UNIVERSITY OF CAPE TOWN

CENTRE FOR

SOCIAL SCIENCE RESEARCH

A rapid review of research methodologies for remote research among young people in sub-Saharan Africa

Yasmin Dunkley, Nontokozo Langwenya, Samantha Malunga, Jane Kelly, Elona Toska

CSSR Working Paper No. 463

May 2021

Published by the Centre for Social Science Research
University of Cape Town
2021

http://www.cssr.uct.ac.za

This Working Paper can be downloaded from:
http://cssr.uct.ac.za/pub/wp/463

ISBN: 978-1-77011-451-7

About the authors:

Yasmin Dunkley is a consultant in Public Health, specialising in HIV prevention. Her recent roles include building the HIV technical capacity and strategic strength of UK, and West and Southern African community based organisations, and local health departments.

Nontokozo Langwenya is a co-Investigator for the HEY BABY study and Young Carers research of the UKRI GCRF Accelerating Achievement for Africa's Adolescents Hub.

Samantha Malunga is an honorary research associate at the AIDS and Society Research Unit in the Centre for Social Science Research at the University of Cape Town.

Jane Kelly is a research officer in the Centre for Social Science Research at the University of Cape Town and is a member of the UCT team of the UKRI GCRF Accelerating Achievement for Africa's Adolescents Hub.

Elona Toska is a researcher in adolescent health at the Centre for Social Science Research and an Associate Lecturer at the Department of Sociology, University of Cape Town. She is a co-Principal Investigator of the Mzantsi Wakho and HEY BABY studies and leads the UCT team of the UKRI GCRF Accelerating Achievement for Africa's Adolescents Hub.

A rapid review of research methodologies for remote research among young people in sub-Saharan Africa

The COVID-19 pandemic represents a new opportunity for survey research teams to design, pilot and implement remote survey modalities. The Helping Empower Youth Brought Up in Adversity with their Babies and Young Children (HEY BABY) study planned to collect follow-up data in 2020 to follow up on 1,046 adolescent mothers and their children. While baseline data collection involved a combination of face-to-face interviews supported by audio computer-assisted self-interviewing (ACASI), follow-up data collection needed to happen remotely due to mobility restrictions and pandemic control regulations. To inform this methodological shift, a rigorous literature review was conducted in 2020. Despite a dearth of literature assessing remote survey modalities with young people in South Africa, this review – based on general concept papers, case studies across low- and middle-income countries and think-pieces – demonstrates that remote survey methodologies may be both feasible and acceptable in Africa. Hybrid models, based on computer assisted technology interviewing *(CATI), with* Short Message Service *(SMS) follow up and higher frequency data collection, may be acceptable alternatives for field-research compared to face-to-face models that are currently unavailable. Question topics will need to be screened to ensure they do not distress participants. CATI models of remote data collection should average 30 minutes; however, frequent check-ins are acceptable in a South African context. Incentivisation and the use of different survey modes will encourage survey response uptake, but care must be taken to ensure the new remote data methodology questions are as alike as possible to the face-to-face baseline. Intensive tracking and subgroup weighting can address the effect of not including individuals without mobile phones in the study. Data analysis and results interpretation will have to factor in mode effects of a mixed modality research study.*

1. Background

HEY BABY – Helping Empower Youth Brought up in Adversity with their Babies and Young Children – is a research project with the overarching goal of understanding what can promote resilience for adolescent parents and their children in South Africa, and thus mitigate intergenerational cycles of adversity (https://www.heybaby.org.za/). In 2018-2019, the study team conducted in-person data collection with 1,046 adolescent mothers and their children. In January 2020, the HEY BABY team was ready to start re-enrolling research participants in the first study follow-up data collection, with questionnaires piloted and team training starting in February-March 2020. However, in March 2020, the World Health Organization declared COVID-19 a global pandemic and, in the absence of a vaccine, urged countries to adopt social distancing strategies to mitigate the spread of the virus (World Health Organisation, 2020). Due to the social distancing and lockdown measures in South Africa, the scheduled follow-up data collection for HEY BABY had to change modality.

The power of research studies like the HEY BABY study is the ability to document changes over time. These changes become all the more important in the context of the COVID-19 pandemic given that the pandemic can – and is – impacting all facets of life, including people's resilience. Many of the topics covered in the HEY BABY study are of a sensitive nature (including, for example, experiences of violence, and sexual and reproductive health), and participants include young parents and young people living with human immunodeficiency virus (HIV). This led to the question: How can the modality of data collection in the HEY BABY study be changed while still capturing these important topics that could be exacerbated during a pandemic? In order to answer this question, we undertook the current review.

This review considers the practicalities and feasibility of carrying out the follow-up of research participants via remote data collection methods with young people in South Africa. It focuses on current academic literature and recent think-pieces from the research sector to look at the ethics of carrying out remote research, the different remote research modalities available, data quality, response and attrition rates, survey length, and the implications of each survey modality choice.

In essence, there were two overarching questions we sought to answer in undertaking this review:

- How do you do remote research within South Africa and – within Sub-Saharan Africa (SSA) more broadly – with young people, because you have chosen to, not because of a pandemic?
- How do you do remote research within SSA with young people, having been forced to, because of a pandemic (e.g., by potentially looking at work in settings with Ebola)?

Two related sub-questions were:

- What is the duration of the interviews that achieves the highest response rates?
- What platforms are the research studies using/ what is the model being used to conduct the interview?

In addition, we were interested in[1]:

- maintaining high response rates (90+%) for interviews;
- experiences from SSA; and
- evidence from research or engagement with adolescent mothers (13-24 year olds).

2. Search strategy and considerations

The global nature of the COVID-19 pandemic means that many research projects are switching to remote data methodologies. There is a large amount of conversation and discussion taking place at universities and within field teams. Think-pieces on this topic, although not peer-reviewed, have been included in this review to provide further clarity on issues as they arise. In turn, the rapidly increasing amount of literature on this topic means that this review may need to be updated again before it is used to justify evidence-based decisions for changing fieldwork practice.

The first search was restricted to young mothers and young women in SSA and focused on remote data collection modalities for topics with similar sensitivity as topics covered in the HEY BABY study (including resilience, reproductive health and victimisation). However, this search yielded very few results and included only thought pieces (Mercer, Lau & Kennedy, 2018; Cullen & Mahmud, 2020; Innovation for Poverty Action [IPA], 2020; Carranza et al., 2020; Peterman et al., 2020; Rogers, 2020).

[1]It is important to note that at present, UCT has three reports that are looking at interviews during COVID-19.

The search was thus widened as follows: "(young people or youths or teen or teenager or teenage or young adult or adolescents or adolescent) AND (remote data collection or CASIC or CATI or IVR or SMS or USSD or WAP or MIMI[2]) AND Africa". Filters were applied for age (13-24 years). This generated 304 articles; abstracts were screened, and no literature was of relevance. The phrase "young people" was removed and this still did not yield relevant results. When "Africa" was removed from the search string some literature surfaced, which is discussed in this paper. In light of how rapidly research in remote data collection has been progressing, grey literature sourced through the HEY BABY team and shared via networks such as Sexual Violence Research Initiative (SVRI) and the Abdul Latif Jameel Poverty Action Lab Africa (J-PAL Africa) are also included in this paper.

While there is literature on remote research methodologies, not all of it is of relevance to the HEY BABY project. Issues emerged around the practicality of the modalities in our context – for example, the distribution of phones to all participants. Not all the ways of conducting remote research were practical for the HEY BABY study; for example, high frequency panel studies normally distribute phones to their participants as part of their research.

PubMed was queried, but of the 304 results, there was a lack of literature on remote collection methodologies with young people in SSA. As a result, this review was expanded to include general concept papers on conducting remote research, as well as practical implementation guides specific to low- and middle-income countries (LMIC), and case examples of remote data collection studies conducted with adults in LMIC. Papers where the focus was on the study outcomes alone, rather than the data collection method, were excluded.

