Survey Research Methods (2019)
Vol. 13, No. 3, pp. 305-318
doi:10.18148/srm/2019.v13i3.7375

ISSN 1864-3361
http://www.surveymethods.org

In Search of the Optimal Mode for Mobile Phone Surveys in Developing Countries. A Comparison of IVR, SMS, and CATI in Nigeria

Charles Q. Lau
RTI International

Alexandra Cronberg
Kantar Public

Leenisha Marks
RTI International

Ashley Amaya
RTI International

Mobile phone surveys are increasingly prevalent in low- and middle-income countries. The main modes include computer-assisted telephone interviewing (CATI), interactive voice response (IVR), and short message service (SMS, or text messaging). But there is surprisingly little research to guide researchers in selecting the optimal mode for a particular survey. To address this gap, this study compares cross-sectional CATI, IVR, SMS, and face-to-face (FTF) surveys of the general population in Nigeria. We ask four research questions: (1) What are production and response rates to CATI, IVR, SMS, and FTF surveys? (2) How representative (age, gender, education, marital status, literacy, household assets, urbanicity) are CATI, IVR, and SMS respondents relative to FTF respondents? (3) Can IVR and SMS provide an unbiased estimate of voting behavior? If there is bias, to what extent can weights reduce bias? (4) How does the cost and time differ across mobile phone survey modes?
We find that FTF had the highest response rate (99%), followed by CATI (15%), IVR (3%) and SMS (0.2%). All mobile phone modes had substantial deficiencies with representativeness: mobile phones underrepresented women, older people, the less educated, and people in rural areas. There were differences in representativeness among mobile phone modes, but differences were relatively small and inconsistent. Both SMS and IVR produced biased estimates of voting relative to official statistics—but SMS was less biased than IVR. Weighting SMS and IVR data for demographic characteristics did not reduce bias. With regard to cost, we find that CATI is the most expensive mobile phone survey mode. For a survey of 3,000 completes, IVR is 43% the cost of CATI, and SMS is 24% the cost of CATI. SMS is significantly less expensive than IVR. We discuss the implications of these results for research and practice.

Keywords: telephone, CATI, IVR, SMS, Africa, mode

1 Introduction

Large-scale social surveys emerged in low- and middle-income countries in the 1950s (Murthy & Roy, 1983). For the rest of the 20th century, virtually all surveys in low- and middle-income countries used face-to-face, paper-and-pencil interviewing (PAPI). In the 2000s, computer-assisted personal interviewing (CAPI) began to replace PAPI (Caeyers, Chalmers, & De Weerdt, 2012), but almost all surveys remained face-to-face (FTF). Survey modes such as web and telephone were not yet feasible at scale due to limited infrastructure (Dabalen et al., 2016).

In the past decade, however, rapid growth in telecommunications have sparked a boom in mobile phone surveys, particularly in Africa. Between 2010 and 2016, mobile phone subscriptions nearly doubled in Africa (GMSA Intelligence, 2017). Mobile broadband (3G or faster) is projected to increase from 33% to 60% of African connections between 2016 and 2020 (GMSA Intelligence, 2017). Researchers have embraced mobile phones to collect data for public opinion surveys (GeoPoll, Inc., 2015), health interventions (Brinkel, Krämer, Krumkamp, May, & Fobil, 2014), and citizen feedback (UNICEF, 2015), among other applications. Mobile phone surveys have grown in popularity in part because FTF surveys are expensive and time consuming to plan and implement.

There are three main mobile phone survey modes: (1) Computer-assisted telephone interviewing (CATI), which uses live interviewers to call and interview respondents; (2) interactive voice response (IVR), which are automated voice surveys, and (3) short message service (SMS) surveys that use text messages. Mobile web surveys (Lau, Johnson, Amaya, LeBaron, & Sanders, 2018) are less feasible due to

Contact information: Charles Q. Lau, RTI Intarnational, 3040 Cornwallis Road, Research Triangle Park, NC 27709, USA (email: clau@rti.org)

low smartphone penetration in Africa (GMSA Intelligence, 2017). Landline penetration is near zero in Africa (Pew Research Center, 2015). Despite their widespread use, there is limited research comparing FTF to mobile phone modes—as well as mobile phone modes to each other (Gibson et al., 2017). The lack of information means that researchers have little empirical basis for selecting one mobile phone mode over another, and typically select mode based on intuition or personal experience.

In this study, we explore the representativeness, bias, and cost of mobile phone surveys in Nigeria, a country with high mobile phone coverage and the most populous in Africa (United Nations, 2018). Using random digit dialing, we conducted cross-sectional surveys via IVR ($n = 1,818$ completes) and SMS ($n = 2,759$ completes). We then compared those surveys with a CATI ($n = 3,785$ completes) survey and a gold standard FTF survey, the Nigeria Demographic and Health Survey ($n = 79,216$ completes). The paper investigates these specific questions:

1. What are production rates to CATI, IVR, SMS, and FTF surveys?

2. How representative (age, gender, education, marital status, literacy, household assets, urbanicity) are CATI, IVR, and SMS respondents relative to FTF respondents?

3. Can IVR and SMS provide an unbiased estimate of voting behavior? If there is bias, to what extent can weights reduce bias?

4. How does the cost and time differ across mobile phone survey modes?

By answering these questions, we aim to provide information to help researchers make an informed choice when selecting the optimal mode in Nigeria and similar countries.

2 Background: Mobile Phone Surveys

We begin this section with a description of each mobile phone mode (section 2.1) and a discussion of their coverage and non-response errors (section 2.2). We then summarize previous research and highlight contributions of the present study (section 2.3).

2.1 Mobile Phone Survey Modes

In CATI, live interviewers use interactive computer systems to collect data from respondents. CATI interviewers place calls, encourage sample members to participate, and administer the survey. In many countries, call centers are staffed with multi-lingual interviewers. If the interviewer and respondent do not speak the same language, the interviewer can quickly handoff a case to another interviewer. In low- and middle-income countries, CATI is most often used to re-contact participants from previous face-to-face surveys (see review of World Bank-led studies in Dabalen et al., 2016)—but cross-sectional CATI surveys also exist (e.g., Jodice, Solomon, & Peng, 2014).

IVR delivers surveys through voice (like CATI), but lacks a human interviewer. IVR surveys are automated voice surveys, in which respondents listen to prerecorded questions and answer by pressing numbers on the telephone keypad (e.g., "Press 1 for Yes, Press 2 for No"). In high-income countries, IVR also uses voice recognition to code responses, but this technology is rarely used in lower income countries. After picking up the phone, individuals hear a pre-recorded message introducing the survey. Respondents select their preferred language, consent to participate, and then answer questions. IVR surveys are less expensive than CATI because IVR does not require a physical call center or human interviewers and managers. Of course, the lack of human interviewers means IVR cannot answer questions or address respondent concerns.

SMS surveys are text-based. Respondents receive a text message introducing the survey, select their preferred language, opt into the survey, and then receive one text message per survey question. To provide their response, respondents reply with a number associated with the response (e.g., "1" for yes) or a free-text response (e.g., "yes"), and then receive the next question. Like IVR, SMS surveys are automated surveys; there is no human interviewer for large-scale surveys. Of course, respondents may assume they are exchanging text messages with a live interviewer rather than an SMS computer system, and may become frustrated if the SMS system cannot answer questions or understand the respondent's answer. SMS technology is enabled on virtually all phones, even the least sophisticated devices in Africa. In low- and middle-income countries, SMS surveys are cross-sectional (e.g., Broich, 2015; Lombaard & Richman, 2015) or used in a panel design (e.g., Lau, Johnson, et al., 2018). This stands in contrast to high-income countries, where SMS is mainly used in panel surveys or to remind participants about interview appointments (e.g., Conrad, Schober, Antoun, Hupp, & Yan, 2017; Hoe & Grunwald, 2015; Steeh, Buskirk, & Callegaro, 2007).

