

Original Paper

Building the Evidence Base for Remote Data Collection in Low- and Middle-Income Countries: Comparing Reliability and Accuracy Across Survey Modalities

Abigail R Greenleaf¹, MPH; Dustin G Gibson², MS, PhD; Christelle Khattar²; Alain B Labrique², MHS, MS, PhD; George W Pariyo², MBChB, PhD

¹Johns Hopkins Bloomberg School of Public Health, Department of Population, Family and Reproductive Health, Baltimore, MD, United States

²Johns Hopkins Bloomberg School of Public Health, Department of International Health, Baltimore, MD, United States

Corresponding Author:

Abigail R Greenleaf, MPH

Johns Hopkins Bloomberg School of Public Health

Department of Population, Family and Reproductive Health

615 N Wolfe Street

Baltimore, MD, 21205

United States

Phone: 1 410 955 3543

Fax: 1 410 955 2303

Email: agreenleaf@jhu.edu

Abstract

Background: Given the growing interest in mobile data collection due to the proliferation of mobile phone ownership and network coverage in low- and middle-income countries (LMICs), we synthesized the evidence comparing estimates of health outcomes from multiple modes of data collection. In particular, we reviewed studies that compared a mode of remote data collection with at least one other mode of data collection to identify mode effects and areas for further research.

Objective: The study systematically reviewed and summarized the findings from articles and reports that compare a mode of remote data collection to at least one other mode. The aim of this synthesis was to assess the reliability and accuracy of results.

Methods: Seven online databases were systematically searched for primary and grey literature pertaining to remote data collection in LMICs. Remote data collection included interactive voice response (IVR), computer-assisted telephone interviews (CATI), short message service (SMS), self-administered questionnaires (SAQ), and Web surveys. Two authors of this study reviewed the abstracts to identify articles which met the primary inclusion criteria. These criteria required that the survey collected the data from the respondent via mobile phone or landline. Articles that met the primary screening criteria were read in full and were screened using secondary inclusion criteria. The four secondary inclusion criteria were that two or more modes of data collection were compared, at least one mode of data collection in the study was a mobile phone survey, the study had to be conducted in a LMIC, and finally, the study should include a health component.

Results: Of the 11,568 articles screened, 10 articles were included in this study. Seven distinct modes of remote data collection were identified: CATI, SMS (singular sitting and modular design), IVR, SAQ, and Web surveys (mobile phone and personal computer). CATI was the most frequent remote mode (n=5 articles). Of the three in-person modes (face-to-face [FTF], in-person SAQ, and in-person IVR), FTF was the most common (n=11) mode. The 10 articles made 25 mode comparisons, of which 12 comparisons were from a single article. Six of the 10 articles included sensitive questions.

Conclusions: This literature review summarizes the existing research about remote data collection in LMICs. Due to both heterogeneity of outcomes and the limited number of comparisons, this literature review is best positioned to present the current evidence and knowledge gaps rather than attempt to draw conclusions. In order to advance the field of remote data collection, studies that employ standardized sampling methodologies and study designs are necessary to evaluate the potential for differences by survey modality.

(*J Med Internet Res* 2017;19(5):e140) doi: [10.2196/jmir.7331](https://doi.org/10.2196/jmir.7331)

KEYWORDS

mHealth; developing countries; Africa South of the Sahara; cell phones; health surveys; reproducibility of results; surveys and questionnaires; text messaging; interviews as topic; humans; research design; data collection methods

Introduction

In low- and middle-income countries (LMICs), where vital registration, surveillance, and health record systems are underdeveloped [1], improved modes of data collection are needed [2]. Public health practitioners could benefit from more timely estimates and indicators to better plan programs, design interventions, and assess progress. The financial and human resource burden of a large survey as well as the need for more frequent data collection, particularly as mandated by the Sustainable Development Goals [3], all justify an improved system to monitor health indicators in LMICs.

The rapid increase of mobile phone ownership in LMICs offers a platform for low-cost, frequent data collection. Urbanization, increased mobile phone network coverage, and the low cost of purchasing a mobile phone have contributed to increased mobile phone ownership in LMICs [4]. According to the International Telecommunications Union, in 2015 the number of mobile subscriptions worldwide was 98.66 per 100 people [5]. Increased mobile phone ownership presents the opportunity to survey respondents remotely, whether via short message service (SMS), computer-assisted telephone interview (CATI), interactive voice response (IVR), or Web surveys. The advantages and disadvantages of these interview modalities are discussed by Gibson et al [6]. As remote data collection becomes more common in LMICs, the reliability and accuracy of data collected should be compared with established methods, including the reference-standard household survey.

There is a well-established body of literature on mobile phones as survey instruments in high-income countries [7-11], but there is a dearth of rigorous research that compares the quality, reliability, and accuracy of remote data collection modes in LMICs. Data collection mode can influence social desirability bias, can impact response rates, or can change the cognitive process for answer retrieval [9]. Although misreporting of sensitive behaviors has long been of interest to survey methodologists, a superior interviewing tool is yet to be identified for use in LMICs [12]. Cognitive models illustrate

how mode of data collection affects information retrieval, judgments about the appropriate responses, and answer choices [11]. Notably, response rates are traditionally lower for remote data collection compared with face-to-face (FTF) data collection [10].

The purpose of this literature review was to identify and synthesize the available literature from LMICs that compare a mode of remote health data collection with at least one other data collection mode. We also discuss reliability and construct validity across measures. By synthesizing the research that compares a mode of remote data collection to another mode, we identify the strengths and limitations of remote modes in LMICs as well as areas for future research.

Methods

This literature review utilized the search terms and primary inclusion and exclusion criteria from a previously conducted literature review in March and April 2015 [6]. We adapted search terms for mobile phone, IVR, text message, survey, questionnaire, and data collection to each database's classification system to query seven databases of peer-reviewed and grey literature. The abstracts and titles were screened against a set of primary inclusion and exclusion criteria [6]. The primary inclusion criteria required that the research was collected from the respondent by SMS, IVR, CATI, or via mobile phone.