Using recently published concept papers on switching to remote data collection methods, the following key terms were found for the search question: "(young people or youths or teen or teenager or young adults or adolescents or adolescent or teenager or teenage) AND (remote data collection or CASIC or CATI or IVR or SMS or USSD or WAP or MIMI) AND Africa". Filters were applied for African and European languages, and age. In this review 33 papers and think-pieces have been reviewed. Twenty-one were abstract screened as not fitting the review's aims.

[2] See Table 2 for explanation of each of these abbreviations.

3. Review structure and limitations

This paper starts by reviewing the literature around the ethics of switching to remote data collection methodologies. It then presents different remote methods and makes some high-level comparisons. Case examples of different methodologies used in LMIC – including SSA – are expanded in the sections that follow. Attrition and non-response effects are discussed, as well as mode effects, including a commentary around typology. Each section starts with an introduction into the section theme, then has a table of articles on the topic, a further elaboration on the articles, and then ends with a commentary around the implications for the HEY BABY study.

The research summarised in this review did not screen included papers by their individual topic areas (for example, employment status, livelihood, HIV and sexual health behaviours, intimate partner violence). There are, however, clear implications for survey mode choice with topic area, as some questions may not be acceptable to be asked remotely. Once the decision has been made on which modality the research team will explore further, additional research will be required to understand how to best switch the baseline questionnaire to a new survey mode while controlling for mode effects with question structure and typological considerations.

We acknowledge that changes in methodologies result in changes in interpretation or bias, so this search was carried out to determine the following:

- Is a change practical?
- What do we need to be aware of?
- What could influence the interpretation of our data?
- How can we best mitigate around these known potential confounders?

4. The ethics of changing data collection methodologies given the sensitivity of the research topics

When considering whether to change data collection methods – particularly when the information being requested of the participants is sensitive – there is a need to consider (i) how best to minimise distress for the participants, (ii) methodological and technical considerations around conducting the research remotely, and (iii) what modalities are best to use to collect the sensitive information, such as violence against women and girls (VAWG).

Table 1: Methodological and technical considerations and methods to mitigate distress for research on VAWG

Study	Methods to mitigate distress	Methodological and technical considerations for remote research
Peterman, Bhatia & Guedes, 2020	Special selection and training of interviewers Ensuring that there is robust attention to confidentiality and referral mechanisms Eligibility requirements to participate in research on VAWG ensuring that participants have access to safe devices to ensure confidentiality Confidentiality with survey platforms	Data accuracy Question design must be simplified for telephonic conversations Sampling must be appropriate for remote data collection Internet administered surveys could include visual components
Rogers, 2020	Carry out a risk assessment for data collection and question choice: Do you need to collect this data? Have you avoided adding questions about violence to socioeconomic assessments? Could the data be collected at a later stage?	Implement graduated informed consent – the initial person who answers the phone is provided general non-specific information about the survey topic. The specific topics of the survey are only revealed to the individual respondent selected to participate in the survey
IPA, 2020	Piloting a survey tool is critical to identify any potential adverse effects of the survey, noting how respondents react to sensitive questions Piloting will also allow space for ethical reflection Piloting can be used to refine enumerator scripts	Non-interviewer-facilitated options like computer-assisted online surveys, e.g. SurveyMonkey, or mobile instant messaging interviews, e.g. via WhatsApp Pilot questions through a small-scale survey pilot, which should be followed by a critical reflection period. For example, if you plan a survey with 2,000 respondents, first pilot the survey with 50 to 100 as a general best practice

Much of the literature and recent think-pieces around switching to remote data collection methodologies forces researchers to ask the question, 'can this data be carried out remotely, while maintaining the same ethical standards'? The research can be done, as long as the methods to mitigate distress are robust, including taking into consideration who the participant may be sharing a space with when data is being collected, being able to pick up on distress cues, and having interventions in place to ensure the participant is not re-traumatised in the process of data collection (Peterman et al., 2020).

As noted in Table 1, conducting remote research has associative methodological and technical considerations, namely, data accuracy and question design (Peterman et al.,2020). In order to carry out remote research with a sensitive topic, assessment for the data collection and question choice should be conducted (Rogers, 2020). The ability to collect data does not necessitate that data should be collected. Data should directly contribute to improving the wellbeing of women who take part in the survey, immediately and beyond by putting women first (Peterman et al.,2020).

4.1 Implications for the HEY BABY study

Ethical considerations were taken into account and questions in the papers that were found were compared to the HEY BABY questionnaires on topics that focus on violence, intimate partner violence, sexual assault, sexual abuse, and antiretroviral therapy (ART) medication among other such sensitive topics. Consideration was given to implications this had on the follow up of the research participants – would there be people who were less likely to participate based on the response options that were being given? Research demonstrated different characteristics of those that were more likely to be lost to follow up during remote data collection. Mobile phone modes over-represent young, urban and well-educated respondents. Conversely, it is more challenging to retain participants who are rural, or who migrate, although the urban bias isn't always the case, as the Dillon study (2012) in Tanzania demonstrates. In some of the papers reviewed, researchers showed the type of group that was lost to follow up during remote data collection.

5. What remote options are available?

Current remote survey methodologies that are available are summarised in Table 2.

Table 2: Current remote survey methodologies

Abbre-viation	Full title	Brief description	References
CATI	Computer Assisted Technology Interviewing	Enumerators in a call centre (or at home) call respondents, ask the relevant questions and enter the responses into a database using a CATI system	Dillon, 2012; Himelein & Kastelic, 2014; Pariyo et al., 2019; Garlick et al., 2020
SMS	Short Message Service - Text message based surveys	This is a text messaging service component of mobile device systems	Berman et al., 2017; Feng et al., 2018; Lau et al., 2018; Owsley et al., 2020
IVR	Interactive Voice Response	A technology that allows humans to interact with a computer-operated phone system through the use of voice and dual-tone multi-frequency signalling (DTMF) tones input via a keypad	Croke et al., 2012; Gallup, 2012; L'Engle et al., 2018; Lau et al., 2019; Pariyo et al., 2019
WAP	Web Assisted Platforms	These are web-based mobile phone surveys, which are best suited for high-end phones with internet capability	Croke et al., 2012; IPA, 2020
USSD	Unstructured Supplementary Service Data	Allows direct transmission of questions from a phone company server to the respondent's phone	Croke et al., 2012
MIMIs	Mobile Instant Messaging Interviews	Interviews conducted via instant mobile text messaging with other users on their smartphones, in real time as well as asynchronously	*this was not covered in this paper
CASIC	Computer-assisted online surveys	An internet surveying technique in which the interviewee follows a script provided in a website	*this was not covered in this paper
ACASI	Audio-computer assisted self-interviewing	Questions are pre-recorded in the local language, and tell respondents to select their answer by touching the corresponding button on the screen of the device (e.g., a smartphone, computer or tablet) that they are given	Van de Wijgert et al. 2000; Langhaug et al., 2010; Beauclair et al., 2013; Brown et al. 2013; Cullen & Mahmud, 2020

6. Contexts in which remote data collection methods were used

This section will explore recommendations found in (i) interrupted studies that pivoted to remote data collection, and (ii) studies that had remote data collection methods as part of the original study design, to see if there were different conclusions around modality.

6.1 Pivoting to remote data collection methods when a study is interrupted

CATI, IVR and SMS are data collection modalities suitable for use on a mobile phone, and lend themselves well to follow-up on an existing baseline survey in LMIC (Himelein et al., 2020). Web surveys completed by mobile devices and requiring smartphones are typically out of reach of many poor households. Non-responders to CATI can receive a text message describing an incentive payment, as was the case in the Himelein et al. (2020) study where participants received one US dollar in phone credit for completed interviews. Himelein et al. (2020) concluded that the ideal situation for CATI is one where researchers have an existing sample frame, with phone contact information, and want to follow up to see whether the intervention affected responses during the COVID-19 pandemic, or to measure other outcomes.