2.2 Errors in Mobile Phone Modes

We focus on errors of representation in this article—i.e., errors that impact the ability to draw inferences from survey respondents to the target population (Groves & Lyberg, 2010). In this section, we focus on non-random coverage and non-response errors. We highlight errors that affect all mobile phone modes equally and also describe errors specific to a particular mode.

Coverage bias emerges when there is a mismatch between the sampling frame and target population, and there are non-random differences between units on the frame and population. Mobile phone surveys exclude people without mobile phones, but mobile phone surveys often draw inference to the general population. Across sub-Saharan Africa, 16% of adults do not have a mobile phone (Mitullah & Kamau,

2013). Women, less educated individuals, and those living in rural areas are less likely to have mobile phones (Pew Research Center, 2015), which affects representativeness of the frame and can lead to coverage bias. Some surveys attempt to minimize coverage error by providing individuals with phones during face-to-face recruitment for use in subsequent mobile phone surveys (Demombynes, Gubbins, & Romeo, 2013), but this may be cost prohibitive and is not viable for studies without an in-person component.

Non-response can cause bias when there is incomplete information on sampled individuals, and there are differences between respondents and non-respondents. Relative to FTF, CATI, IVR, and SMS modes share some non-response errors that may lead to bias. All mobile phone modes lack in-person interviewers who can enhance the legitimacy of the survey and persuade reluctant people to participate using tailored approaches (Jäckle, Lynn, Sinibaldi, & Tipping, 2013). Limited telecommunications infrastructure in many countries means that phone calls may fail. Due to limited access to reliable electricity, many people only power on phones when needed—causing people to miss calls and delay receipt of text messages (Dillon, 2012). Individuals from rural areas are less likely to have electricity (W. Mitullah, Samson, Wambua, & Balongo, 2016), which can contribute to bias.

Self-administered modes (SMS and IVR) may experience more non-response errors compared to CATI. CATI interviewers perform a variety of functions that assist people who are less familiar with survey research and have less cognitive ability (i.e., older, less educated, and rural people). These functions include answering questions about the survey's legitimacy, clarifying questions, and probing respondents. Older and less educated respondents also sometimes have difficulty understanding the concept of inputting numbers into the phone to provide responses (Lerer, Ward, & Amarasinghe, 2010), as is required with IVR and SMS. In CATI, however, these respondents simply need to provide a verbal response to the interviewer, who then codes the response. Further, CATI may also produce higher screener and interview completion rates (and therefore higher response rates) than IVR and SMS because interviewers can tailor requests for survey participation, set appointments, reduce breakoffs through periodic encouragements to the respondent, and perform real-time handoffs in the event of a language barrier.

The text-based format of SMS may also lead to more non-response errors compared to voice-based CATI and IVR. SMS excludes people with limited or no literacy. Advocates of IVR point to the literacy requirement of SMS as evidence that IVR is superior for reaching low-literacy populations in Africa (Human Network International, 2016). In addition, many phones have limited inbox size. Because SMS surveys send each question as a separate message, respondents' inboxes may fill up, preventing people from answering questions (Gibson et al., 2017). Further, lower income respondents in particular may be concerned about the potential cost of SMS surveys. Although most SMS surveys provide for toll-free access, respondents may still believe that they will be charged to participate (Lau, Baker, Eyerman, Lombaard, & Thalji, 2018). These downsides of SMS may be partially offset by the fact that SMS respondents can reply to the SMS survey at their leisure. In contrast, IVR and CATI respondents can only complete the survey when they are called.

2.3 Previous Research and Contributions of the Present Study

In this section, we review research on three aspects of mobile phone surveys in low- and middle-income countries: response rates, representativeness and bias, and cost. For each aspect, we describe how the present study addresses gaps in our understanding.

Production Rates. Cross-sectional surveys using SMS and IVR in sub-Saharan Africa have low response rates, generally ranging from 0.7% - 4.8% (Broich, 2015; Lau, Baker, et al., 2018; Leo, Morello, Mellon, Peixoto, & Davenport, 2015). One outlier is an SMS survey in Kenya, with a response rate of 12% (Lau, Baker, et al., 2018). It is difficult to compare SMS and IVR response rates because these studies differ in country, length, topic, population, and sample frame. Although we are aware of CATI surveys being implemented in practice, there is scant information about production rates to cross-sectional CATI surveys in low- and middle-income countries. Further, existing research focuses on response rates—but has little information about other production rates (e.g., noncontact rate, screener completion rate, interview completion rate)—that could shed light on why modes have different response rates.

Representativeness and Bias. Most literature on representativeness compares FTF with a single mobile phone mode. This literature shows that SMS and IVR respondents are generally not representative of the larger population. For example, SMS surveys in Kenya, Ghana, Nigeria, and Uganda, underrepresent women, less educated, older, and less technologically savvy people (Lau, Baker, et al., 2018). In South Africa, SMS surveys also underrepresent young (Lombaard & Richman, 2015) and less educated people (Broich, 2015). Similarly, IVR surveys underrepresent people in rural areas, women, younger, and less educated in Afghanistan, Ethiopia, Mozambique, and Zimbabwe (Leo et al., 2015). Literature on CATI survey representativeness is largely based on panel studies that call respondents originally recruited from a face-to-face survey. These studies show few consistent associations between demographic characteristics and attrition—potentially because wide variation in country context—ranging from Peru and Honduras (Ballivian, Azevedo, & Durbin, 2015) to South Sudan (Demombynes et al., 2013) to Lebanon (Mahfoud, Ghandour, Ghandour,

Mokdad, & Sibai, 2015) to South Africa (Garlick, Orkin, & Quinn, 2017).

Studies comparing the representativeness among mobile phone modes are rare. Two studies are relevant. First, a randomized experiment in Zimbabwe showed that IVR respondents were more likely to have lower socioeconomic status, headed by women, and had more food insecurity compared to SMS (mVAM, 2017a). Lower socioeconomic status people are harder to reach in mobile (Lau, Baker, et al., 2018), suggesting that IVR reaches a more representative sample. The authors speculate this is due to IVR's ability to include illiterate people. However, the IVR and SMS surveys were not comparable: SMS was English-only, but IVR included both English and Shona, a widely spoken Zimbabwean language; most IVR respondents chose Shona. English speakers are more likely to be higher socioeconomic status, so it is difficult to attribute the differences to mode. SMS also included an incentive, whereas IVR had no incentive. Another experiment in Malawi compared SMS and CATI (mVAM, 2017b). SMS respondents were more likely to be younger than CATI respondents. SMS respondents also reported more food insecurity than CATI respondents. This result is contrary to expectations because SMS and CATI respondents had similar socioeconomic status. SMS requires literacy, and literate people are less likely to have food insecurity. One possibility, according to the authors, is that the greater food insecurity among SMS respondents reflects more measurement error in SMS.