Once all articles that met the primary inclusion and exclusion criteria were identified, two of the authors (AG and CK) independently reviewed the articles using the secondary inclusion and exclusion criteria, as listed in [Textbox 1](#).

Summary briefs, inaccessible full texts, and a number of manuscripts that used a remote form of data collection but did not compare the method against a standard were excluded. The references of articles included in this review were searched to find relevant publications that were not identified by the literature search. Surveys included in this review were not required to be nationally representative.

Textbox 1. Secondary inclusion and exclusion criteria.

Secondary inclusion criteria

- Study conducted in LMIC as defined by the World Bank
- Two or more modes of data collection are compared in the study
- At least one of the data collection modes is remote
- The survey includes questions about health

Secondary exclusion criteria

- Studies without a human component to the research
- Studies that collect only adherence information or that strictly examine the use of reminders for health-seeking behaviors and outcomes
- Studies that compare modes of facility-based surveillance data collection

Table 1. Categories of data collection included in this literature review.

Survey administration	Remote	In-person
Interviewer administered	CATI ^a	FTF ^b
Self-administered	IVR ^c	IVR ^d
	SMS ^{e,f}	SAQ ^g
	Postal SAQ	
	Web MP ^h	
	Web PC ⁱ	

^aCATI: computer-assisted telephone interview.

^bFTF: face-to-face.

^cIVR: interactive voice response.

^dIVR is traditionally administered remotely but can be administered in-person by the interviewer by handing a phone to the respondent that plays an IVR survey.

^eSMS: short message service.

^fTwo forms of SMS surveys were included in this study: single sitting (survey completed at one time) and modular (an SMS sent each day until survey is completed).

^gSAQ: self-administered questionnaire.

ⁱWeb PC: Web on personal computer.

^hWeb MP: Web on mobile phone.

A standardized data collection tool containing inclusion and exclusion criteria was completed by two reviewers for each screened article. Variables in the extraction form included study design, study location, data collection mode, response rates, sensitive questions, study limitations, cost, and findings. Once compiled, the two reviewers discussed any differences in their respective reviews to make final inclusion or exclusion decisions. The two reviewers relied on a third person to clarify any inclusion disagreements.

Included articles were grouped by location of respondent in relation to interviewer (remote or in-person) and by the person administering the questionnaire (self-administered or interviewer administered; see [Table 1](#)). If the participant was in a different geographic location from the interviewer while administering the survey (no FTF interaction), data collection was classified as remote. In-person data collection was defined as FTF interviewer-respondent interaction. Self-administered was defined as surveys where respondents answer without questions from the interviewer. Interviewer administered was defined as the interviewer speaking with the respondent to elicit responses. FTF surveys were defined as an in-person, interviewer-administered survey. CATI was the only form of remote interviewer-administered survey included in this literature review. IVR, SMS, Web surveys, and a self-administered questionnaire (SAQ) sent back via post were defined as remote, self-administered surveys. There were two examples of an in-person self-administered survey: SAQ and IVR administered on-site.

We reported inter-method reliability for articles that compared the same respondents across two or more modes. For articles that did not compare the same population across the two modes, we considered external construct validity, the degree to which a measure satisfies theoretical predictions about a measurement. Included articles were reviewed and marked for sensitive

questions, defined as asking about a subject that is private or taboo [13]. In assessing the results, we considered not only statistical significance, but effect size, direction of difference, and potential confounding factors [11].

Results

Overview

The parent systematic literature search identified 11,568 records, which after removing the duplicates and adding 6 articles identified by the authors, was reduced to 6625 records (see [Figure 1](#)). The primary inclusion and exclusion criteria further decreased the number of articles to 145. After removing 126 articles that did not include a comparison mode and 9 surveillance articles, we conducted full-text abstraction on 10 articles that compared two or more modes of data collection in a LMIC, with at least one form being remote data collection (see [Table 2](#)). All but one of the articles were published between 2011 and 2015. The 10 articles collected data in 7 countries, across 4 regions (Asia, Latin America, Europe, and the Middle East); notably none took place in Sub-Saharan Africa (SSA). One article reported data from 2 countries, Honduras and Peru [14].

Ten distinct modes of data collection were used in the 10 studies (see [Table 3](#)). The three in-person modes were FTF [14-21], SAQ, and in-person IVR [18]. The 7 types of remote data collection included two types of phone calls, IVR (remote) [14,22] and CATI [14,16,17,20,23,24]; and three modes that required respondents to type their responses into a mobile phone or computer, including SMS (all but one were singular-design) [14,15,19,21,24] and two types of Web surveys [25,26], (one administered on a personal computer (PC) and the other taken via a mobile browser on a smartphone). The most frequent mode of data collection was FTF (5/10 studies) and second most

frequent was SMS (4/10 studies). The 10 articles made 25 comparisons, of which 12 were from a World Bank study in Latin America [14]. The most common comparison was FTF and CATI [14,16,17,20] (compared 5 times in 4 articles) and the second most common comparison was FTF to SMS [14,15,19] (compared 4 times in 3 articles).

The majority (8/10) of the articles collected cross-sectional data and did not provide respondents with a mobile phone; only the

studies in Nepal and those in Honduras and Peru provided respondents with mobile phones [14,24]. Respondents were sampled in a variety of ways. Four of the identified articles were population-based studies, all of which enrolled participants FTF [14,16,17,20]. Five studies compared the same population across methods of data collection [14,15,19,20,25]. Finally, 6 of the 10 articles included sensitive questions [14,18,20,24,25,26].

Figure 1. Flowchart of articles identified and included in review.