Table 3: CATI survey length and response rates

Study	Survey length	Response rate
Himelein et al., 2020	20 - 25 minutes	30%, and all numbers were attempted nine times over a 14 day period The existence of baseline data on individuals, along with some pre-existing relationship through past surveying, may increase response rates
IPA, 2020	20 minutes for best results. No longer than an hour	CATI response rate is faster than IVR and SMS whose response rates are slow – only use for longer surveys when quantity and speed are valued over quality

Innovation for Poverty Action (IPA) recommends CATI for remote data collection on a SurveyCTO platform due to the quality of data that can be

collected using this method (IPA, 2020). IPA notes that although IVR, SMS and WAP are more scalable than CATI, and less expensive, the responses collected are less representative and less likely to be honest than when CATI is used. Thus, for the best results one could consider shorter, more frequent surveys. The CATI survey length and response rates from the Himelein et al. (2020) and IPA (2020) papers are summarised in Table 3.

6.1.1 Implications for HEY BABY

The HEY BABY researchers suggested the best way forward was to try to use like-for-like survey methods for follow up from baseline – CATI is demonstrated as best for an interrupted study when a face-to-face (FTF) baseline has already been established. This is the case in the HEY BABY study where researchers already have an established relationship with the research participants rather than having to do random digit dialling. This prior relationship that participants have to the study may encourage higher response rates than typical with remote modalities. However, survey length may need to be shortened.

A novel approach would be to use ACASI (using the SMS/CATI hybrid) to replace the ACASI parts of the HEY BABY survey. This proposal will be discussed later in the paper.

6.2 Remote data collection as part of the original study design

The papers included here are about conducting high frequency panel surveys, often on topics about health, to draw nationally representative samples. Although these panel surveys are not directly related to the HEY BABY research aims, because the HEY BABY researchers do not desire to ensure that their survey matches a general population but rather matches their baseline, there is still overlap with modality considerations, which means these papers have been included in this review.

Panel studies are a particular design of longitudinal study in which the unit of analysis is followed at specified intervals over a long period, often many years, similar to the HEY BABY study. The key feature of panel studies is that they collect repeated measures from the same sample at different points in time. Most panel studies are designed for quantitative analysis and use structured survey data. In turn, most panel studies provide phones to participants to facilitate follow up. The typical panel data methodology is to collect data FTF as a baseline. After the baseline data collection, interviews are conducted regularly (weekly, bi-weekly,

monthly) over the phone. Phones are typically distributed to participants as part of uptake in the study.

The first series of studies looks at response rates across survey modalities (section 6.2.1). The second series looks at 'concordance' across survey modalities, as well as at response rates (section 6.2.3).

6.2.1. Response rates across survey modalities

This section of the paper will review response rates across survey modalities and the implications of using multiple survey modalities in a study. The papers referred to in this section have been summarised in Table 4.

Table 4: Mixed survey modalities and response rates

Study	Survey modalities and context	Response rate
Croke et al., 2012	Households were assigned one of four technologies: IVR, USSD, WAP, and CATI In Tanzania, all switched to CATI by round 8	Using CATI, 66%, moving to 75% under World Bank management after a 4-month delay Response rates could have been improved through distributing phones and solar chargers Those with access to electricity answered on average in 18.6 rounds versus 16.4 for those without access to electricity
Lau et al., 2019	Study explored the production and response rates to CATI, IVR, SMS, and FTF surveys in Nigeria	Survey found that FTF had the highest response rates at 99%, followed by CATI 15%, IVR 3% and SMS 0.2%.
Gallup, 2012	SMS, IVR and CATI within a panel survey in Peru and Honduras	IVR 20%, SMS 30%, CATI 51%. At the end of six waves, IVR decreased to 19%, SMS to 21%, and CATI to 39%

The paper by Croke et al. (2012) drew on two high frequency panel surveys in Africa (Tanzania and South Sudan). The authors argued that CATI with mobile phones is the best way of carrying out high frequency panel research in SSA.

When comparing the modalities, the fraction of respondents owning internet-enabled phones turned out to be very low (eliminating WAP), support

from the phone company to run USSD was minimal, and IVR turned out to be clumsy as questions had to be broken down to avoid too many response options. CATI did not have these drawbacks. Live phone interviews became the technology of choice, and all those who were reachable and had access to a mobile phone were put through a basic call centre, which consisted of a group of enumerators who each had multiple phones (one for each phone network allowing cheaper within network calls) and a computer with a standard data entry screen. SMS was not used as it was seen as having too many limitations.

With Tanzania, between round 8 and round 26, before the survey was transferred to the World Bank, an average of 304 respondents, or 66%, participated in the survey. After the survey had been put under World Bank management, after a four month gap in interviews, the number of respondents increased to 343 (75% of the sample). After 33 rounds of mobile interviews, the overall non-response rate was 25% of the 458 households in the sample that had access to phones. Four percent never responded to a request for a mobile phone interview.

6.2.2. Implications for HEY BABY

The best way of carrying out high frequency panel research in SSA is CATI with mobile phones (Croke et al., 2012). The Tanzanian example has the most relevance for the HEY BABY study as the initial research did not distribute phones, and attempted to compare WAP, USSD and IVR with CATI. However, after 8 rounds of high panel survey data, all participants were switched to CATI. Economic status was a significant predictor of survey participation: Households without a phone, those using non-premium phone providers, and those in the second poorest income quintiles were significantly less likely to participate, relative to households of median wealth. Wealth is found to be correlated with survey participation. However, when the regression is restricted to households that were identified as reachable in round 8, the impact of wealth largely disappears, and location (living in rural Dar es Salaam) and using the premium provider remain significant variables. Response rates could have been improved through distributing phones and solar chargers; those with access to electricity answered on average in 18.6 rounds versus 16.4 for those without access to electricity.

At first glance, the relevance of a high frequency panel survey for HEY BABY may not be clear. The survey was trying to find a nationally representative sample on which to conduct a household survey. However, the comparison of survey modalities (CATI, IVR, WAP or USSD) in the Tanzanian survey may be generalizable to HEY BABY, given the contexts are similar since Tanzania is also a LMIC in SSA. Furthermore, the Tanzanian example has the most relevance for

the HEY BABY study as the initial research did not distribute phones, and it had delays between baseline and follow up**.**

The study also raises additional relevant questions for HEY BABY in terms of phone ownership: there was an initial high attrition rate in Tanzania because a baseline was administered without reference to how many participants had phone access. The time between the survey of baseline data within Tanzania, (at 4 months) achieving a 66% response rate, moving to a 75% response rate, may be more typical of the HEY BABY experience. However, given we will not be distributing mobile phones, we can expect lower response rates than 75% or 66%. The potential for a mixed modality with SMS reminders that a CATI is due may be an easily replicable means of increasing response rate with HEY BABY. The suggestion that distributing solar chargers could help response rates as electricity may be a barrier for involvement in our research is worth considering, although the distribution of solar charges and credit could also be classed as an intervention for HEY BABY. Since they are the same tools to be used in the HEY BABY study, it could be seen as an intervention in itself which may have some impact on data integrity.

The higher response rate of CATI compared to IVR and SMS in the study by Lau et al. (2019) is due to two factors: (i) among all numbers dialled, more sample members agreed to participate in CATI (due to the ability of CATI interviewers to persuade people to participate); and (ii) 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 (86 questions for CATI versus 12 questions for SMS and IVR). By mixing modes – for example by starting with a primary mode (in this case it was IVR and SMS) and then contacting non-respondents with a secondary mode (in this case CATI) – researchers can leverage the strengths of each mode (Lau et al., 2019). 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.

As HEY BABY has an established baseline, there would be no need for random digit dialling. This means that HEY BABY's CATI response rate should be higher than in the Lau et al. (2019) study. There are two interesting areas of future research here: re-mixing modes to leverage the strengths of each mode, and the use of modular designs with SMS and IVR using repeated measurements to collect data from the same person or from independent samples.

