories**
--	Greeting Message	Hello, we are calling you from the Institute of Public Opinion and Research and are conducting a short research project on the impact of COVID-19 in Malawi. We would first like to ask in which language we can best communicate.	
Q1	Language filter	Please have a look at the numbers on your telephone, and press …. If you wish to proceed in ……	1 = Chichewa 2 = Chisena 3 = Chiyao 4 = Chitumbuka
--	Project Information	For this research project, we are looking for people living in rural areas to participate. Your number has been chosen by chance.	
~~Q2~~	~~Sex~~	~~Are you male, or female? Press 1 if male, press 2 if female~~	~~1 Male~~ ~~2 Female~~
Q3	Area of Residence	Would you say the place you usually live is a city, a boma or elsewhere? Press 1 for city, press 2 for boma, press 3 for elsewhere	1 = City 2 = Boma 3 = Elsewhere (Rural)
--	Goodbye Message	"Thank you for your time. A member of our team will contact you at a later stage to participate in a 20-minute survey on the impact of COVID-19."	
--	Error Message	"That wasn't a valid reply, please try again"	
--	Ineligible Message	"Thank you very much for your time. Unfortunately, you don't fit the criteria for our survey at this time. Thank you for your participation. Goodbye"	

Findings

Among the mobile numbers received, 13 were duplicates, and removed from analysis, leaving a total sample size of 25,524. Among these, 12,906 (50.6%) were answered after all IVR call attempts were made (table 3). Of the respondents who answered the call during the IVR survey, 7,193 (55.7%) went on to answer the language question, indicating engagement with the survey. Among them, the vast majority spoke Chichewa (83.1%). Where respondents were asked their sex prior to indicating their area of residence (batch 1), most of them were found to be male (72.4%), and thus the survey ended there for them. Overall, among the respondents who answered the question on place of residence, approximately 40% were found to live in a rural area compared to 84% as reported in the 2018 national census [11]. IVR completion was estimated as 24.1%, indicating the proportion of respondents who reached the end of the survey. Most completions were found to occur during the first two call attempts, declining thereafter.

Table 3: Yield of IVR survey seeking to identify rural respondents for a MPS in Malawi

	Batch 1	**Batch 2**	**Combined**
Unique mobile numbers	**600***	**24924^**	**25,524**
Outcome			
Answered	346 (62.0%)	12,560 (51.1%)	12,906 (50.6%)
Not Answered	212 (38.0%)	12,011 (48.9%)	12,618 (49.4%)
Language Question Answered ǂ			
Yes	116 (33.5%)	7,077 (56.4%)	7,193 (55.7%)
No	230 (66.5%)	5,483 (43.7%)	5,713 (44.3%)
Sex			
Female	29 (27.6%)	-	-
Male	76 (72.4%)	-	-
Place of Residence			
City	9 (31.0%)	2,462 (41.0%)	2,471 (41.0%)
Boma ()	7 (24.1%)	1,140 (19.0%)	1,147 (19.0%)
Rural	13 (44.8%)	2,401 (40.0%)	2,414 (40.0%)
Overall IVR yield per number	2.2%	9.6%	9.5%

ǂ Proportion among those who answered call

*Outcome data missing for 42 individuals; ^ 13 of the numbers already used in previous batches

From the IVR survey, we selected two groups of respondents for follow-up CATI interviews. Batch 1 consisted of respondents who identified as female and lived in a rural area (n=13). Batch 2 consisting of all rural respondents, both men and women (n=2,401), creating a total of 2,414 respondents. Figure 1 illustrates the CATI call outcomes among the pre-screened mobile phone numbers, compared to those from the RDD sampling frame. Approximately 50% of the IVR numbers resulted in a completed CATI interview (table 3), compared to 24% of RDD numbers (figure 1, figure 2). In addition, a larger proportion of the RDD numbers were found to be inaccessible at the time of the CATI interview, compared to IVR (overall difference in proportions $p<0.001$).