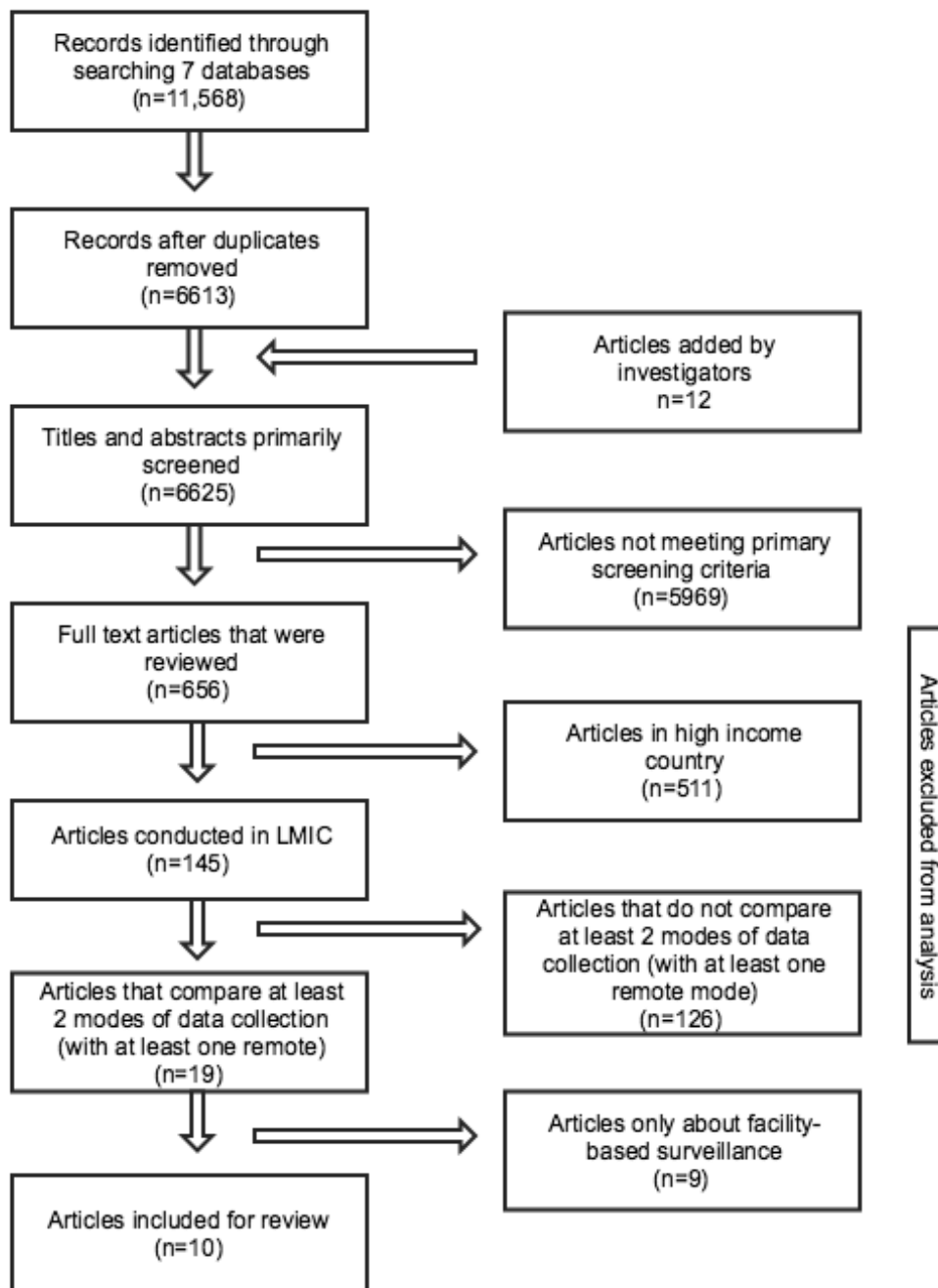


Table 2. Types of data collection which are compared and the number of included studies.

Method #1	Method #2						
	CATI ^a	IVR ^b remote	SMS ^c : singular	SMS: modular	Web MP ^d	IVR in-person	SAQ ^e remote
CATI			1	1			
FTF ^f	5	2	4				
SAQ in-person						1	1
IVR remote	2						
Web PC ^g					2		
SAQ remote						1	
SMS: singular	2	2		1			

^aCATI: computer-assisted telephone interview.

^fFTF: face-to-face.

^bIVR: interactive voice response.

^eSAQ: self-administered questionnaire.

^cSMS: short message service.

^gWeb PC: Web on personal computer.

^dWeb MP: Web on mobile phone.

Table 3. Comparison of in-person interviewer administered compared with remote interviewer administered.

Author (year)	Country (sample type) ^a	Data collection #1 (sample size)	Data collection #2 (sample size)
Ballivian [14] (2013)	Honduras (dependent)	FTF ^b (1500)	CATI ^c (600)
Ballivian [14] (2013)	Peru (dependent)	FTF (1500)	CATI (384)
Ferreira [16] (2011)	Brazil (independent)	FTF (4048)	CATI (440)
Francisco [17] (2011)	Brazil (independent)	FTF (2636)	CATI (2015)
Mahfoud [20] (2014)	Lebanon (dependent)	FTF (2836)	CATI (771)

^aWhen participants were the same across modes, we classified the sample as dependent. Different participants across modes was labeled as an independent sample.

^bFTF: face-to-face.

^cCATI: computer-assisted telephone interview.

Comparison of Modes of Data Collection

In-Person Interviewer Administered Compared With Remote Interviewer Administered

We identified 5 comparisons of FTF interviews with CATI surveys in 4 articles [14,16,17,20]. One article included comparisons of FTF and CATI in both Peru and Honduras [14]. Of these 5 comparisons, 2 compared responses in independent samples [16,17] and 3 used the same population across the two modalities [14,20]. All comparisons generally showed concordance of results between modes.