The Gallup study (2012) found that using a field partner that is aware of the research site and has credibility in the community could result in a lower non-response rate. Further, it was noted that there was likely to be lower response rates from participants living in rural areas due to more restricted cellular phone coverage. There were also lower response rates and survey completions noted among participants who were exposed to self-administered methods of IVR and SMS (Gallup, 2012). Lastly, when the participants completed the survey without an interviewer present it seemed to affect the response to the first re-contact survey, as the groups exposed to the CATI methodology had a higher response rate (Gallup, 2012). The problem is more severe in the case of IVR where survey calls are lost for good when not answered immediately.

6.2.3. Concordance across survey modes: Establishing survey mode effects

Greenleaf et al. (2017) conducted a literature review of 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 in LMIC looking at IVR, CATI, SMS, or web surveys. The authors recognised that data collection mode can influence social desirability bias, and impact response rates, i.e. data collection mode can change the cognitive process for answer retrieval. They conducted a 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. No papers were found for SSA, and all but one paper was published in the last 5 years. The main themes drawn from these 10 papers were considerations around: (i) data concordance across survey mode, and (ii) social desirability bias across survey mode.

The Mahfoud et al. (2014) paper looked at a FTF survey in Lebanon. Half of the respondents were called on their mobile phones and asked an abridged version of a FTF questionnaire. There was high concordance (kappa) between the CATI and FTF surveys. Kappa was above 0.8 for measures including age, health insurance, diabetes, current cigarette-smoking (highest agreement and kappa: agreement=95.6%, κ=0.91), and ever cigarette-smoking (second highest agreement=93.5%, κ=0.87). Kappa was between 0.6 and 0.8 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. 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. More on data concordance and social desirability across survey mode is reflected in Table 5.

Table 5: Data concordance and social desirability across survey mode

Study	Data concordance across survey mode	Social desirability bias across survey mode
Greenleaf et al., 2017	FTF interviews with CATI surveys in 4 of the 10 articles. All comparisons generally showed concordance of results between modes Three articles made six comparisons of in-person interviewer administered (all FTF) and either IVR or SMS, with one of the three studies having moderate to good agreement and 62.4% of questions having the same answers for both surveys There was one study where there was inconsistent data agreement, with a kappa of .02 to .36. The authors proposed it was because of the unfamiliar terminology that was used in the SMS that in a FTF survey could be clarified by the interviewer	Item nonresponse rates and frequency of self-reported, sensitive sexual behaviours were statistically significantly different across the three methods for all reported questions (IVR, self-administered questionnaire (SAQ) and FTF SAQ) Subjective psychological responses, such as the perception of confidentiality, explain the lower report of undesirable behaviours in the self-administered in-person article survey compared with the other two modes Authors note that they are not sure whether sensitive behaviour is reported more frequently via text due to decreased time pressure or increased privacy
Mahfoud et al., 2014	After a FTF survey in Lebanon, half of the respondents were called on their mobile phones and asked an abridged version of a FTF questionnaire. There was high concordance (kappa) between the CATI and FTF surveys	Reports of past-year alcohol consumption was slightly higher via CATI compared with FTF. 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

The issue of concordance in the two studies have clear implications that need to be worked through for HEY BABY.

- How does changing survey mode change data reliability?
- How does changing survey mode increase or decrease social desirability bias?

Both of these considerations are explored further later in the paper in section 13 on mode effects.

7. Individual remote data collection in LMIC: case studies and examples

Once it has been decided that research can safely and ethically take place, there are a range of remote survey modalities to consider, which are explored in more detail within the research articles. Each survey modality has its own associated benefits and best practice recommendations. This section provides a high-level overview of survey modalities, as well as some initial comparisons between survey modes. The papers and think-pieces also highlight practical considerations on how these technologies have been used, both in cases where research has been interrupted and had to move to remote technologies, and in cases where the original design was to switch from FTF modalities to remote technologies.

7.1 CATI case studies

Table 6 captures CATI case studies, their response rates, incentives and weighting.

Table 6: CATI case studies response rate, incentives and weighting

Study	Survey length and frequency	Response rate and completion rate	Incentives	Weighting
Himelein & Kastelic, 2014	20-25 minutes	30% Calls were not answered because of suspicion of unknown foreign numbers, particularly during a crisis situation	Non-responders received a text message describing an incentive payment of one US dollar in phone credit for completed interviews. All numbers were again attempted five times, but the exercise netted only 48 additional households (2.2%)	Attrition correction was applied in addition to the probability weights Researchers calculated a propensity score adjustment and a post-stratification adjustment was applied

Study	Survey length and frequency	Response rate and completion rate	Incentives	Weighting
Dillon, 2012	10 minutes-1 hour Average: 27 minutes Participants were called once every 3 weeks and were given datasheets about when to expect calls from researchers	98% of participants were interviewed each round	Phones were supplied to study participants. Phone endowment appears to engender a deep sense of commitment to the project. US$0.76 of credit to each phone after each interview. Respondents viewed the free battery charge as an additional form of compensation	High participation rates: introduced as a study of cotton production. Farmers excited to see interest in their cultivation. Small, culturally homogenous villages. Villagers helped missing respondents
Garlick et al., 2020	20 minutes, with three groups of varying frequency: in-person every fourth week for 12 weeks; in person every week for 12 weeks; every week by mobile phone for 12 weeks	51%	Mobile credit voucher of US$1.17 for every fourth interview	Frequency effects on response rate, but higher frequency overall higher numbers Objective indicators of data quality did not differ systematically between different interview frequencies or survey mode

7.2 SMS case Studies

Table 7 reflects SMS case studies showing response and completion rates, and incentives.

Table 7: SMS case studies response and completion rate, and incentives

Study	Survey length	Response rate and completion rate	Incentives
Owsley et al., 2020	Up to 25 questions	Survey completion conditional on starting is 94% Completion rate of ~20% for SMS surveys with a low-income urban sample unaccustomed to SMS surveys Survey completion is similar for surveys of 25 questions compared to surveys of 5 questions — especially when you offer incentives	Small incentives for completing surveys (as low as US$0.25) increase your completion rate by ~5-6 percentage-points, 40% Higher incentives (only up to US$1.00 in this case) encourage completion when the survey is longer
Lau et al., 2018	N/A	18%, but 39% of these respondents were not the intended participants	US$1 – without verification wrong participants took the survey to get the phone credit
Feng et al., 2018	10 questions	Quota based 10,000 completed responses; 2,500 respondents per round; four rounds of surveys; one month apart each; 6,694 unique respondents, some did the survey more than once – perhaps because of phone sharing	US$0.50
Berman et al., 2017	16 questions (recommendation 12-15)	34% Questions sent to 3,000 people and 1,000 people completed survey	US$0.50 in mobile credit on survey completion

If respondents are told how long the SMS survey will be it increases completion by 2 percentage points (pp) (Owsley et al., 2020). The findings by Owsley et al. (2020) also showed that it is best to randomize the order of questions because of order effects and because the order affects which responses are picked.

If a significant amount of time has lapsed since follow up when conducting SMS surveys, a participant should be asked to validate their identity to avoid the high rate of wrong participants taking part in the survey, as seen in the study by Lau et al. (2018). The pre-SMS survey telephone calls in the study by Lau et al. (2018) were inexpensive and effective in getting participants to participate. The phone call check-in may have been a better mode than SMS because of its higher response rate (52%) and ability to confirm that the correct person was answering the survey.

The study by Feng et al. (2018) deployed a national SMS-poll in Liberia and collected data from 6,694 individuals about individual-level health and health-seeking behaviour throughout the Ebola outbreak from March to June 2015. The survey used propensity score matching from a recent household survey (Liberia Institute of Statistics and Geo-Information Services et al., 2014) to generate balanced samples to match and compare variables from the Demographic Health Survey (DHS), to measure changes.