Cost. Survey researchers attempt to maximize quality within cost and time constraints. However, there are few studies that directly compare costs of different mobile phone survey modes. One example is Ballivian et al. (2015), who report cost for monthly panel surveys in Peru and Honduras for CATI ($25 per interview), IVR ($17 per interview), and SMS ($8 per interview). Even though they are both self-administered modes, the cost for IVR is considerably more than for SMS because of the cost of recording and managing audio. To our knowledge, there are no published studies comparing the cost of cross-sectional mobile phone survey modes.

Contributions of the Present Study. Our study extends previous research by providing comparable response rates between CATI, IVR, SMS, and FTF surveys. We also contribute to the literature on representativeness in three main ways. First, we present pairwise comparisons of 4 modes (IVR, SMS, CATI, and FTF). This design allows us to compare mobile phone modes with each other, as well as with a FTF survey—painting a more comprehensive picture of the quality of each mode. Second, we purposefully designed the sample, questionnaires, and data collection to be comparable, allowing us to make direct comparisons across modes. Third, our study moves beyond demographic representativeness, and investigates bias in SMS and IVR surveys. We compare weighted estimates of voting from SMS and IVR surveys with true values from official statistics. This analysis helps us understand how demographic skews in mobile phone surveys can impact survey statistics, though we are limited by our focus on a single item. Our study also extends work by Ballivian et al. (2015) by providing additional comparisons of cost for CATI, IVR, and SMS cross-sectional surveys.

3 Data and Method

3.1 Data

IVR and SMS. We conducted IVR and SMS surveys of the Nigerian adult population (age 18-64) with mobile phones. The sample was based on a random-digit dialing (RDD) sample of the four main mobile phone networks in Nigeria (Airtel, Etisalat, MTN, and Globacom), which comprise over 99% of the market share (according to Nigerian Communication Commissions data, available at: https://www.ncc.gov.ng/stakeholder/statistics-reports/industry-overview). The RDD sample was proportionally stratified by the market share of the mobile network operator. We assigned each sampled phone number to either SMS or IVR. Data were collected in June, July, and August 2017 for IVR—and September and October 2017 for SMS. Contact times were the same for SMS and IVR (8am-8pm Monday-Friday, 11am-8pm Saturday, 2pm-8pm Sunday). After the initial contact, non-responding phone numbers were contacted up to three additional times before being finalized. Individuals who completed the survey were provided with a small incentive of approximately 0.50 US dollars.

The IVR and SMS surveys were offered in four languages (English, Hausa, Igbo, Yoruba) and asked the same 12 questions about demographics, voting, and attitudes towards gender relations. See Appendix 1 for exact question wording. The questionnaire began with a short introduction in English that described the study and allowed the respondent to select a preferred language. The IVR survey presented all four languages as options. The SMS vendor's platform, however, could only present two languages at a time, so we randomly assigned respondents to receive one of three versions of the instrument—with English and one Nigerian language (i.e., English/Hausa, English/Igbo, English/Yoruba). This design may have decreased response rates among people who cannot read English—e.g., a Hausa-only speaker who happened to receive the English/Igbo version would not have an opportunity to respond.

Supplementary analysis of SMS data showed that selecting English was associated with being younger and male—but not other variables (education, rural residence, radio ownership, or literacy). Because the SMS survey encouraged English (by always presenting it as an option), the SMS sur-

vey likely overrepresents younger people and men—an issue we return to in the Results section. In the Results section, we also present findings from a robustness check, where we conducted the analysis limited to individuals who selected English.

CATI. We analyzed data from a CATI survey of the Nigerian adult population (age 18-64). This survey was conducted between April and June, 2016. Like the IVR and SMS survey, the CATI survey was based on an RDD sample of mobile phone numbers in Nigeria, and used proportional stratification by the market share of the mobile operator. The sample was based on the four main operators (Airtel, Etisalat, MTN, and Globacom), but also included Visafone, a small operator with approximately 1% market share (Nigerian Communications Commission data; https://www.ncc.gov.ng/stakeholder/statistics-reports/industry-overview).[1] Calls were placed between 8:30am and 8:00pm Monday - Friday, and 8:30am and 12pm on Saturday.[2] The survey was conducted in the same four languages as IVR and SMS (English, Hausa, Igbo, and Yoruba) and included 86 questions about technology use. If there was a language barrier between an interviewer and respondent, the interviewer handed the call off in real-time to another interviewer.

FTF. The FTF survey is the 2013 Nigeria Demographic and Health Survey (NDHS). Part of the international Demographic and Health Surveys Program, the NDHS is a gold standard survey implemented by the Nigerian National Population Commission (Nigeria's national statistical agency). The NDHS is a face-to-face survey based on a stratified, multi-stage area probability sample of households. The sample design included all 36 states and the Federal Capital Territory of Abuja. Within each state's urban and rural areas, the sample selected localities and enumeration areas. Interviewers then conducted a household listing to enumerate and sample households. Among the 38,522 households in the NDHS, interviewers used paper-and-pencil interviewing to collect data from multiple household members. For this article, we use data from the Household Questionnaire, in which a household member reports about the characteristics of the household and household members. We do not use data from the Women's or Men's Questionnaire, in part to avoid sensitive questions about health and reproductive issues. For more information, see National Population Commission and ICF International (2014). It is worth noting that the FTF survey used pencil-and-paper interviewing rather than electronic data capture (as in mobile phone surveys), and also used proxy reporting (whereas the mobile phone surveys did not.)

3.2 Analysis

The analysis proceeds in four stages, one for each research question. Research Question 1 asks how production rates differ by modes. We present response rates for CATI, IVR, SMS, and FTF using Response Rate #3 from American Association for Public Opinion Research (2016). For all mobile phone modes, it was difficult to distinguish working from non-working numbers because of technical issues from the mobile network operators, leading to a large number of cases of unknown eligibility (Lau & DiTada, 2018). To address this issue, we use AAPOR Response Rate #3 because it adjusts for the proportion of unknown eligibility cases that are eligible. For the "Unknown Other" (UO) group, we used an adjustment factor (e) of 0.642. This factor was derived by manually dialing 600 Unknown Other cases (from the CATI survey) within three months of the CATI survey, and coding whether the number was working versus not working. For an example of a survey using this method of calculating e, see Keenedy, Keeter, and Dimock (2008). We did not include an adjustment for Unknown Eligibility (UE) cases; we took a conservative approach and assumed these cases are all eligible.

In addition to response rates, we present three production rates: (1) the rate of individuals agreeing to participate in the screener (defined as agreeing to participate among all numbers dialed)[3], (2) the screener response rate (completing the age question, conditional on agreeing to participate in the screener), and (3) the interview response rate (answering all questions in the survey, conditional on completing the screener). These three rates show the stages at which nonresponse occurs: although not comprehensive, these rates help us begin got understand reasons for response rates by mode.

For Research Question 2—about the representativeness of survey mode—we present the socio-demographic composition (age, gender, education, relationship status, rural versus urban, radio ownership, and literacy) of respondents who completed CATI, IVR, SMS, and FTF surveys. Radio ownership is one indicator of socioeconomic status in Africa. The mobile phone survey samples are all unweighted because they are random samples of phone numbers stratified proportionally by mobile network operator. The FTF data are base-weighted to account for the complex sample design. We present distribution of each variable for each mode, percentage point differences across modes, and p-values from pairwise χ^2 tests of differences.

[1] Although Visafone was not included in the samples for SMS and IVR, its small market share means the inclusion of Visafone does not affect the results.

[2] There was no Sunday data collection in CATI, unlike in IVR and SMS. This difference does not impact our findings, however, because there were no differences in sample composition between Sunday and Monday-Saturday completers in IVR or SMS—with one exception. We describe this exception in the Results section.