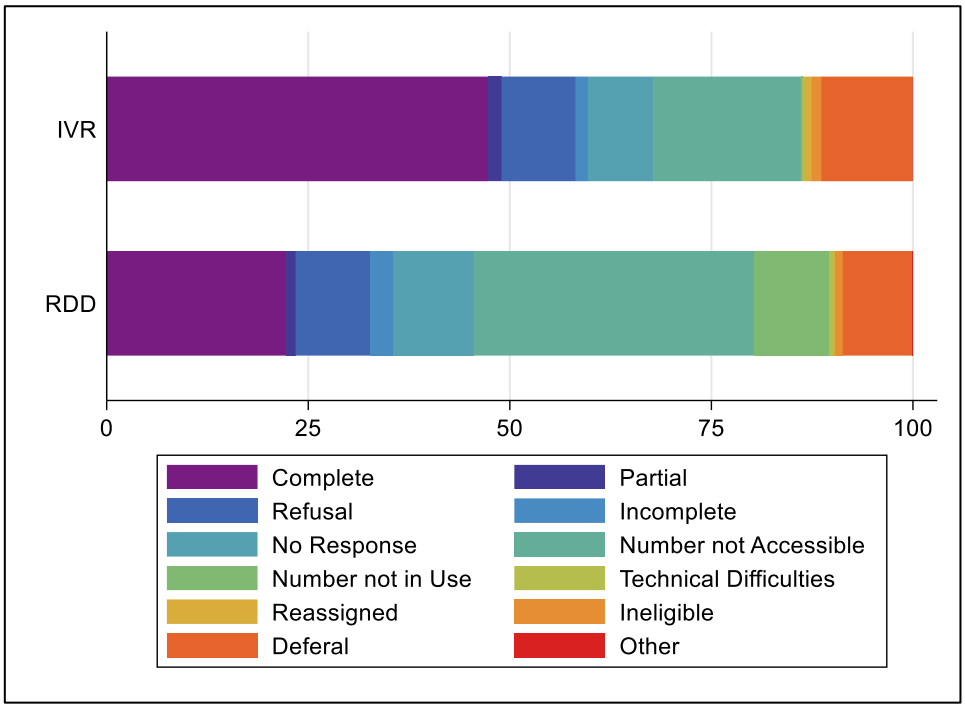

Figure 1: Mortality mobile-phone CATI call outcomes among RDD and IVR pre-screened mobile phone numbers in Malawi.

In total, 1,204 CATI interviews had been completed using mobile numbers derived from the IVR survey. Figure 2 compares the CATI call attempts and interview completion among the RDD and IVR screened numbers. In both instances we found that most CATI completions occurred during the first two call attempts and declined over time.

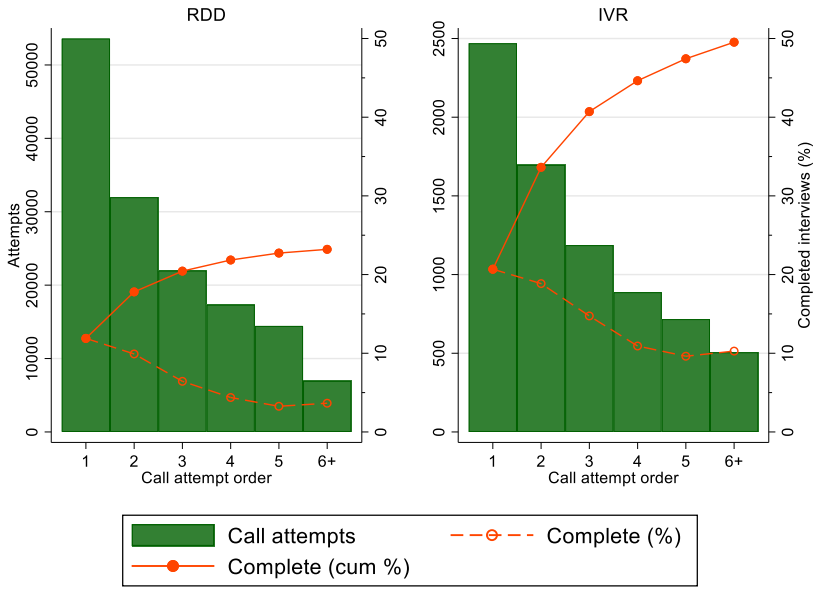

Figure 2: Mortality mobile-phone CATI completion among RDD and IVR pre-screened mobile phone numbers in Malawi.

Table 4 provides the costs related to (i) generating the mobile numbers derived through the IVR, and (ii) the costs of the IVR, per completed CATI interview. The total cost for the IVR survey was US$21,518.39, with a total mean duration of 46.1 seconds per number dialled. As such, the total cost of the IVR per rural respondent identified was US$8.91. This was lowest in the batches where all rural respondents were eligible for inclusion (both men and women). The overall cost of the IVR per completed CATI was calculated at US$17.9, ranging from US$51.1 where the IVR survey was targeting only women living in rural areas (Batch 1), to US$17.6 where all rural respondents were included (table 4).