SMS was also used during the Ebola outbreak in Liberia in order to collect information from communities on the front lines of the outbreak, to feed that information back into the community response planning process, and to start a dialogue with affected stakeholders (Berman et al., 2017). At the time, no data was available about the level of public knowledge of the Ebola Virus Disease. The SMS questionnaire was designed to produce a rapid assessment of ideational elements that could inform the content and strategy of the communication response. The limit of SMS messages to 160 characters constrained the space available for the questions to be asked. Crafting the question properly may be the biggest challenge in conducting such a survey. If HEY BABY is going to include a focus on COVID-19, it will need a theoretical framework that is the basis on which to design the questions. Rather than just asking directly about a particular health issue (such as COVID-19), Berman et al. (2017) indicated the importance of including questions that are based on an established theoretical framework. In their study they used an adaptation of "a metatheory of health communciation" (Kincaid, 2000) to know which types of questions to ask.

7.3 ACASI case studies

The Beauclair et al. (2013) paper laid out that ACASI is acceptable in urban disadvantaged communities in South Africa. The study looked at FTF interviewing versus audio computer-assisted self-interviewing (ACASI) in order to reduce social desirability bias within a specific South African context. A total of 878 participants responded to questions about their sexual relationships on a touch screen computer in a private office. The survey was carried out in Cape Town. At the end, participants were asked about the ease of use, privacy, and truthfulness. The results were compared to reported behaviours in FTF interviews (demographic and health surveys) from Lesotho, Swaziland and Zimbabwe.

The study had the following results:

- 85.9% of participants found methods easy to use – unemployed participants and those in the 40–70 year-old age group were the least likely to find their methods easy to use;
- 96.3% perceived privacy; and
- 82.5% preferred ACASI to other modes of inquiry when reporting on sexual behaviour.

With the ACASI model, participants disclosed having more partners and more concordant relationships (i.e., more than one relationship at the same time) than the numbers disclosed through the DHS which was a FTF survey modality, asking the same questions. The ACASI survey also found that there was a lower difference ratio in the responses between male and female participants as related to relationship concordance and partner number (i.e., males and females were more likely to have similar responses in terms of numbers of partners and relationship overlap) than disclosed in the DHS FTF survey. This has implications for desirability bias. It is important to note that this was a "relatively short survey" (Beauclair et al., 2013), but item length or response rate are not known to the HEY BABY researchers.

The Kane et al. (2016) paper was a practical guide on how to convert a complex paper-based questionnaire focused on mental health, substance use, and HIV risk behaviours into an ACASI system. The paper highlighted the benefits of ACASI-based interviews relative to other modalities (according to Van de Wijgert et al., 2000; Brown et al., 2013) as:

- increased data validity for sensitive measures – removal of social desirability and interviewer biases;

- increased participant privacy;
- functionality for illiterate participants;
- ability to be programmed in multiple languages;
- automatic and accurate programming of skip/logic patterns;
- fewer missing data and better non-response rates;
- no effect of inter-rater variance;
- reduction in staff time needed for interviewing; and
- obviation of the need for costly, time-intensive and (possibly) inaccurate data entry.

7.3.1. Social desirability bias in ACASI case studies

Langhaug et al. (2010) identified 26 studies, conducted in LMICs, which focused on sexual behaviours and examined the use of ACASI as compared with other interview techniques. The review found strong evidence that ACASI-administered interviews resulted in lower rates of reporting bias than comparison methods, including self-administered questionnaires and FTF interviewing. More recent studies published after the review have found similar results (Beauclair et al., 2013, Adebajo et al., 2014). Participants in the Beauclair et al. (2013) study perceived privacy (96.3%) when responding to questions around their sexual relationships because they were responding to questions on a touch screen computer in a private office. In addition, they reported preferring ACASI to other modes of inquiry (82.5%) when reporting on sexual behaviours. Furthermore, more partners and more relationship concordance were disclosed, and there were lower difference ratios between male and female participants than observed in the DHS.

There is potential here to use a hybrid of CATI with SMS responses – that is, use CATI interviews to ask questions and get respondents to respond via SMS rather than vocally over the phone to increase disclosures (i.e. a remote ACASI). Opting for this would give the privacy benefits of ACASI, while avoiding social desirability issues. A potential issue here relates to baseline comparisons: if ACASI was not used in the baseline, and one modeled a ‘remote ACASI’ one may get enhanced disclosures because of the modality switch. However, since this review, HEY BABY researchers went through the FTF survey and found that there were parts of the survey that were ACASI in origin. Therefore, trying to find a remote alternative to ACASI makes even more sense in the HEY BABY context.

7.4 IVR case studies

According to L'Engle et al. (2018), in order to promote participant understanding, questions should be concise when using IVR. For responses about age, multiple choice format improved call continuation and data quality in comparison to entering a specific number. The L'Engle et al. (2018) study was conducted over a 27-day period in February-March 2017. There were 1,076,258 calls placed in order to get 46,849 people to start the survey (which equates to a productivity rate of 23 calls dialled for a survey start). A total of 16,003 eligible respondents participated in the survey: 9,469 completed the full interview which was a 31% response rate. Sixty-seven calls were dialled to yield an eligible contact, and 83 calls were dialled to yield an interview. Respondents who missed the call or were unable or unwilling to complete the survey at the time of the call could call back using their phone's missed call or redial feature. Nonresponse was substantial due to the automated calling system and dialling of many unassigned or non-working numbers. Younger, urban, better educated, and male respondents were overrepresented in the sample and so it could be concluded that this survey methodology is helpful for reaching that demographic.

7.5 Mixed modalities case studies

Mixed mode data collection can draw from the best aspects of each method, such as the combination of CATI and SMS, where a follow up message is sent once a respondent has completed a CATI survey (IPA, 2020).

The Carranza et al. (2020) study is a hybrid model of FTF, SMS and CATI: an FTF baseline, an SMS survey 2-3 days later, followed up by a phone-based 3-4 month midline and 10-12 month endline in Johannesburg with 6,891 jobseekers in Johannesburg. There was a response rate of ~83%, after 2-3 days of an in-person baseline and a small incentive of US$1. The survey consisted of two questions with specific responses. For the phone-based midline follow-up 3-4 months after recruitment, there was 96% completion. For the phone based endline follow-up 10-12 months after recruitment, there was 85% completion. Both follow-ups were 30-40 minutes in length.

Key learnings from a J-PAL presentation (Carranza et al. 2020) include realising that phone numbers in the study demographic change frequently. It is good practice to send an SMS to remind the participant that the survey will be happening soon. Keeping the survey within 30 to 40 minutes and giving the US$1 phone credit as an incentive for completing the survey at the end ensured participants were more likely to finish the survey. CATI can be used in a highly effective and efficient manner through services like SurveyCTO on tablets,

phones or desktop computers. Using CATI allows for data to be available either instantly or at the end of the day, depending on connectivity constraints – and data quality checks can be performed immediately. However, there is a need for a longer training process for research assistants to conduct this kind of survey (Carranza et al.(2020) held a 3-day training session).

8. Reducing non-response bias and attrition effects

One of the issues with mobile phone survey modes is that it can be more difficult to persuade reluctant people to participate over the phone, thus contributing to programme attrition. Programme attrition is where those enrolled in a programme drop out before it ends, which leads to non-response bias. Attrition and non-response are only problematic when they occur in a non-random manner. If attrition and non-response are associated with observable characteristics of respondents which have been recorded in the baseline survey, then it is manageable, and can be addressed by re-weighting the remaining respondents by the inverse of the probability of attrition.

8.1. Implications for HEY BABY

Since HEY BABY is a follow up study, there is already a baseline group, some of whom may be unable to be reached by the new survey modality, and yet need to be compared with the baseline. If one cannot reach them in this second round, how does one compare their changing needs and demographics and how does one ensure any changes in data are the result of real change rather than non-response bias?

The ideal would be to use incentivisation for higher response rates (explored in the next section of the paper). The question, then, is how much do incentives have an effect in terms of data reliability, as well as how much is needed to incentivise participation?

However, if some individuals still cannot be reached or do not complete the survey, we will need to correct for their non-representation. With what we know about survey modes, we can expect that mobile phone modes will over-represent young, urban and well-educated respondents. We can correct for these through using demographics to predict who is less likely to respond; using call attempts to match subgroups within treatment and control groups (Behaghel et al., 2015); and lastly through intensively following up a sub-sample of hard-to-reach people and weight those people in the survey (Millán & Macours, 2017).