[3] Ideally, we would disaggregate this rate further into noncontact and nonresponse. However, our data do not contain information to distinguish noncontacts (e.g., no one picking up phone) from unit nonresponse (hanging up or refusal).

Research Question 3 asks whether IVR and SMS can provide unbiased estimates of voting—and whether weighting can reduce bias. This analysis involves comparisons in estimates of voting—i.e., whether the respondent voted in the 2015 Presidential election. We chose voting because of the availability of reliable official statistics for the comparison. We did not include the other substantive questions in SMS and IVR because we lacked adequate comparison data. Unfortunately, we are not able to include CATI in this analysis: the CATI survey did not include a measure of voting. We derive the true value for voting from official statistics (International Institute for Democracy and Electoral Assistance or IIDEA, https://www.idea.int/data-tools/country-view/231/40).[4] (The CAPI survey did not measure voting.) In this analysis, we compare unweighted estimates of voting between the IVR survey, SMS survey, and the true value from official statistics. Next, we create weights to align the SMS and IVR data to population totals from the FTF survey. We use an iterative proportional fitting algorithm (*ipfweight* in Stata) to create weights that align the SMS and IVR surveys to benchmark values from the FTF survey. We weighted by age, gender, education, and village residence. We trimmed weights to have an upper limit of 5 ($n = 67$ for IVR; $n = 115$ for SMS). We report point estimates and 95% confidence intervals (CI) for both SMS and IVR (weighted and unweighted), and use t-tests to compare (1) weighted to unweighted estimates for SMS and IVR; (2) weighted IVR to weighted SMS; and (3) weighted IVR and weighted SMS against the true value. We also report deff statistics in the text to show the impact of weighting on variance.

Next, we consider the relative cost of CATI, IVR, and SMS (but not FTF because cost data were not available). This analysis is focused on the relative cost across modes; we could not present actual costs due to arrangements with vendors. The mobile phone surveys we conducted each have a different number of completed interviews. To address this issue, we present both fixed costs and variable costs—and use this information to plot costs by mode and number of completed interviews ($n = 1,000$, $n = 3,000$, $n = 5,000$, $n = 10,000$). "Fixed costs" are not sensitive to number of completes and include project management, programming, data collection monitoring, professional audio recording (IVR only), IT support, and data file preparation. "Variable costs" are costs-per-complete—including voice airtime (CATI and IVR), charges for SMS, respondent incentives, call center staff (CATI only). We exclude our own labor in drafting the questionnaire, translation, and designing this methodological research.

This analysis provides comparable cost estimates for IVR and SMS because they both include 12 questions and otherwise have identical design features. This analysis shows how cost is sensitive to changes in number of completes. However, the analysis does not permit a completely balanced comparison between CATI and IVR/SMS because our CATI survey includes 86 questions. One potential solution is to present "cost-per-question" measures by dividing the total cost by the number of questions. However, as Dabalen et al. (2016) note, cost-per-question analyses are based on a number of questionable assumptions and should be regarded as "back of the envelope" calculations. Nevertheless, the cost for CATI would have been lower if it had 12 (rather than 86) questions, meaning this analysis overstates cost differences between CATI and IVR/SMS.

4 Results

4.1 Response and Production Rates

Table 1 shows the final dispositions, response rates, and production rates, by mode. Response rates for all mobile phone modes are lower than for the FTF survey, which had a response rate of 99%.[5] Among mobile phone modes, CATI has the highest response rate (14.55%), followed by IVR (2.89%) and then by SMS (0.23%). The higher response rate of CATI compared to IVR and SMS is due to two factors. First, among all numbers dialed, more sample members agreed to participate in CATI; we speculate this is due to the ability of CATI interviewers to persuade people to participate.[6] Second, the interview completion rate (i.e., the number of people who completed among those that completed the screener) was actually higher for CATI than for SMS (and to a lesser extent, IVR)—despite the significantly longer interview length of CATI.

The IVR response rate is nearly 13 times higher than SMS (2.89 / 0.23)—potentially because IVR can reach illiterate sample members. The IVR survey also included all four languages as options, whereas SMS only included two languages due to limitations in the SMS platform. Breakoff is another reason for SMS's lower response rate compared to IVR. The interview completion rate was higher for IVR

[4]According to IIDEA, 29,432,083 people voted in the 2015 Presidential election, out of a voting age population of 91,669,312 (32%). IIDEA statistics are based on all adults (age 18 and over), whereas the SMS and IVR surveys are based on ages 18-64. Our calculations from the 2012 Afrobarometer Nigeria data show that individuals over 65 are more likely to vote than people 64 or younger. This means the SMS/IVR estimates are lower than they should be because they do not include people age 65 and over (who vote more than their younger peers).

[5]It's also worth noting that 99% is the household-level response rate (we use data from the household screener); the response rate for the individual interview is slightly lower (95% for men and 97% for women).

[6]However, we cannot rule out the possibility that CATI had higher contact rates than IVR or SMS. Although contact rates (i.e., the proportion of sample numbers reached) should not theoretically differ by mode (especially between CATI and IVR), it is possible that CATI could have had higher contact rates.

Table 1
Final Dispositions and Response Rates, by Mobile Phone Mode (Nigeria)

Final Disposition	Definition	CATI (%)	IVR (%)	SMS (%)	FTF (%)
A. Completed	Answered all questions	10	2	0	96
B. Ineligible	Screened out due to age or unoccupied (for FTF)	0	1	0	4
C. Eligible, partial interview	Passed age screener, but did not complete survey	1	0	0	-
D. Unknown eligible	Respondent agreed to participate but eligibility not established (due to refusals or no answer for age question)	6	0	0	-
E. Unknown other	No answer	83	97	100	1
Total		100	100	100	100
AAPOR Response Rate #3[a]	$\frac{A}{(A+C+D+(e\cdot E))}$	15	3	0	99
Agree to Participate Rate	$\frac{(A+B+C)}{(A+B+C+D+E)}$	17	3	0	99
Screener Completion Rate, conditional on agreeing to participate	$\frac{(A+B+C)}{(A+B+C+D)}$	67	92	96	100
Interview Completion Rate, conditional completing screener	$\frac{A}{(A+C)}$	89	85	63	100
Number of sample members contacted		36,992	97,092	1,842,017	40,320
Number of completes		3,785	1,818	2,759	38,522

[a] See Methods section for more detail about the response rate calculation.

(85%) than SMS (63%). Lower interview completion rates for SMS may stem from respondents taking breaks and forgetting to complete the survey: SMS respondents can complete the survey at their leisure, whereas IVR does not offer this possibility.

4.2 Sample Representativeness

Research Question 2 asks whether the representativeness of mobile phone surveys (CATI, IVR, SMS) differs from FTF. Table 2 shows the socio-demographic composition of CATI, IVR, SMS, and FTF modes, percentage point differences, and p values from statistical tests. Unless noted, all differences noted in the text below are statistically significant at $p < 0.01$.

Age. All mobile phone modes under-represent older people relative to the FTF survey—particularly in the 50-64 age group, which represents 17% according to the FTF data, but only 7% in CATI, 2% in IVR, and 3% in SMS. There were also differences among mobile phone modes. The skew towards respondents age 18-29 was less evident in CATI compared to IVR and SMS. Considering the entire age distribution, IVR produced more representative data than SMS [$\chi^2(11); p = 0.01$]. SMS respondents were the youngest—65% were 18-29—though the prominence of English in SMS language selection may contribute some to the young age profile of SMS respondents (see Methods section).