An independent probability sample of respondents in Brazil who were over 18, had a landline phone, and were interviewed

via CATI, produced estimates similar to an independent sample of respondents who also have a landline and who answered a household survey (FTF) [16]. The respondents contacted via CATI had statistically significant different estimates for 5 (number of household residents, mean age, schooling, smoking, health insurance) of the 18 measures compared with the FTF respondents with a landline. CATI respondents when compared with all FTF respondents (regardless of landline ownership) differed on 8 of the 18 variables, but after applying poststratification weights, only three estimates were biased.

The second study in Brazil compared FTF respondents with CATI landline respondents, sampled in the same manner as the aforementioned article. Two of the four estimates (diabetes,

asthma, bronchitis, or emphysema) were the same between the two samples. The other two estimates (hypertension and osteoporosis) had a higher reported prevalence among the CATI respondents [17]. Nonetheless, authors from both the studies concluded that the telephone survey was a rapid alternative to FTF surveys to provide global prevalence estimates.

After a FTF survey in Lebanon, half of the respondents were called on their mobile phones and asked an abridged version of the FTF questionnaire [20]. Overall, there was high concordance (kappa) between the CATI and FTF surveys. Kappa was above .8 for measures including age, health insurance, diabetes, current cigarette-smoking (highest agreement and kappa: agreement=95.6%, $\kappa=.91$), and ever cigarette-smoking (second highest agreement=93.5%, $\kappa=.87$). Kappa was between .6 and .8 for questions about current water-pipe smoking and past-year alcohol consumption. Reports of past-year alcohol consumption was slightly higher via CATI compared with FTF [20]. The authors concluded that estimates from the modes are reasonably comparable when data were stratified by age, gender, and education and that the difference in past-year alcohol consumption may be caused by social desirability bias [20].

The World Bank's study in Honduras and Peru aimed to validate a survey across four modalities: FTF, IVR, CATI, and SMS [14]. The study enrolled households who answered a 7-question FTF survey on household assets and poverty into a panel survey. The survey included questions about water and sanitation. Participants in Peru were sampled using the National Statistics office sampling frame, and households below the poverty line were oversampled. Honduras also used probabilistic sampling but used the Gallup World Poll Sampling Frame and did not oversample households below the poverty line. Regardless of mode, attrition was highest among less educated, less affluent, older, rural participants [14]. In both the countries, CATI estimates were very similar to the estimates collected FTF. Compared with the FTF survey, discordant responses from the panel in Honduras ranged from -2.1% to 0% for CATI (unreported for Peru). Furthermore, CATI had the lowest discordance with the FTF survey compared with SMS and IVR [14].

In-Person Interviewer Administered Compared With Remote Self-Administered

Three articles made 6 comparisons of in-person interviewer administered (all FTF) and either IVR or SMS (see Table 4). Two articles from China, both about infant feeding practices, used the test-retest method to compare FTF and SMS surveys [15,19]. One of the articles began with a FTF interview then followed up with a SMS survey [15], and the other article interviewed participants in the opposite order [19]. Both articles administered the second survey after a short time period (less than 24 h and less than 3 days). The study that sent 10 text messages to participants then followed up with FTF surveys had moderate to good agreement and 62.4% of questions had the same answers for both surveys [19]. All but one kappa and inter-class correlation were between .56 and .76; (the outlier kappa=.23 was for a question about the usefulness of a feeding calendar). The last question, which was a multiple-choice categorical question, about the source of feeding knowledge,

had the highest agreement (85% of the 33 responses were the same across methods) and a kappa value of .76.

Data agreement in the other Chinese feeding study was inconsistent [15]. The highest agreement was a kappa of .86 for the first question on the survey which was about breastfeeding the day before. The other 4 questions had moderate to poor agreement. Data agreement was worst for dietary recall, with a kappa ranging from .02 to .36 for the 7 food categories. The authors proposed that certain terms were difficult for mothers to understand (eg, iron-fortified food, solid or semi-solid food) and that during the FTF survey, the interviewers could explain these concepts, a feat that SMS cannot achieve due to the limited characters in a text and constraints on the number of texts a respondent is willing to receive. Du et al also explored the length of time between the two surveys (3 hours compared to 8 hours) but did not find a statistically significant difference in reported measure by the length of time between surveys.

The World Bank study in Honduras and Peru compared FTF with two modes of remote self-administered: IVR and SMS. Compared with the FTF survey, discordant responses from the panel in Honduras ranged from -14.6% to 12.7% for IVR, and from -15.6% to 15.3% for SMS. Discordant responses were similar for IVR and SMS on a per question basis. The IVR and SMS responses were statistically significantly different from the FTF estimates. To assess the reliability, the same respondents were asked a question second time, within 10 weeks of the first administration of the questions. The total reliability coefficient in Honduras for SMS and FTF were quite similar (.74 and .77, respectively). IVR had the highest reliability coefficient (.86); but because the IVR results were most discordant with the other modes, the authors concluded that IVR was not a suitable mode for this survey. However, authors were satisfied with the reliability of SMS surveys in their study's context.

In-Person Self-Administered Compared With Remote Self-Administered

To assess human immunodeficiency virus (HIV)-related risk behaviors among Hong Kong migrant men who were aged 18-60 years, authors systematically sampled 2416 migrants at a customs check point in Hong Kong. Authors compared three modes of data collection (see Table 5). First, all participants completed a FTF demographic survey and then were randomized to one of the three modes (in-person IVR, SAQ returned on-site in a self-sealing envelope, and SAQ to be completed off-site and returned via post) [18]. This article also included a comparison of two in-person self-administered questionnaires (in-person IVR compared with in-person SAQ; see Table 6). The authors found differential reporting of sensitive behaviors by mode. The low response rate (only 36% of men randomized to complete the postal SAQ returned the questionnaire) limits analysis of results. Item nonresponse rates and frequency of self-reported, sensitive sexual behaviors were statistically significantly different across the three methods for all reported questions. The IVR estimates were more similar to the remote (postal) SAQ than to the in-person SAQ. The postal mode reported socially desirable answers more frequently than the other two modes (both in-person and self-administered) [18]. This study's authors suppose that subjective psychological

responses, such as the perception of confidentiality, explain the lower report of undesirable behaviors in the self-administered in-person article survey compared with the other two modes.