9. Incentives

The Leo et al. (2015) study took place in four low-income countries. Data relevant to the HEY BABY study spoke to the research question: 'How effectively does monetary compensation impact survey completion patterns?' Most studies point to a positive effect of incentives on response rates across most related modalities, including web-based, cross-sectional, and panel studies. However, evidence is mixed when examining specifically the case of mobile phone-based surveys, with some cases reporting no effects of monetary compensations.

To further test the effect of incentives on mobile surveys, Leo et al. (2015) randomized the introduction that respondents received when they started the survey, testing two treatment conditions and having a control group that did not receive either of the treatment conditions. The first treatment was a transfer condition of receiving four minutes of mobile airtime upon completion of the survey. The second treatment condition was the raffle condition of being entered into a raffle to win a substantial amount of mobile airtime at the end of the survey. Pooling the data from all four surveys together shows that both the raffle and transfer conditions have significantly higher completion rates than the control conditions. However, the pooled analysis does not indicate that either treatment condition was more effective than the other. It must be noted that failure of random assignment in Zimbabwe and Ethiopia caused an unintended selection effect in the study. We therefore cannot exclude the possibility that apparent treatment effects may instead actually be selection effects. Individuals across different income groups seem to be equally affected by incentives. In a similar vein, the Leo et al. (2015) study found no difference in age, gender or education groups with regard to their response to the incentives treatment. The only exception found was in Afghanistan, with urban dwellers more likely to react to monetary incentives than rural ones. Offering compensation for survey completion mitigated attrition rates in several of the pilot countries.

10. Matching subgroups within treatment and control groups

The Behaghel et al. (2012) paper uses basic information on the number of attempts that were made to obtain a response to the survey from each individual, in order to compare balanced groups. The method can be applied whenever data collection entails sequential effort to obtain a response; for instance, trying to call several times in a phone survey, making several visits to the respondent's home, or even gradually offering higher incentives (gifts) to potential respondents. This is most

effectively done with a treatment and control group. The results showed that individuals respond to the survey depending on an unobserved variable – call it V – that can be interpreted as reluctance to respond. Analysts can form a subset of treatment and control individuals that responded to the survey and have the same distribution of V. If 60% of the respondents in a group answered before the 18th call, they must have the 60% lowest values of V (those who responded after the 18th call or never responded must have the 40% highest values of V). If we have two groups (treatment and control) in which the answering behaviour is different, it remains true that, if 60% of the respondents in the treatment group answered before the 18th call, and 60% of the respondents in the control group answered before the 20th call, each of these two subsamples contain the 60% lowest values of V within their group. When the groups are randomized (or can be considered initially similar for any other reason), the 60% lowest Vs in each group represent the same population. The insight of the paper therefore is that when survey attempts are recorded, two groups can be identified in the data, and compared with no risk of selection bias, because they have the same distribution of V.

11. Intensive tracking to weight demographics differently

To correct for attrition in the Özler and Cuevas (2019) study, more calls were made – they increased number of calls at different times of the day and managed to reach 112 of 150 households, which is a 75% rate of successful follow-up. That is an 18 pp increase from the 18-month follow-up, in a striking reversal of trends. Had the study stopped at three calls, they would have had a gap of about 13 pp in the rate of success between beneficiaries and non-beneficiaries: 55% in the former vs. 68% in the latter. But, by 10 calls, the gap had narrowed down to just 8 pp (70% vs. 78%). There was a boost to success between calls 4 and 5, which is when the first evening call happened. So, the evening call was a boost to calling people in work hours. In contrast, there was no similar boost between calls 7 and 8, which is when the first weekend call happened, suggesting that conditional on having tried all hours of the day, adding weekends did not add that much value (which may have cost implications). Papers showing the impact of intensive tracking on attrition rates are shown in Table 8.

For the HEY BABY study, a similar tracking method could be used by using the number of calls it took to reach a certain percentage of beneficiaries (or non-beneficiaries), as described in Behaghel et al. (2015). HEY BABY successfully interviewed 70% of non-beneficiaries after 10 calls, while it took only three to four calls to reach the same percentage of beneficiary households. Limiting the

sample of beneficiaries to those reached within four calls and doing some standard bounding analysis would give us another way of assessing attrition bias.

Instead, Özler and Cuevas (2019) carried out a community tracking protocol to find the 38 individuals that could not be reached over the phone by visiting their houses. This led to a 57% rate of success. The intensive tracking protocol may not be ethical or feasible within the HEY BABY study, given COVID-19, and the nature of our topic, but it is a useful protocol in terms of phone calls.

The Baird et al. (2008) report found that although it was costly to track the individuals over long distances, movers and non-movers[3] look quite different along a number of observable characteristics, reinforcing the importance of respondent tracking for credible econometric inference.

Hill (2004) found that tracking can reduce attrition and is feasible if procedures are locally appropriate, well planned, involve the community, collect as much locating data as possible, and have explicit criteria; if tracking is done at regular intervals; and if interviewers are well trained, supervised, and motivated.

Table 8: Impact of intensive tracking on attrition rates

Study	Attrition rates	Tracking to correct for attrition	Success rate of follow up
Özler & Cuevas, 2019	57% of initial participants remained after 18 months	They increased number of calls at different times of the day	Effective success rate made it to 89.5% (0.75 + 0.58* (1-0.75)).
Baird et al., 2008	16%	Tracking respondents across country borders in order to reduce attrition bias by using a two-stage tracking methodology, using a team-oriented approach, and cell phone technology. This paper did not weight those found during ITP differently to the complete tracking phase sample	Contact was made with 84% of target respondents, and 83% were surveyed (the remaining 1% either had died or refused to participate in the survey)

[3] Movers and non-movers in this context refer to the differences between those who move home and those who don't move at all.

Study	Attrition rates	Tracking to correct for attrition	Success rate of follow up
Hill, 2004	Studies with no tracking had 9–21% attrition per year compared with only 1–8% for those with tracking.	Tracking can reduce attrition by up to 45%	The number of respondents located by tracking who would have otherwise been lost ranged from 5% to 45% of the sample.
Millán & Macours, 2017	Attrition was almost 30% after the regular tracking phase (RTP), similar to attrition rates also found in young mobile population in other studies	Intense tracking phase: 90% Not including those found during the intensive tracking leads to an overestimate of the Intent-to-treat (ITT) effects for the target population by more than 35% Used inverse probability weighting with estimates of weights that exploit the similarities between missing individuals and those found during an intensive tracking phase. Compared these estimates with alternative strategies using regression adjustment, standard weights, bounds or proxy information	Attrition falls to 10% after the intense tracking phase (ITP), and the data collected during the ITP allows quantifying the attrition bias obtained after regular tracking only

12. Correcting for non-response bias

Mercer, Lau & Kennedy (2018) look at an online opt-in survey in the United States which tries to reduce opt-in bias through a range of statistical measures. This paper was unable to completely compensate for these challenges through raking, propensity weighting and matching, to adjust samples so that they more closely match the population on a chosen set of dimensions. Although this paper is trying to increase sample reliability vs. the American public, it is a useful general discussion in terms of matching the HEY BABY remote data collection

group with the original baseline, given that any remote data collection method will inevitably exclude those who do not have phones.

The study tested a variety of elaborate weighting adjustments to online opt-in surveys with sample sizes as large as 8,000 interviews. Across all of these scenarios, none of the evaluated procedures reduced the average estimated bias across 24 benchmarks below 6 pp – down from 8.4 points unweighted. Even the most effective adjustment strategy was only able to remove about 30% of the original bias. When it comes to accuracy, choosing the right variables for weighting is more important than choosing the right statistical method. The most basic weighting method (raking) performs nearly as well as more elaborate techniques based on matching. Whatever method used, successfully correcting of bias in opt-in samples requires having the right adjustment variables. The 'right' adjustment variables include more than the standard set of core demographics. A careful consideration of the factors that differentiate the sample from the population, and their association with the survey topic, is far more important. This is what would guide us on which variables to weight.