Gender. Women represented 53% of the population according to the FTF data, but only 33% in CATI, 28% in IVR, and 36% in SMS. Surprisingly, SMS was more representative than CATI and IVR. The reason for the better gender representation in SMS is not entirely clear. One possibility is that some women may hesitate to pick up an incoming voice call from an unknown number—whereas SMS allows people to see the content of the message without taking any action.

Socioeconomic Status. Mobile phone modes all underrepresent less educated people relative to the FTF benchmark: In the FTF survey, 39% had no school—substantially higher than all mobile phone modes. Among the mobile phone modes, IVR is the "least bad" mode at capturing the lower end of the educational distribution: 8% of IVR respondents had no school compared to 5% in CATI, and 3% in

Table 2
Composition of FTF, CATI, IVR, and SMS General Population Surveys in Nigeria

	Survey Estimates				Percentage Point Differences					
	FTF (%)	CATI (%)	IVR (%)	SMS (%)	FTF-CATI	FTF-IVR	FTF-SMS	CATI-IVR	CATI-SMS	IVR-SMS
Age										
18-29	40	54	62	65	−14**	−22**	−25**	−8**	−11**	−3
30-39	25	27	27	23	−2**	−2**	2**	0	4**	4**
40-49	18	12	8	9	6**	10**	9**	4**	3**	−1
50-64	17	7	2	3	10**	15**	14**	5**	4**	−1
Female	53	33	28	36	20**	25**	17**	5**	−3**	−8**
Education										
No school	39	5	8	3	34**	31**	36**	−3**	2**	5**
Primary school	26	9	11	6	17**	15**	20**	−2*	3**	5**
Secondary school	22	46	32	41	−24**	−10**	−19**	14**	5**	−9**
Post-secondary school	13	40	50	51	−27**	−37**	−38**	−10**	−11**	−1
Household owns radio	70	79	71	69	−9**	−1**	1	8**	10**	2
Read very well	-	-	81	85	-	-	-	-	-	−4**
Live in Village	57	19	24	21	38**	33**	36**	−5**	−2	3**
Relationship Status										
Married or cohabiting	71	46	46	34	25**	25**	37**	0	12**	12**
Divorced, Separated, Wid.	5	2	7	6	3**	−2**	−1**	−5**	−4**	1
Single	23	52	47	60	−29**	−24**	−37**	5**	−8**	−13**
Language Selection[a]										
English	22	77	60	94	−55	−38	−72	17	−17	−34
Hausa	46	13	25	3	33	21	43	−12	10	22
Igbo	10	1	3	1	9	7	9	−2	0	2
Yoruba	15	10	11	3	5	4	12	−1	7	8
Other	6	n/a	n/a	n/a	−14	−22	−25	−8	−11	−3
Unweighted N[b]	79,216	3,785	1,818	2,759						

[1] Statistical tests not shown. [2] Sample sizes vary slightly because "don't know" responses are excluded.
* $p < 0.05$ ** $p < 0.01$

SMS.[7] IVR may be more effective at capturing less educated people it can reach illiterate sample members, whereas SMS cannot. However, we were surprised that IVR had a slight edge over CATI with respect to education: we expected CATI interviewers to be useful in explaining the survey to less educated respondents. In contrast, there is little variation in household radio ownership estimates across modes.

Literacy. One of the widely-cited benefits of IVR is that it can capture illiterate or semi-literate people, whereas SMS cannot (Human Network International, 2016). This idea has weak support in our data, however. On one hand, it is true that SMS respondents are more likely to say they read "very well" compared to IVR respondents (85% versus 81%; $p < 0.01$). But this difference is small in magnitude. More importantly, over four-fifths of IVR respondents read "very well."—suggesting that cross-sectional population based surveys are overwhelmingly comprised of literate (and highly educated) respondents.

Rural Residence. All mobile phone modes underrepresent people from rural areas. Rural residence was 57% in FTF data, but less than 25% in all mobile phone modes, potentially because of weaker mobile phone signals in rural areas or the characteristics of rural respondents (who tend to be less educated). IVR produces the most representative data (24% rural) compared to SMS (21%) and CATI (19%). These differences are small, however, and deserve further investigation.

Marital Status. All mobile phone modes underrepresent married people compared to the FTF benchmark, due in part to the younger ages among mobile phone survey respondents.

Survey Language. We include the language selected by respondents in this table for descriptive interest; comparisons

[7] Recall that interviewing occurred on Sundays for IVR and SMS, but not CATI. This difference had no impact on any study findings—except for the education comparison between SMS and CATI. This is because education was positively associated with Sunday interviews for SMS [$\chi^2(2) = 7.1; p = 0.028$]. However, this "Sunday effect" does not change our general conclusion that SMS underrepresents educated people relative to CATI. This difference remains statistically significant even when the SMS Sunday interviews are excluded [$\chi^2(3) = 90.2; p < 0.001$].

across mode do not necessarily indicate which mode produces most representative data. English was selected by 94% of SMS respondents, 77% of CATI respondents, and 60% of IVR respondents. The greater prevalence of English in SMS may reflect a preference to read and text in English compared to non-English languages (but also the languages available in SMS; see Methods section.)

Robustness Check for Language. To help understand how the available languages in SMS affected the results, we re-ran the analysis in Table 2 but restricted the sample to individuals who chose English as the survey language. The pattern of mode differences was the same as in Table 2—meaning the results are at least consistent within English language, and suggesting the language availability in SMS did not alter the overall conclusions.

Summary. The starkest differences in Table 2 are the large, statistically differences between FTF and mobile phone survey respondents. For mobile phone modes, there were substantial underrepresentation of women, older people, less educated, and rural people relative to FTF. This underrepresentation may be due to a variety of undercoverage and non-response errors. We had expected that representativeness would also vary among mobile phone modes—with CATI expected to be the most representative, followed by IVR, and then by SMS. These expectations found little support in the data. Differences across mobile phone modes were relatively small and not consistent. Each mode performed well on several demographic variables—e.g., age for CATI, gender for SMS, education and rural residence for IVR. SMS also had less bias in voting estimates compared to IVR. But there was not a single mode that outperformed others. We return to this issue in greater detail in the Discussion.

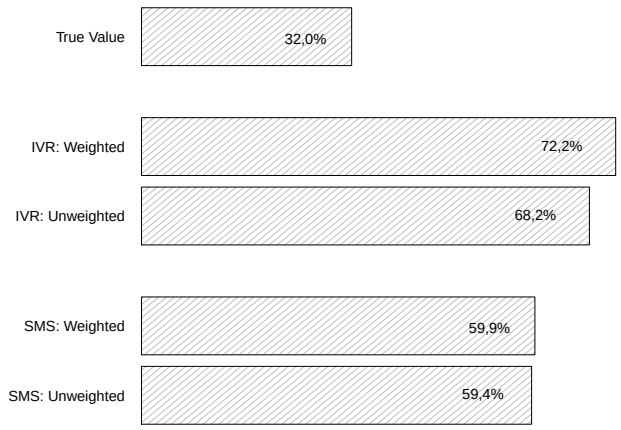
4.3 Bias in Estimates of Voting

Figure 1 shows point estimates and 95% CI of voting from the SMS and IVR surveys (unweighted and weighted), as well as the true value.