Table 4. In-person interviewer administered compared with remote self-administered.

Author (year)	Country (sample type) ^a	Data collection #1 (sample size)	Data collection #2 (sample size)
Ballivian [14] (2013)	Honduras (dependent)	FTF ^b (1500)	SMS ^c (900)
Ballivian [14] (2013)	Honduras (dependent)	FTF (1500)	IVR ^d (600)
Ballivian [14] (2013)	Peru (dependent)	FTF (1500)	SMS (677)
Ballivian [14] (2013)	Peru (dependent)	FTF (1500)	IVR (383)
Du [15] (2013)	China (dependent)	FTF (591)	SMS (591)
Li [19] (2013)	China (dependent)	FTF (177)	SMS (99)

^aWhen participants were the same across modes, we classified the sample as dependent. Different participants across modes was labeled as an independent sample.

^bFTF: face-to-face.

^cSMS: short message service.

^dIVR: interactive voice response.

Table 5. In-person self-administered compared with remote self-administered.

Author (year)	Country (sample type) ^a	Data collection #1 (n)	Data collection #2 (n)
Lau [18] (2000)	Hong Kong (dependent)	In-person IVR ^b (1254)	Postal SAQ ^c (556)
		In-person SAQ (606)	Postal SAQ (556)

^aWhen participants were the same across modes, we classified the sample as dependent. Different participants across modes were labeled as an independent sample.

^bIVR: interactive voice response.

^cSAQ: self-administered questionnaire.

Table 6. In-person self-administered compared with a second mode of in-person self-administered.

Author (year)	Country (sample type) ^a	Data collection #1 (sample size)	Data collection #2 (sample size)
Lau [18] (2000)	Hong Kong (dependent)	In-person SAQ ^b (606)	In-person IVR (1254)

^aWhen participants were the same across modes, we classified the sample as dependent. Different participants across modes were labeled as an independent sample.

^bSAQ: self-administered questionnaire.

Remote Self-Administered Compared With a Second Mode of Remote Self-Administered

Four articles make five comparisons of remote self-administered modes with a second mode of remote self-administered (see [Table 7](#)). Two articles compare Russian respondents' answers on a Web survey on a mobile phone with a Web survey on a PC. The survey that compared a different population across the two modes did not find any difference in report of sensitive behavior indicators [26]. In the second study that was a cross-over experiment, 2 of the 5 sensitive questions, namely, alcohol consumption and income, were statistically significantly different [25]. The PC-based Web survey reported higher levels of alcohol consumption and higher income. The respondents reported higher trust in data confidentiality on the Web data collected on PC compared with the Web data collected on mobile phone. The authors also tested for interaction between gender and survey mode but did not find any statistically significant gender differences.

The World Bank study also compared SMS and IVR. The study found a higher attrition and lower survey-completion rate among IVR and SMS respondents compared with the other two modes [14]. Furthermore, panelists responding via self-administered mode were more likely to leave questions unanswered compared with CATI. Finally, SMS then IVR were estimated to be the least expensive options for data collection, compared with CATI and FTF.

In addition to the work in Peru, Honduras, and Russia, 1 article compared two modes of remote data collection. This survey was nested in a panel study in Nepal and compared CATI, and two types of SMS surveys. During single-sitting SMS interviews participants completed the survey at one time, and during module-design text interviews participants answered one

question per day [24]. There were very few differences in results when comparing the two SMS modes. The modular survey did have higher nonresponse rate than the single sitting SMS survey, but the respondents in the modular design group found the survey to be significantly easier to complete than persons in the other two groups [24].

Remote Self-Administered Compared With Remote Interviewer-Administered

The research in Nepal compared the two aforementioned modes of SMS to CATI (see [Table 8](#)). They found that both text message modes increased the probability of disclosing sensitive information (eg, age of drinking onset, ever-smoking marijuana) compared with CATI, but mode did not impact the report of factual survey items (eg, marital status, age) [24]. The authors note that they are not sure whether sensitive behavior is reported more frequently via text due to decreased time pressure or increased privacy [24].

The final comparisons from the World Bank study is CATI compared with IVR and SMS. SMS, although the least expensive of the three modes, has twice the attrition rates as CATI [14]. Another important consideration about SMS from the World Bank study is that personal Internet access was reported more frequently via SMS. The World Bank study's authors hypothesize that this could be caused by younger informants who are more likely to respond to an SMS survey compared with other ages. Reliability co-efficients for SMS range from .57 (Do you consider yourself poor?) to .87 (Do you currently have a television at home?) [14]. The Cronbach alpha for IVR is higher (at .86) than SMS, and item-level reliability has a smaller range from .79 (In the last 30 days have you access the Internet or not?) to .93 (Do you currently have a television at home?) [14].

Table 7. Remote self-administered compared with a second mode of remote self-administered.

Author (year)	Country (sample type) ^a	Data collection #1 (sample size)	Data collection #2 (sample size)
Mavletova [26] (2013)	Russia (independent)	Web MP ^b (481)	Web PC ^c (532)
Mavletova and Couper [25] (2013)	Russia (dependent)	Web MP (884)	Web PC (884)
Ballivian [14] (2013)	Peru (dependent)	IVR ^d (383)	SMS ^e (677)
Ballivian [14] (2013)	Honduras (dependent)	IVR (600)	SMS (900)
West [24] (2015)	Nepal (independent)	SMS: singular (150)	SMS: modular (150)

^aWhen participants were the same across modes, we classified the sample as dependent. Different participants across modes was labeled as an independent sample.

^bWeb PC: Web on personal computer.

^cWeb MP: Web on mobile phone.

^dIVR: interactive voice response.

^eSMS: short message service.