13. Mode effects: social desirability bias and privacy considerations

It is important to highlight that a change in research mode may impact the self-reported data of respondents, which has implications for data integrity. The change from FTF modalities to a remote data survey measure may increase self-reported risk behaviour thus introducing privacy bias that may need to be controlled. Most of the research on social desirability bias has looked at ACASI versus FTF models. However, there is an assumption in the literature that the survey modality will change your baseline FTF self-reported data, regardless of what new mode you choose. This can be celebrated in terms of the gains that can be made with remote data collection modalities, including the quality of data entry, rather than bemoaning what is lost by FTF contact. However, this section continues by being cautious about the impact on data integrity of switching mode during a research study, including consideration of frame error, non-response error, and measurement error, because of social desirability bias, primacy effects, or inaccurate recordings by the interviewer.

As mentioned in the Cullen and Mahmud (2020) think piece, ACASI can be a good complement to standard FTF survey questions by giving respondents the option to respond privately, rather than to the enumerator, particularly in low literacy settings. In addition, ACASI allows for delivery of sensitive questions to

be more uniform, potentially minimising any enumerator effects. It is also possible to integrate ACASI-based questions within a FTF survey. This requires adding instructions for the enumerator to ensure privacy, to hand over the tablet to the respondent, and then to step away. By implication, it is possible to add ACASI to a CATI model using SMS survey responses. Table 9 shows studies on mode effects and error types.

Table 9: Mode effects and error types

Study	Mode effects	Error type
Cullen & Mahmud, 2020	FTF surveys versus ACASI	Reporting bias
Langhaug et al., 2010	Despite wide variation in geography and populations sampled, computer-assisted interviewing decrease item non-response rates and increases rates of reporting of sexual behaviours	Data entry errors were also reduced when controlled by a computer programme: in studies with no differences in reporting of sexual behaviours, ACASI still improved the quality of data entry.
Dillman & Christian, 2005	In FTF the interviewer assessment of body language and the building of rapport is important. FTF uses long and complex scales Telephonic interviews tend to be simplified. Telephonic interviews use shortened scales	Survey error due to changing the mode of survey, especially changes in the wording of questions Questions that use vague quantifiers and require the formulation of an answer on the spot are more susceptible to being influenced by context and mode effect.
Pariyo et al., 2019	The mode of data collection (e.g., CATI, IVR, in-person) impacts survey participation and responses	Response and measurement effects introduced by a mode are jointly referred to as mode effects Frame error Non-response error Measurement error

The Langhaug et al. (2010) review found strong evidence that ACASI-administered interviews result in lower rates of reporting bias than comparison methods, including self-administered questionnaires and FTF interviewing. More recent studies published after the review have found similar results (Beauclair et al., 2013; Adebajo et al., 2014).

Dillman and Christian (2005) found that identically worded questions often produce different answers when administered through different modes, although this is not LMIC specific. Changing modes may produce results that provide a false indication of change having occurred between waves of data collection. To control for mode effects one can use unimode construction, i.e. write questions that will work across survey modes and reduce the length of scales. One must also take into account the visual design and layout as this may have paralanguage effects – respondents can draw meaning from visual layout as much as words in self-administered modes. Mode changes between the initial data-collection period and follow-up data collections for panel studies should be evaluated carefully. There is a considerable likelihood that such changes make it difficult to accurately measure change between survey waves. Each mode favours the construction of questions in ways that differ from those favoured by each of the other modes. When data collection for panel studies is left to technical staff associated with each mode, it is likely that question structures will be changed.

The Pariyo et al. (2019) paper outlined that understanding mode effects is important because the mode used can introduce three types of errors:

- frame error, in which certain members of the target population are erroneously included or excluded from the sample frame. Frame error is concerning in countries where cell phone ownership is not above 80%. With the HEY BABY study, 29% of participants reported owning a phone at baseline data collection and 81% had access to a phone;
- non-response error, when those who respond to the survey are different from those who do not respond to the survey; and
- measurement error, due to the responses recorded in the survey not being accurate either due to respondent or interviewer error. Three reasons are social desirability bias, primacy effects, and inaccurate recording by interviewer.

In combining modes or comparing results from surveys collected using different modes, researchers should attempt to design surveys and the questions in such a way as to reduce each of the three sources of error. For example, mobile phone survey participants receiving either an IVR or CATI survey on first contact may provide comparable responses for certain types of questions, but not others, when re-surveyed using a different survey delivery modality. This suggests that mobile phone survey researchers should be wary of lumping together results collected using different modes, even when collected contemporaneously. These findings show that reliability varies by question. Careful design can help ensure clarity of

questions to minimize the cognitive burden for respondents, many of whom may not have prior experience in taking automated surveys.

14. Key take home lessons for HEY BABY

There are several key take home lessons of relevance for the HEY BABY study (and other research studies having to change to remote data collection methodologies).

Firstly, it seems that a hybrid model of CATI combined with SMS is best suited to the HEY BABY study. Evidence suggests that CATI is best for an interrupted study when a FTF relationship has been established during baseline (which is the case for HEY BABY). It must be noted that SMS reminders and follow ups can be used to increase response rates. However, it needs to be kept in mind that there is likely to be higher attrition rates amongst respondents living in rural and/or impoverished areas (which encompasses many of the respondents in the HEY BABY study), as they may not have access to a mobile phone and/or may have restricted network coverage.

Secondly, while economic incentives (such as receiving mobile phone airtime) may lead to a higher response rate, caution needs to be exercised as using economic incentives may impact on data integrity. Psychological incentives – i.e., clear communication of the study's mission to respondents and the important role they have to play – may be of particular relevance here, as well as having a local field partner within the research site who has credibility within the community and who could help strengthen participation rates.

Thirdly, it is inevitable that some respondents will not be reached via mobile phone and as such we need to be able to correct for their non-representation. This can be done through intensive tracking and subgroup weighting.

15. Conclusion

This paper summarises results from a review of quantitative research methods that could be used instead of FTF data collection during the COVID-19 pandemic with a highly vulnerable group: adolescent mothers in South Africa. We review experiences to inform the design of the follow-up data collection activities of the HEY BABY study, identifying practical and methodological issues and recommendations for remote quantitative research.

References

Adebajo, S., Obianwu, O., Eluwa, G., Vu, L., Oginni, A., Tun, W., Sheehy, M., Ahonsi, B., Bashorun, A. & Idogho, O. 2014. Comparison of audio computer assisted self-interview and face-to-face interview methods in eliciting HIV-related risks among men who have sex with men and men who inject drugs in Nigeria. *PLOS One,* 9, e81981.

Baird, S., Hamory, J. & Miguel, E. 2008. *Tracking, attrition and data quality in the Kenyan life panel survey round 1 (KLPS-1).* UC Berkeley: Center for International and Development Economics Research.

Beauclair, R., Meng, F., Deprez, N., Temmerman, M., Welte, A., Hens, N. & Delva, W. 2013. Evaluating audio computer assisted self-interviews in urban South African communities: evidence for good suitability and reduced social desirability bias of a cross-sectional survey on sexual behaviour. *BMC Medical Research Methodology,* 13, 11.

Behaghel, L., Crépon, B., Gurgand, M. & Le Barbanchon, T. 2012. Please call again: Correcting non-response bias in treatment effect models. *IZA Discussion Paper No. 6751*. Bonn: Institute for the Study of Labour.

Behaghel, L., Crépon, B., Gurgand, M. & Le Barbanchon, T. 2015. Please call again: correcting nonresponse bias in treatment effect models. *Review of Economics and Statistics*, 97, 1070-1080.

Berman, A., Figueroa, M.E. & Storey, J.D. 2017. Use of SMS-based surveys in the rapid response to the Ebola outbreak in Liberia: opening community dialogue. *Journal of Health Communication,* 22, 15-23.