Three key results emerge. First, relative to the true value of 32%, the weighted estimates are significantly higher in both SMS (60%; $t = 29.7; p < 0.01$) and IVR (72%; $t = 37.7; p < 0.01$). This difference could be due to sample representativeness or greater participation among people who are civically engaged. Second, the SMS and IVR weighted estimates are different ($t = 8.5; p < 0.01$): there is slightly less bias in SMS compared to IVR – despite the lower response rate in SMS (Table 1). It is worth noting that SMS and IVR both have substantial bias: their estimates of voting are approximately twice as high as the true value.

Third, there is no difference between weighted and unweighted estimates for SMS ($t = 0.38; p = 0.71$). But the weighted estimate is higher than the unweighted estimate for IVR ($t = 2.60; p = 0.01$). This result implies that weighting does not impact on bias for SMS—but actually increases bias for IVR. Supplementary analysis (not shown in table) shows that the deff for weighted estimates is 3.1 for IVR and 4.0 for SMS. Weighting substantially reduces precision (and reduces effective sample size), but weighting does not improve estimates of voting.

Figure 1. Estimates of Voting in 2015 Nigerian Election, by Mode and Weighting



Costs. Table 3 shows the relative costs for CATI, IVR, and SMS surveys we conducted. The fixed costs for IVR and SMS surveys (of any size) are 32% and 14% the fixed costs of CATI. This large difference likely reflects costs in maintaining a call center, hiring interviewers, and training. Variable costs (i.e., the cost for each completed interview) for IVR are half the cost of CATI; SMS variable costs are 30% the cost of CATI—likely because self-administered modes require minimal labor costs. To make these costs more concrete, we show illustrative costs for various sample sizes. For a survey with 3,000 completes, IVR is 43% the cost of CATI. SMS is 24% the cost of CATI. These results also mean that SMS is significantly less expensive than IVR. For a survey of 3,000 completes, SMS is 56% (24/43) the cost of IVR.

5 Discussion

Our study offers direct comparisons among cross-sectional FTF, CATI, IVR, and SMS surveys of the general population in Nigeria. The FTF survey had the highest response rate (99%), followed by CATI (15%), IVR (3%), and SMS (0.2%). Relative to IVR and SMS, CATI surveys had higher agree to participate rates (likely due to the interviewer's ability to persuade sample members) and higher interview completion rates (despite a longer interview). Relative to FTF, all mobile phone surveys had substantial deficiencies with demographic representativeness: men, younger people, educated, and urban people were over-represented in

Table 3
Relative cost per complete interview: CATI, IVR, and SMS (standardized to 3,000 completed interviews)

	CATI	IVR	SMS
Number of Questions	86	12	12
Costs relative to CATI (in %)[a]			
Fixed costs[b]	100	32	14
Variable costs[c]	100	50	30
Costs by number of completes (in %)			
1,000	100	38	20
3,000	100	43	24
5,000	100	45	25
10,000	100	47	27

[a] See text for description of costs. [b] i.e., cost that is not affected by number of completes
[c] i.e., cost per complete

mobile phone surveys. There were also differences among mobile phone modes—but differences were not consistent nor large. Relative to the FTF survey, CATI performed the best with respect to age. SMS had more representative data on gender than either CATI or IVR. IVR was more representative with respect to education and rural residence. IVR was better at capturing less literate respondents than SMS, but the difference was small. Further, IVR had more bias in reports of voting than did IVR.

Based on these results, we can draw three main conclusions. First, we conclude that there are serious concerns about representativeness and bias in cross-sectional mobile phone surveys of the general population. We caution researchers against uncritically drawing inferences based on cross-sectional CATI, IVR, SMS surveys to the general population. In the present day, mobile phone surveys in low- and middle-income countries may be more appropriate for panel surveys (e.g., Dabalen et al., 2016) or list samples, for example, of program beneficiaries (e.g., Lau, Johnson, et al., 2018). Mobile phone surveys have also proven to be invaluable in collecting monitoring and evaluation data for development projects and in health interventions.

Despite this caution, we recognize that cross-sectional mobile phone surveys will continue to be conducted, under the principle that "some data are better than no data" in lower income countries where systematic information about public opinion, health, education, and other topics remains relatively scarce. When surveys are conducted for these reasons, we urge researchers to be transparent about response rates, and consider issues of representativeness and bias when interpreting the data.

Further, we recognize that cross-sectional mobile phone surveys may be the only feasible survey mode after a natural disaster, during a public health emergency, or in a fast-moving political crisis. Under these conditions, face-to-face surveys may not be an option. For example, an SMS media tracking study conducted by GeoPoll in Kenya illustrated how people's access to information changed during a government-imposed media shutdown (Elliott, 2018).

Given these challenges with representativeness and bias, one way to improve the quality of mobile phone surveys is to mix modes (de Leeuw, 2005). By mixing modes (e.g., starting with a primary mode and then contacting nonrespondents with a secondary mode), researchers can leverage the strengths of each mode. As we have observed in this study, SMS may be better able to reach women, whereas IVR can capture people with weaker literacy skills. Currently, mixed mode surveys are uncommon in low- and middle-income countries. But several recent examples show promise. In South Africa, for example, mixing modes in a survey of job training program graduates increased response rates, although did not necessarily improve representativeness (Lau, Johnson, et al., 2018). We look forward to research that identifies the best ways to mix modes.

Our second conclusion is that CATI does not produce substantially more representative data compared to IVR or SMS—despite CATI's significantly higher response rate. We had expected CATI to collect more representative data because interviewers could persuade hard-to-reach groups (e.g., older people, less educated people, people in rural areas) to participate. While CATI had a higher response rate, that higher response rate did not also lead to consistently more representative data. CATI was also significantly more expensive than IVR and SMS, though two caveats are relevant for the CATI cost estimates. First, we used a professional call center from a market research firm. Other organizations that set up temporary call centers (Dabalen et al., 2016) have lower overhead cost and may be less expensive. Second, the CATI survey asked 86 questions, compared to the 12-question IVR and SMS surveys.

To be sure, CATI is generally seen as the only mode for longer or more complex surveys. In practice, researchers rarely use IVR and SMS surveys for surveys longer than 20 or 30 questions because of concerns about respondent breakoff and resulting bias. However, the maximum length for a single IVR or SMS survey has not adequately been tested. It is worth noting that IVR or SMS surveys could also use modular designs using repeated measurements to collect data from the same person (West, Ghimire, & Axinn, 2015) or from independent samples.

We would emphasize that our caution about CATI is limited to simple, cross-sectional, general population surveys in a single country (Nigeria). We recognize that CATI may be more useful for panel surveys, where maintaining engagement with panel members through a live interviewer may be important. CATI may also be the only mode feasible for surveys that pose more complicated questions to respondents

about their economic and social well-being, among other topics. We also suspect that CATI may be more effective than self-administered modes in countries with lower mobile phone penetration and where surveys are less common. We look forward to future comparative research on modes in a wider range of countries.

Third, we conclude that there is not compelling evidence for meaningful differences between IVR and SMS—even though IVR response rates were higher than SMS (3% versus 0.2%). IVR outperformed SMS with regard to some dimensions of representativeness (age, education, and village residence), but SMS captured more female respondents than IVR. Further, IVR had greater bias in estimates of voting than SMS. (Future research should compare bias in IVR and SMS for additional items beyond voting.) IVR's ability to capture illiterate respondents is a commonly-cited argument in favor of IVR over SMS (Human Network International, 2016). However, both IVR and SMS overwhelmingly captured a highly literate population. We would again emphasize that many of these differences between IVR and SMS are small in magnitude—especially when compared to FTF.