Table 8. Remote self-administered compared with remote interviewer-administered.

Author (year)	Country (sample type) ^a	Data collection #1 (sample size)	Data collection #2 (sample size)
West [24] (2015)	Nepal (independent)	CATI ^b (150)	SMS ^c modular (150)
West [24] (2015)	Nepal (independent)	CATI (150)	SMS singular (150)
Ballivian [14] (2013)	Peru (dependent)	CATI (384)	SMS (677)
Ballivian [14] (2013)	Honduras (dependent)	CATI (600)	SMS (900)
Ballivian [14] (2013)	Peru (dependent)	CATI (384)	IVR ^d (383)
Ballivian [14] (2013)	Honduras (dependent)	CATI (600)	IVR (600)

^aWhen participants were the same across modes, we classified the sample as dependent. Different participants across modes was labeled as an independent sample.

^bCATI: computer-assisted telephone interview.

^cSMS: short message service.

^dIVR: interactive voice response.

Discussion

Principal Findings

This literature review synthesizes research that compares two or more modes of data collection in LMICs, with a special focus on remote data collection. We identified 10 articles that covered a range of modes and a variety of sampling methods. Three articles collected data in East Asia, 3 in Central and South America, 2 in Russia, 1 in South Asia, and 1 in Lebanon. No articles were identified that compared in-person self-administered and remote interviewer-administered surveys (eg, CATI). In-person self-administered includes a SAQ or a computer-assisted self-interview, with or without audio. Due to both heterogeneity of outcomes among the studies and the limited number of studies, this literature review is best positioned to present the current evidence rather than attempt to draw conclusions.

The most comprehensive study was conducted by the World Bank, where researchers made six comparisons across four modes of data collection (FTF vs IVR, FTF vs CATI, FTF vs SMS, IVR vs CATI, IVR vs SMS, and CATI vs SMS) in each of two countries [14]. This is the only study in the literature review to compare IVR versus CATI and IVR versus SMS. Comparing IVR and CATI is a particularly useful comparison because if the same sampling method is used for CATI and IVR, the impact of administration of interview can be better isolated and assessed. The finding that CATI and FTF estimates produced the best criterion validity when compared with the other modes in the study is consistent with findings from other studies included in this review.

Half of the studies enrolled participants FTF. Enrolling participants FTF mitigates one of the main benefit of remote

data collection—reduced data collection cost. Only two of the studies used random digit dialing (RDD) and both were limited to landlines [16,17]. No articles in this review were conducted in SSA, but we expect an increasing amount of evidence will be emerging from the area.

A minority of articles explicitly compared the profile of respondents between the two modes of data collection. By comparing sample demographics to the target population, we will better understand the respondent bias a mode may introduce. Ferreira et al found that groups with higher telephone ownership or coverage were more likely to report better health [16] and the World Bank study identified young people as more likely to respond to a SMS survey [14]. Other key information, including cost, length of the questionnaire, and the reliability of measures were not reported consistently but would provide important implementation information. It is imperative to use American Association for Public Opinion Research Reporting Guidelines so that survey metrics are comparable [27].

The impact of mode on reporting sensitive behaviors in LMIC is discrepant [12]. A meta-analysis that compared 15 data sets (which included no forms of remote data collection), mostly comparing FTF and audio computer-assisted self-interview found that non-FTF methods did not consistently produce a significant increase in the reporting of 4 sensitive questions [12]. In this literature review, only 3 articles offered a straight-forward comparison of nondesirable behaviors. In 2 of the articles, remote data collection elicited higher report of nondesirable behaviors compared with in-person data collection [18,20]. The article that compared CATI, single-sitting and modular-design SMS found that the SMS respondents reported more socially undesirable behaviors compared to CATI [24].

Considering that all but one article in this literature review had been published in the past 5 years, we anticipate an increase in publications comparing modes of data collection in the coming years. As evidence continues to emerge, research designed to isolate the cause of differences in measures between modes should be a priority. For example, researchers should ask direct questions around the impact of increased privacy, greater anonymity, or greater convenience of remote data collection. Only 3 of the articles in this literature review included such questions [24,25,26]. By eliciting participant's opinions about the different modes, discrepancies can be better explained. Country context, such as literacy levels, mobile phone ownership, and network coverage are particularly important to note when considering remote data collection in LMICs. Therefore researchers should include aforementioned information in publications so that conditions can be considered. Furthermore, reporting factual, sensitive, or perceptual questions, as well listing the type of question (such as multiple choice, numeric, text) all provides pertinent information for decision making.

Although studying mode effect is an important aspect of remote data collection research, sampling is equally pertinent. RDD functions without a sampling frame in many countries, which means after identifying all mobile network operator prefixes in a country, numbers are randomly generated [28]. It is unknown whether RDD can consistently produce nationally representative estimates of a health outcome or which mode is best suited for RDD. Six of the 10 studies in this literature reviewed enrolled participants in-person. The advantage of FTF enrollment is that the research team has a reference standard against which the remote data collection tool is measured. When enrolling participants for FTF, asking how many mobile phone numbers each participant has and their estimated network coverage helps to estimate how representative the sample will be. Finding out the preferred language of survey while enrolling a participant will allow the first follow-up contact to be in the respondent's language of choice, negating the need for a language question and likely increasing the response rate. It is unknown whether RDD is more likely to enroll respondents who are hard to reach in FTF surveys, but this hypothesized advantage should be assessed. RDD and remote data collection generally have the

advantage of faster collection of data than a FTF survey, thus making this approach particularly useful during a crisis.

To isolate a superior method of data collection, studies that compare more than two modes of remote data collection are preferred. Only three studies in this review compared more than two modes [14,18,24]. Specifically, future research should follow mHealth guidelines [29], incorporating a factorial design where possible. As a minimum, it would be advantageous for authors to identify which questions are sensitive in their context so that mode effects for sensitive questions can be compared, even if the subject matter is different.