	Batch 1	**Batch 2**	**Combined**
Unique mobile numbers	**600***	**24924^**	**25,524**
Respondents living in a rural area	13	2,401	2,414
Completed CATI	10 (76.9%)	1,194 (49.7%)	1,204 (49.9%)
Mean Duration (secs)	48.1	46.0	46.1
Total IVR cost (US$)	511.21	21,007.18	21,518.39
IVR Cost Per Rural Number Generated (US$)	39.3	8.75	8.91
Cost Per Completed CATI (US$)	51.1	17.6	17.9

Table 4: Yield and costs of IVR survey seeking to identify rural respondents for a MPS in Malawi

Preliminary Conclusions

Preliminary findings from this analysis suggest that an IVR approach may be a viable strategy to identify rural respondents in mobile phone surveys that draw their sample through RDD. Although additional resources are needed to run an IVR survey, these may be offset by a reduction in the time required by CATI enumerators to fulfil pre-set sampling criteria. Preliminary data from the CATI completion rates indicated the challenges faced in reaching rural respondents – particularly rural women in mobile phone surveys. The fulfilment of target quota for women remained consistently low across all enumeration areas, with an over-representation of men living in urban areas. Although mobile phone coverage has been increasing within many low and middle-income countries, ownership remains low among women in particular [14, 15]. Findings from the analysis of IVR results obtained in Malawi further illustrate this, as a vast majority of those who engaged with the IVR survey were men. Additional analysis evaluating the costs and resources used in conducting the CATI surveys before and after the introduction of IVR pre-screening would be useful in quantifying the degree to which IVR methods may offset the costs and enumerator time required to fulfil sampling quotas. In addition, further evaluations of the yield generated from referral of rural women by men identified through the IVR would be useful in determining the synergy of these two strategies in increasing representation of women overall [16], which is important in instances where there is a focus on generating maternal and child health statistics.

References

1. Silver KTL. Smartphone Ownership is Growing Rapidly Around the World, but Not Always Equally Pew Research Center; 2019 February 5, 2019.
2. Hu SS, Balluz L, Battaglia MP, Frankel MR. The impact of cell phones on public health surveillance. Bull World Health Organ. 2010;88(11):799.
3. MICS Plus: A step-by-step Guide to Implementation. United Nations International Childrens Emergency Fund; 2020.
4. Phadnis R, Wickramasinghe C, Zevallos JC, Davlin S, Kumarapeli V, Lea V, et al. Leveraging mobile phone surveys during the COVID-19 pandemic in Ecuador and Sri Lanka: Methods, timeline and findings. PLoS One. 2021;16(4):e0250171.
5. Gourlay S, Kilic T, Martuscelli A, Wollburg P, Zezza A. Viewpoint: High-frequency phone surveys on COVID-19: Good practices, open questions. Food Policy. 2021;105:102153.
6. Fu HS-P, C. World Bank. 2020. [cited 2023]. Available from: https://blogs.worldbank.org/opendata/high-frequency-monitoring-covid-19-impacts.
7. L'Engle K, Sefa E, Adimazoya EA, Yartey E, Lenzi R, Tarpo C, et al. Survey research with a random digit dial national mobile phone sample in Ghana: Methods and sample quality. PloS one. 2018;13(1):e0190902.
8. KRISTEN HIMELEIN SE, CHARLES LAU, DAVID MCKENZIE. World Bank. 2020. Available from: https://blogs.worldbank.org/impactevaluations/mobile-phone-surveys-understanding-covid-19-impacts-part-i-sampling-and-mode.
9. Labrique A, Blynn E, Ahmed S, Gibson D, Pariyo G, Hyder AA. 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. 2017;19(5):e121.
10. Rapid Mortality Mobile Phone Surveys: London School of Hygiene and Tropical Medicine; 2022 [Available from: https://www.lshtm.ac.uk/research/centres-projects-groups/rapid-mortality-mobile-phone-survey.
11. 2018 Malawi Population and Housing Census. National Statistical Office; 2019.
12. Marron O, Thomas G, Burdon Bailey JL, Mayer D, Grossman PO, Lohr F, et al. Factors associated with mobile phone ownership and potential use for rabies vaccination campaigns in southern Malawi. Infectious Diseases of Poverty. 2020;9(1):62.
13. engageSpark 2023 [Available from: https://www.engagespark.com/.
14. Blumenstock JE, Eagle N. Divided we call: disparities in access and use of mobile phones in Rwanda. Information Technologies & International Development. 2012;8(2):pp. 1-16.
15. The Mobile Gender Gap. GSMA; 2021.
16. Glazerman S, Grépin KA, Mueller V, Rosenbaum M, Wu N. Do referrals improve the representation of women in mobile phone surveys? Journal of Development Economics. 2023;162:103077.