There was a major difference in cost between IVR and SMS, however. Despite the fact that the surveys were comparable (e.g., same length, same questions, both were conducted by best-in-class international vendors, used the same sample), IVR is significantly more expensive than SMS. For a survey of 3,000 completes, SMS is 55% the cost of IVR. Based on our operational experience in other studies, we have found that IVR data collection costs more than SMS because (1) IVR requires audio-recording by professional voice talent; (2) mobile phone operators charge more for voice than text; and (3) IVR requires more start-up costs when setting up connection with mobile operators compared to SMS. Further, IVR requires more time from professional staff (not included in our cost analysis here) to conduct quality assurance of recordings and manage the audio content.

In sum, IVR and SMS performed roughly similarly with regard to representativeness and bias, but IVR cost substantially more. When designing short, simple surveys, researchers can combine our conclusions with their own research requirements and budget to draw conclusions about the best mode for their study. For example, projects with very constrained budgets would likely opt for SMS over IVR. Projects that emphasize on reaching the least educated and rural populations may opt for IVR over SMS—provided they are willing to pay significantly more for relatively minor improvements in representativeness.

We encourage readers to consider three main limitations when interpreting our conclusions. Our study was based in Nigeria only and asked simple and mostly attitude-based questions. Further, while we have aimed to produce direct comparisons across modes, there were subtle differences across modes that may have impacted the results. For example, the FTF survey used proxy reporting, whereas the mobile surveys did not. There were minor differences in the RDD sample. Different organizations administered each survey, raising the possibility of house effects. While we do not believe these limitations would meaningfully alter the study's conclusions, we recognize these are potential threats to comparability.

We are living in exciting period for survey research in low- and middle-income countries. Whereas FTF used to the sole mode of data collection, researchers now enjoy of bevy of options for survey modes—including CATI, IVR, and SMS—but also other modes including mobile web and chatbots. The ever-growing menu of survey modes offers opportunities but also challenges to researchers, who presently make decisions about modes largely based on intuition and operational experience. Researchers can use information in this study to make evidence-based decisions about survey modes. But we also recognize that mobile phone surveys in low- and middle-income countries are in their infancy. This study captures a snapshot of CATI, IVR, and SMS during a dynamic period of change. We view this study not as offering authoritative conclusions about the optimal mode, but as a first step in systematically understanding the strengths and weaknesses of each mode. Future research should address some of the limitations of this study by studying a broader range of demographic variables and comparing bias in variables beyond voting. We also encourage research in a broader range of countries—including middle-income countries in regions beyond Africa. We also look forward to research on comparing measurement error across modes, as well as research that eliminates house effects as a potential source of difference across modes.

References

American Association for Public Opinion Research. (2016). Standard definitions: Final dispositions of case codes and outcome rates for surveys, 9th edition. Retrieved from https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

Ballivian, A., Azevedo, J. P., & Durbin, W. (2015). Using mobile phones for high-frequency data collection. In D. Toninelli, R. Pinter, & P. de Pedraza (Eds.), *Mobile research methods: Opportunities and challenges of mobile research methodologies*. London: Ubiquity Press.

Brinkel, J., Krämer, A., Krumkamp, R., May, J., & Fobil, J. (2014). Mobile Phone-Based mHealth Approaches for Public Health Surveillance in Sub-Saharan Africa: A Systematic Review. *International Journal of Environmental Research and Public Health*, 11559–11582. doi:doi:10.3390/ijerph111111559

Broich, C. (2015). Offline data collection in Sub-Saharan Africa using SMS surveys: lessons learned. Paper presented at the meeting of the American Association for Public Opinion Research, Hollywood, FL.

Caeyers, B., Chalmers, N., & De Weerdt, J. (2012). Improving consumption measurement and other survey data through CAPI: Evidence from a randomized experiment. *Journal of Development Economics*, *98*, 19–33.

Conrad, F., Schober, M., Antoun, C., Hupp, A., & Yan, H. (2017). Text interviews on mobile devices. In P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, ...B. West (Eds.), *Total survey error in practice*. Hoboken: John Wiley & Sons, Inc. doi:doi:10.1002/9781119041702.ch14

Dabalen, A., Etang, A., Hoogeveen, J., Mushi, E., Schipper, Y., & von Engelhardt, J. (2016). Mobile phone panel surveys in developing countries: A practical guide for microdata collection. In *Directions in development–poverty*. Washington, DC: World Bank.

de Leeuw, E. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, *21*(5), 233–255.

Demombynes, G., Gubbins, P., & Romeo, A. (2013). Challenges and opportunities of mobile phone-based data collection: Evidence from South Sudan. Retrieved from http://documents.worldbank.org/curated/en/385371468102891764/pdf/wps6321.pdf

Dillon, B. (2012). Using mobile phones to collect panel data in developing countries. *Journal of International Development*, *24(4)*, 518–527. doi:doi:10.1002/jid.1771

Elliott, R. (2018). Geopoll report on the Kenya media shutdown. GeoPoll Blog, February 21, 2018. Retrieved from https://blog.geopoll.com/kenya-media-shutdown-tv-switch-off-kenya

Garlick, R., Orkin, K., & Quinn, S. (2017). Call me maybe: Experimental evidence on using mobile phones to survey microenterprises. Retrieved from https://pedl.cepr.org/sites/default/files/Garlick

GeoPoll, Inc. (2015). Perceptions of south africa's state of the nation: Tns. Retrieved from https://research.geopoll.com/case-studies/tns-state-of-the-nation-south-africa.html

Gibson, D. G., Pereira, A., Farrenkopf, B. A., Labrique, A. B., Pariyo, G. W., & Hyder, A. A. (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. doi:doi:10.2196/jmir.7428

GMSA Intelligence. (2017). The mobile economy Sub-Saharan Africa 2017. Retrieved from https://www.gsmaintelligence.com/research/?file=7bf3592e6d750144e58d9dcfac6adfab&download

Groves, R. M. & Lyberg, L. (2010). Total Survey Error: Past, present, and future. *Public Opinion Quarterly*, 849–879. doi:https://doi.org/10.1093/poq/nfq065

Hoe, N. & Grunwald, H. (2015). The role of automated SMS text messaging in survey research. *Survey Practice*, *8(6)*, 1–15.

Human Network International. (2016). How we became sms skeptics. Human Network International blog post. Retrieved from http://hni.org/blog/2016/08/11/how-we-became-sms-skeptics/

Jäckle, A., Lynn, P., Sinibaldi, J., & Tipping, S. (2013). The effect of interviewer experience, attitudes, personality and skills on respondent co-operation with face-to-face surveys. *Survey Research Methods*, *7(1)*, 1–15.

Jodice, D., Solomon, S., & Peng, D. (2014). Understanding egyptian public opinion: Setting the demographic framework for telephone and internet research. Paper presented at the annual meeting of the World Association for Public Opinion Research, Nice, France.

Keenedy, C., Keeter, S., & Dimock, M. (2008). A "brute force" estimation of the residency rate for undetermined telephone numbers in an RDD Survey". *Public Opinion Quarterly*, *72*, 28–39.

Lau, C., Baker, M., Eyerman, J. D., Lombaard, A., & Thalji, L. M. (2018). How representative are SMS Surveys in Africa? experimental evidence from four countries. *International Journal of Public Opinion Research*. doi:doi:10.1093/ijpor/edy008

Lau, C. & DiTada, N. (2018). Identifying non-working phone numbers for response rate calculations in Africa. *Survey Practice*, 11–2.