Limitations

We note three main limitations to this literature review. First, the small number of studies (n=10) that compared two modes of data collection or more (n=25 comparisons), made it difficult to draw conclusions. The included research used a wide variety of data collection modes and sampling techniques and covered a plethora of topics and populations, thus negating the ability to make conclusions. Furthermore, the nonlinear relationship of effects can make pattern identification a challenge [11]. Second, owing to the inherent limitations of searching for grey literature, our search strategy could have missed important articles. A third limitation pertains to the inconsistent reporting of key survey metrics as well as lack of a formal statistical test to analyze the variation between the results of each article. Regardless of these limitations, this literature review contributes to efforts to characterize current evidence on the effect of remote data collection mode on data quality in LMICs.

Conclusions

Due to the nascent state of remote data collection in LMICs, several research areas merit further investigation. The advantages of remote data collection are presented, but a superior mode for a population has yet to be established due to a dearth of evidence. We encourage randomized control trials with multiple arms to identify a mode appropriate for the context. Ultimately, researchers must balance the desire for more efficient, cost-effective data collection methods with study aims and the limitations of a novel mode of data collection.

Acknowledgments

The studies described as part of the research agenda are funded by the Bloomberg Philanthropies. This funding agency had no role in the preparation of this manuscript. We would like to thank Victoria Gammino for a careful review of the manuscript.

Conflicts of Interest

None declared.

References

1. Okonjo-Iweala N, Osafo-Kwaako P. Improving health statistics in Africa. *Lancet* 2007 Nov 03;370(9598):1527-1528. [doi: [10.1016/S0140-6736\(07\)61644-4](https://doi.org/10.1016/S0140-6736(07)61644-4)] [Medline: [17980722](https://pubmed.ncbi.nlm.nih.gov/17980722/)]
2. Hyder A, Wosu A, Gibson D, Labrique A, Ali J, Pariyo G. Noncommunicable disease risk factors and mobile phones: a proposed research agenda. *J Med Internet Res* 2017;19(5):e133 [FREE Full text] [doi: [10.2196/jmir.7246](https://doi.org/10.2196/jmir.7246)]
3. UN. Goal 3 Ensure healthy lives and promote well-being for all at all ages 2016 URL: <http://www.un.org/sustainabledevelopment/health/> [accessed 2017-04-19] [WebCite Cache ID 6pqltShQG]