Brown, J.L., Swartzendruber, A. & DiClemente, R.J. 2013. Application of audio computer-assisted self-interviews to collect self-reported health data: an overview. *Caries research*, 47(Suppl. 1), 40-45.

Carranza, E., Garlick, R., Orkin, K. & Rankin, N., 2020. *Job search and hiring with two-sided limited information about workseekers' skills*. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3634690

Croke, K., Dabalen, A., Demombynes, G., Giugale, M. & Hoogevee, J. 2012. Collecting high frequency panel data in Africa using mobile phone interviews. *Policy Research Working Paper Series No. 6097*. Washington, D.C.: World Bank Group.

Cullen, C. & Mahmud, M. 2020. *Surveying on sensitive topics: using audio computer assisted self-interviewing.* [Online] University of Oxford: Mind and Behaviour Research Group. Available: https://mbrg.bsg.ox.ac.uk/method/surveying-sensitive-topics-using-audio-computer-assisted-self-interviewing [Accessed 12 January 2021].

Dillman, D.A. & Christian, L.M. 2005. Survey mode as a source of instability in responses across surveys. *Field Methods*, 17, 30-52.

Dillon, B. 2012. Using mobile phones to collect panel data in developing countries. *Journal of International Development,* 24, 518-527.

Feng, S., Grépin, K.A. & Chunara, R. 2018. Tracking health seeking behavior during an Ebola outbreak via mobile phones and SMS. *NPJ Digital Medicine,* 1, 51.

Gallup. 2012. *The World Bank Listening to LAC (L2L) Pilot Project. Report on Attrition of Panel Participants in Peru and Honduras*. https://microdata.worldbank.org/index.php/catalog/2022/download/30813 [Accessed 12 January 2021]

Garlick, R., Orkin, K. & Quinn, S. 2020. Call me maybe: experimental evidence on frequency and medium effects in microenterprise surveys. *The World Bank Economic Review,* 34, 418-443.

Greenleaf, A.R., Gibson, D.G., Khattar, C., Labrique, A.B. & Pariyo, G.W. 2017. Building the evidence base for remote data collection in low- and middle-income countries: comparing reliability and accuracy across survey modalities. *Journal of Medical Internet Research,* 19, e140.

Hill, Z. 2004. Reducing attrition in panel studies in developing countries. *International Journal of Epidemiology,* 33, 493-498.

Himelein, K., Eckman, S., Lau, C. & Mckenzie, D. 2020. *Mobile phone surveys for understanding COVID-19 impacts: part I sampling and mode.* [Online] Available:https://blogs.worldbank.org/impactevaluations/mobile-phone-surveys-understanding-covid-19-impacts-part-i-sampling-and-mode [Accessed 7 November 2020].

Himelein, K. & Kastelic, J.G. 2014. *The socio-economic impacts of Ebola in Liberia: results from a high frequency cell phone survey*. World Bank Group.

Innovation for Poverty Action [IPA]. 2020. Remote surveying in a pandemic: handbook. In Glazerman, S., Rosenbaum, M., Sandino, R. & Shaughnessy, L.

(eds) *Innovations for Poverty Action*. 1-43.

Kane, J.C., Murray, L.K., Sughrue, S., Demulder, J., Skavenski Van Wyk, S., Queenan, J., Tang, A. & Bolton, P. 2016. Process and implementation of audio computer assisted self-interviewing (ACASI) assessments in low resource settings: a case example from Zambia. *Global Mental Health*, 3, e24.

Kincaid, D.L. 2000. Mass media, ideation, and behavior: a longitudinal analysis of contraceptive change in the Philippines. *Journal of Communication Research,* 27, 723-763.

L'Engle, K., Sefa, E., Adimazoya, E.A., Yartey, E., Lenzi, R., Tarpo, C., Heward-Mills, N.L., Lew, K. & Ampeh, Y. 2018. Survey research with a random digit dial national mobile phone sample in Ghana: methods and sample quality. *PLOS One,* 13, e0190902.

Langhaug, L.F., Sherr, L. & Cowan, F.M. 2010. How to improve the validity of sexual behaviour reporting: systematic review of questionnaire delivery modes in developing countries. *Tropical Medicine & International Health,* 15, 362-381.

Lau, C., Cronberg, A., Marks, L. & Amaya, A. 2019. In search of the optimal mode for mobile phone surveys in developing countries. A comparison of IVR, SMS, and CATI in Nigeria. *Survey Research Methods*, 13(3), 305-318.

Lau, C., Gachugu, E., Johnson, E. & Marks, L. 2018. Using SMS technology to survey low-income youth: lessons from a vocational education tracking study in Kenya. *Journal of International Development,* 30, 1060-1063.

Leo, B., Morello, R., Mellon, J., Peixoto, T. & Davenport, S.T. 2015. Do mobile phone surveys work in poor countries? *Center for Global Development Working Paper No. 398*. Available at SSRN: http://dx.doi.org/10.2139/ssrn.2623097

Liberia Institute of Statistics and Geo-Information Services, Ministry of Health and Social Welfare [Liberia], National AIDS Control Program [Liberia], and ICF International. 2014. *Liberia Demographic and Health Survey 2013*. Monrovia, Liberia: Liberia Institute of Statistics and GeoInformation Services and ICF International.

Mahfoud, Z., Ghandour, L., Ghandour, B., Mokdad, A.H. & Sibai, A.M. 2014. Cell phone and face-to-face interview responses in population-based surveys: how do they compare? *Field Methods,* 27, 39-54.

Mercer, A., Lau, A. & Kennedy, C. 2018. *For weighting online opt-in samples, what matters most?* [Online] Pew Research Center. Available:

https://www.pewresearch.org/methods/2018/01/26/for-weighting-online-opt-in-samples-what-matters-most/ [Accessed 19 November 2020].

Millán, T.M. & Macours, K. 2017. Attrition in randomized control trials: using tracking information to correct bias. *IZA Discussion Paper No. 10711*. Bonn: IZA Institute of Labor Economics.

Owsley, N., Jang, C., Kamande, S., Hassanali, K. & Leidich, A. 2020. *Getting the most out of your SMS survey. Results from an experiment testing the effects of SMS survey design on response rates and patterns.* [Online] Busara Center. Available: https://www.busaracenter.org/getting-the-most-out-of-your-sms-survey-experimental-results [Accessed 23 Dec 2020].

Özler, B. & Cuevas, P. 2019. *Reducing attrition in phone surveys*. [Online] Available: https://blogs.worldbank.org/impactevaluations/reducing-attrition-phone-surveys [Accessed 23 December 2020].

Pariyo, G.W., Greenleaf, A.R., Gibson, D.G., Ali, J., Selig, H., Labrique, A.B., Al Kibria, G.M., Khan, I.A., Masanja, H. & Flora, M.S. 2019. Does mobile phone survey method matter? Reliability of computer-assisted telephone interviews and interactive voice response non-communicable diseases risk factor surveys in low and middle income countries. *PLOS One*, 14, e0214450.

Peterman, A., Bhatia, A. & Guedes, A. 2020. *Remote data collection on violence against women during COVID-19: a conversation with experts on ethics, measurement & research priorities.* [Online] Available: https://www.unicef-irc.org/article/1997-remote-data-collection-on-violence-against-women-during-covid-19-a-conversation-with.html [Accessed 23 December 2020].

Rogers, K. 2020. *Has it become too dangerous to measure violence against women?* [Online] Devex. Available: https://www.devex.com/news/has-it-become-too-dangerous-to-measure-violence-against-women-97112 [Accessed 23 December 2020].

Van de Wijgert, J., Padian, N., Shiboski, S. & Turner, C. 2000. Is audio computer-assisted self-interviewing a feasible method of surveying in Zimbabwe? *International Journal of Epidemiology*, 29(5), 885-890.

World Health Organisation. 2020. *Coronavirus disease (COVID-19) pandemic.* [Online] World Health Organization. Available: https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/novel-coronavirus-2019-ncov [Accessed 22 December 2020].