Lau, C., Johnson, E., Amaya, A., LeBaron, P., & Sanders, H. (2018). High stakes, low resources: What mode(s) should youth employment training programs use to track alumni? evidence from South Africa. *Journal of International Development*. doi:doi:10.1002/jid.3359

Leo, B., Morello, R., Mellon, J., Peixoto, T., & Davenport, S. (2015). Do mobile phone surveys work in poor countries? Retrieved from https://www.cgdev.org/sites/default/files/CGD-Working-Paper-398-Mobile-Phones.pdf

Lerer, A., Ward, M., & Amarasinghe. (2010). Evaluation of IVR data collection uis for untrained rural users. Proceedings of the First ACM Symposium on Computing for Development, London, United Kingdom, ACM Press.

Lombaard, A. & Richman, M. (2015). Mobile research in emerging markets: Taking the step into the world of probability sampling. Paper presented at the meeting of the World Association for Public Opinion Research, Buenos Aires, Argentina.

Mahfoud, Z., Ghandour, L., Ghandour, B., Mokdad, A., & Sibai, A. M. (2015). Cell phone and face-to-face inter-

view responses in population-based surveys: How do they compare? *Field Methods*, *27(1)*, 39–54. doi:doi:10.1177/1525822X14540084

Mitullah, W., Samson, R., Wambua, P., & Balongo, S. (2016). Building on progress: Infrastructure development still a major challenge in africa. afrobarometer dispatch no. 69. Retrieved from http://afrobarometer.org/sites/default/files/publications/Dispatches/ab_r6_dispatchno69_infrastructur%20remains_challenge_en.pdf

Mitullah & Kamau. (2013). The partnership of free speech & good governance in africa. Retrieved from http://afrobarometer.org/sites/default/files/publications/Briefing

Murthy, M. N. & Roy, A. S. (1983). Development of the sample design of the indian national sample survey during its first 25 rounds. In M. Bulmer & D. P. Warwick (Eds.), *Social research in developing countries*. London: John Wiley & Sons, Ltd.

mVAM. (2017a). Mind the Mode IVR vs SMS in Zimbabwe. Retrieved from http://mvam.org/2016/12/15/mind-the-mode-ivr-vs-sms-in-zimbabwe/

mVAM. (2017b). Mind the mode: Who's testing & who's talking in malawi? Retrieved from http://mvam.org/2017/07/28/mind-the-mode/

National Population Commission and ICF International. (2014). Nigeria demographic and health survey 2013.

Pew Research Center. (2015). Cell phones in Africa: Communication lifeline. Retrieved from http://www.pewglobal.org/2015/04/15/cell-phones-in-africa-communication-lifeline/

Steeh, C., Buskirk, T., & Callegaro, M. (2007). Using text messages in u.s. mobile phone surveys. *Field Methods*, *19(10)*, 59–75. doi:doi:10.1177/1525822X06292852

UNICEF. (2015). UNICEF's U-Report social platform hits 1 million active users. Retrieved from https://www.unicef.org/media/media_82583.html

United Nations. (2018). World statistics pocketbook: 2018 edition.

West, B. T., Ghimire, D., & Axinn, W. G. (2015). Evaluating a modular design approach to collecting survey data using text messages. *Survey Research Methods*, *9(2)*, 111–123.

Appendix
Question wordings
(Appendix follows on next page)

Table A1
Question Wording

Question	SMS Survey	IVR Survey	CATI Survey	CAPI Survey
1. Age	How old are you? Reply with a number. 1)17 or younger 2)18-29 3)30-39 4)40-49 5)50-64 6)65 or older	How old are you? For 17 or younger, press 1. For 18-29, press 2. For 30-39, press 3. For 40-49, press 4. For 50-64, press 5. For 65 or older, press 6. If you don't know, press 9.	In which of these age categories do you fall? 17 or younger, 18-24, 25-34, 35-44, 45-54, 55-64, or 65 or over? Then: What is your exact age?	Recorded during household roster.
2. Gender	Are you male or female? Reply with a number. 1)Male 2)Female	Are you male or female? For Male, press 1. For Female, press 2.	Coded by interviewer	Coded by interviewer
3. Education	What is the highest level of education you have completed? 1)No School 2)Primary school 3)Secondary school 4)Post-secondary school 5)Don't know	What is the highest level of education you have completed? For No School, press 1. For Primary school, press 2. For Secondary school, press 3. For Post-secondary school, press 4. For Don't know, press 9.	What is the highest level of school you have completed? No formal school or not completed primary, primary school, secondary school, any post-secondary education.	What is the highest level of school you have completed? Interviewer codes into 11 categories (not read to respondent).
4. Marital	What is your marital status? Reply with a number. 1)Married or living with partner 2)Divorced or separated 3)Widowed 4)Single 5)Don't know	What is your marital status? For Married or living with partner, press 1. For Divorced or separated, press 2. For Widowed, press 3. For single, press 4. For Don't know, press 9.	What is your current marital status? Married or living with partner, divorced or separated, widowed, or never married and never lived with partner	What is your current marital status? Married or living with partner, divorced or separated, widowed, or never married and never lived with partner.
5. Urban	Do you live in a village or in a city? Reply with a number. 1)Village 2)City 3)Don't know	Do you live in a village or in a city? For Village, press 1. For City, press 2. For Don't know, press 9.	n/a	n/a
6. Radio	Do you have a radio in your household? Reply with a number. 1)Yes 2)No 3)Don't know	Do you have a radio in your household? For Yes, press 1. For No, press 2. For Don't Know, press 9.	Do you have a radio in your household? Yes, No	Do you have a radio in your household? YES, NO.
7. Read	How well can you read English? Reply with a number. 1)Very well 2)Somewhat well 3)Not at all 4)Don't know	How well can you read [FILL LANGUAGE OF INTERVIEW]? For Very well, press 1. For Somewhat well, press 2. For Not at all, press 3. For Don't know, press 9.	n/a	How well do you read English? Would you say excellent, good, fair, poor, or not at all?
8 Vote	Did you vote in the 2015 presidential election? Reply with a number. 1)Yes 2)No 3)Don't know	Did you vote in the 2015 presidential election? For Yes, press 1. For No, press 2. For Don't Know, press 9.	n/a	n/a
9. Housework	In your opinion, should men share housework with women like cleaning and cooking? Reply with a number. 1)Yes 2)No 3)Don't know	In your opinion, should men share housework with women like cleaning and cooking? For men should share, press 1. For women should do all, press 2. For Don't Know, press 9.	n/a	n/a
10. Gender discrimination	In your opinion, how often are women treated unequally by employers? Reply with a number. 1)Always 2)Often 3)Rarely 4)Never 5)Don't know	In your opinion, how often are women treated unequally by employers? For Always, press 1. For Often, press 2. For Rarely, press 3. For Never, press 4. For Don't Know, press 9.	n/a	n/a
11. Inheritance	Do you agree or disagree that women's share of inheritance should equal men's share? Reply with a number. 1)Agree 2)Disagree 3)Don't know	Do you agree or disagree that women's share of inheritance should equal men's share? For Agree, press 1. For Disagree, press 2. For Don't Know, press 9.	n/a	n/a
12. Intimate Partner Violence	Do you think a husband is justified in hitting or beating his wife if she goes out without telling him? Reply with a number. 1)Yes 2)No 3)Don't know	Do you think a husband is justified in hitting or beating his wife if she goes out without telling him? For Yes, press 1. For No, press 2. For Don't Know, press 9.	n/a	n/a