4. Poushter J, Oates R. Pew Research Center. 2015. Cell phones in Africa: communication lifeline URL: <http://www.pewglobal.org/2015/04/15/cell-phones-in-africa-communication-lifeline/> [accessed 2017-04-19] [WebCite Cache ID 6pqnC5jhS]
5. World Bank. Mobile cellular subscriptions (per 100 people) URL: <http://data.worldbank.org/indicator/IT.CEL.SETS.P2> [accessed 2017-04-19] [WebCite Cache ID 6pqnLVdpq]
6. Gibson D, Pereira A, Farrenkopf B, Pariyo G, Labrique A, Hyder A. Mobile phone surveys for collecting population-level estimates in low- and middle-income countries: a literature review. *J Med Internet Res* 2017;19(5):e139 [FREE Full text] [doi: [10.2196/jmir.7428](https://doi.org/10.2196/jmir.7428)]
7. Corkrey R, Parkinson L. Interactive voice response: review of studies 1989-2000. *Behav Res Methods Instrum Comput* 2002 Aug;34(3):342-353. [Medline: [12395550](https://pubmed.ncbi.nlm.nih.gov/12395550/)]
8. Kempf A, Remington P. New challenges for telephone survey research in the twenty-first century. *Annu Rev Public Health* 2007;28:113-126. [doi: [10.1146/annurev.publhealth.28.021406.144059](https://doi.org/10.1146/annurev.publhealth.28.021406.144059)] [Medline: [17094769](https://pubmed.ncbi.nlm.nih.gov/17094769/)]
9. Cannell C, Miller P, Oksenberg L. Research on interviewing techniques. *Sociol. Methodol* 1981;12:389-437. [doi: [10.2307/270748](https://doi.org/10.2307/270748)]
10. Groves R. Three eras of survey research. *Public Opin Q* 2011 Dec 15;75(5):861-871. [doi: [10.1093/poq/nfr057](https://doi.org/10.1093/poq/nfr057)]
11. Jackle A, Roberts C, Lynn P. Assessing the effect of data collection mode on measurement. *Int Stat Rev* 2010;78(1):3-20. [doi: [10.1111/j.1751-5823.2010.00102.x](https://doi.org/10.1111/j.1751-5823.2010.00102.x)]
12. Phillips A, Gomez G, Boily M, Garnett G. A systematic review and meta-analysis of quantitative interviewing tools to investigate self-reported HIV and STI associated behaviours in low- and middle-income countries. *Int J Epidemiol* 2010;39(6):1541-1555. [doi: [10.1093/ije/dyq114](https://doi.org/10.1093/ije/dyq114)]
13. Tourangeau R, Smith T. Asking sensitive questions: the impact of data collection mode, question format, and question context. *Public Opin Q* 1996;60(2):275-304. [doi: [10.1086/297751](https://doi.org/10.1086/297751)]
14. Ballivian A, Azevedo J, Durbin W, Rios J, Godoy J, Borisova C. Using mobile phones for high-frequency data collection. In: Toninelli D, Pinter R, de Pedraza P, editors. *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies*. London: Ubiquity Press; 2015:21-39.
15. Du X, Wang W, van Velthoven MH, Chen L, Scherpbier RW, Zhang Y, et al. mHealth series: text messaging data collection of infant and young child feeding practice in rural China - a feasibility study. *J Glob Health* 2013 Dec;3(2):020403 [FREE Full text] [doi: [10.7189/jogh.03.020403](https://doi.org/10.7189/jogh.03.020403)] [Medline: [24363921](https://pubmed.ncbi.nlm.nih.gov/24363921/)]
16. Ferreira A, César CC, Malta DC, Souza Andrade AC, Ramos CG, Proietti FA, et al. Validade de estimativas obtidas por inquérito telefônico: comparação entre VIGITEL 2008 e inquérito Saúde em Beagá. *Rev bras epidemiol* 2011 Sep;14(1):16-30. [doi: [10.1590/S1415-790X2011000500003](https://doi.org/10.1590/S1415-790X2011000500003)]
17. Francisco P, Azevedo Barros MB, Segri N, Alves MC, Cesar C, Malta D. Comparação de estimativas para o auto-relato de condições crônicas entre inquérito domiciliar e telefônico - Campinas (SP), Brasil. *Rev bras epidemiol* 2011 Sep;14(3):5-15. [doi: [10.1590/S1415-790X2011000500002](https://doi.org/10.1590/S1415-790X2011000500002)]
18. Lau JTF, Thomas J, Liu JLY. Mobile phone and interactive computer interviewing to measure HIV-related risk behaviors: the impacts of data collection methods on research results. *AIDS* 2000 Jun 16;14(9):1277-1280.
19. Li Y, Wang W, van Velthoven MH, Chen L, Car J, Rudan I, et al. Text messaging data collection for monitoring an infant feeding intervention program in rural China: feasibility study. *J Med Internet Res* 2013 Dec 04;15(12):e269 [FREE Full text] [doi: [10.2196/jmir.2906](https://doi.org/10.2196/jmir.2906)] [Medline: [24305514](https://pubmed.ncbi.nlm.nih.gov/24305514/)]
20. Mahfoud Z, Ghandour L, Ghandour B, Mokdad A, Sibai A. Cell phone and face-to-face interview responses in population-based surveys: how do they compare? *Field Methods* 2014 Jul 01;27(1):39-54. [doi: [10.1177/1525822X14540084](https://doi.org/10.1177/1525822X14540084)]
21. van Velthoven MH, Li Y, Wang W, Du X, Wu Q, Che N. mHealth project in Zhao County, rural China - Description of objectives, field site and methods. *J Glob Health* 2013;3(2):020401. [doi: [10.7189/jogh.03.020401](https://doi.org/10.7189/jogh.03.020401)]
22. Abbott SA, Friedland BA, Sarna A, Katzen LL, Rawiel U, Srikrishnan AK, et al. An evaluation of methods to improve the reporting of adherence in a placebo gel trial in Andhra Pradesh, India. *AIDS Behav* 2013 Jul;17(6):2222-2236 [FREE Full text] [doi: [10.1007/s10461-012-0402-z](https://doi.org/10.1007/s10461-012-0402-z)] [Medline: [23299877](https://pubmed.ncbi.nlm.nih.gov/23299877/)]
23. Moura E, Claro RM, Bernal R, Ribeiro J, Malta D, Morais Neto O. A feasibility study of cell phone and landline phone interviews for monitoring of risk and protection factors for chronic diseases in Brazil. *Cad Saúde Pública* 2011 Feb;27(2):277-286. [doi: [10.1590/S0102-311X2011000200009](https://doi.org/10.1590/S0102-311X2011000200009)]
24. West BT, Ghimire D, Axinn WG. Evaluating a modular design approach to collecting survey data using text messages. *Surv Res Methods* 2015;9(2):111-123 [FREE Full text] [Medline: [26322137](https://pubmed.ncbi.nlm.nih.gov/26322137/)]
25. Mavletova A, Couper M. Sensitive topics in PC web and mobile web surveys: is there a difference? *Surv Res Meth* 2013;7(3):191-205. [doi: [10.18148/srm/2013.v7i3.5458](https://doi.org/10.18148/srm/2013.v7i3.5458)]
26. Mavletova A. Data quality in PC and mobile web surveys. *Soc Sci Comput Rev* 2013;31(6):19.
27. The American Association for Public Opinion Research.: Washington, DC; 2016. Standard definitions: final dispositions of case codes and outcome rates for surveys URL: http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf [accessed 2017-04-19] [WebCite Cache ID 6pr64foLk]
28. Gibson D, Pariyo G, Wosu A, Greenleaf A, Ali J, Ahmed S, et al. Evaluation of mechanisms to improve performance of mobile phone surveys: a research protocol. *JMIR Res Protoc* 2017;6(5) [FREE Full text] [doi: [10.2196/resprot.7534](https://doi.org/10.2196/resprot.7534)]

29. Agarwal S, LeFevre A, Lee J, L'Engle K, Mehl G, Sinha C, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. *BMJ* 2016 Mar 17;352:i1174. [doi: [10.1136/bmj.i1174](https://doi.org/10.1136/bmj.i1174)]

Abbreviations

CATI: Computer-assisted telephone interview
FTF: face-to-face
HIV: Human Immunodeficiency Virus
IVR: interactive voice response
LMIC: low- and middle-income countries
PC: personal computer
RDD: random digit dialing
SAQ: self-administered questionnaire
SMS: short message service
SSA: Sub-Saharan Africa

Edited by E Rosskam, A Hyder; submitted 16.01.17; peer-reviewed by J de Jong, TR Soron; comments to author 28.02.17; revised version received 22.03.17; accepted 23.03.17; published 05.05.17

Please cite as:

Greenleaf AR, Gibson DG, Khattar C, Labrique AB, Pariyo GW

Building the Evidence Base for Remote Data Collection in Low- and Middle-Income Countries: Comparing Reliability and Accuracy Across Survey Modalities

J Med Internet Res 2017;19(5):e140

URL: <http://www.jmir.org/2017/5/e140/>

doi: [10.2196/jmir.7331](https://doi.org/10.2196/jmir.7331)

PMID: [28476728](https://pubmed.ncbi.nlm.nih.gov/28476728/)

©Abigail R Greenleaf, Dustin G Gibson, Christelle Khattar, Alain B Labrique, George W Pariyo. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 05.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.