Sustainable data collection for development: mixed modes and statistical modelling

February 2021

Authors

Shirley Jeoffreys-Leach, Bobby Berkowitz, Irma Grundling, Jonathan Gellar

About FinMark Trust

FinMark Trust is an independent non-profit trust whose purpose is 'Making financial markets work for the poor, by promoting financial inclusion and regional financial integration'. We pursue our core objective of making financial markets work for the poor through two principle programmes. The first is through the creation and analysis of financial services consumer data to provide in depth insights on both served and unserved consumers across the developing world. The second is through systematic financial sector inclusion and deepening programs to overcome regulatory, supplier and other market level barriers hampering the effective provision of services. Together, these programmes unlock financial inclusion and sector development through a symbiotic relationship between rigorous data collection and research activities. Our work can be found in South Africa, throughout the SADC region and the global arena.

For more information:

Visit our website at www.finmark.org.za
Email Grant Robertson (Executive of Data and Analytics) on
grantr@finmark.org.za
Call us on +27 11 315 9197

CONTENTS

1. Introduction

The groundwork for meaningful policymaking and appropriate and adequate allocation of resources for development is done through the collection of data. Only through the evidence provided by this data can policymakers and regulators gain the insights they require to create a conducive environment for effective and sustainable economic development. If used appropriately, data has transformative power as it facilitates the development of indicators to measure the status quo, set targets, and to track change over time periods and hold governments and service providers accountable.

Collecting data on financial inclusion and progress in terms of bringing financial services to those without access in a cost-effective and innovative manner has become essential for governments committed to economic development. This research has provided significant evidence that financial system development not only precedes significant economic growth, but also that the relationship is causal[1].

The cost, especially of demand-side data collection, is arguably the greatest barrier to more and sustainable financial inclusion research, especially in developing countries that are dependent on donor funding to conduct the research. As a result, demand-side research occurs less frequently than is needed to effectively facilitate early-warning signals for interventions to timeously enable the achievement of financial inclusion targets.

With the sustainability of data collection in mind, the insight2impact facility has set out to test whether or not mobile technology could be used to offer faster and more cost-effective data collection. It aims to determine whether it can do this with the same reliability and precision as the more common face-to-face demand-side financial inclusion surveys.

The cost, especially of demand-side data collection, is arguably the greatest barrier to more and sustainable financial inclusion research, especially in developing countries that are dependent on donor funding to conduct the research.

[1] https://www.cgap.org/blog/learning-financial-inclusion-research-what-should-we-expect

2. Executive Summary

The core strength of the mixed modal approach lies in it still producing reliable estimates of the indicator values at sub-group levels of analysis.

The i2i facility set out to pilot a mixed modal sample design, together with a calibrated multilevel regression and post stratification (MRP) model. This was done as an alternative to a face-to-face only option for digital financial services and gender indicator data collection. The pilots were conducted using a mix of SMS and telephonic interviewing methods, along with in person face-to-face surveys in four African and four Asian markets (Kenya, Nigeria, Tanzania, Uganda, Bangladesh, India, Indonesia, and Pakistan).

The calibrated MRP provided an innovative solution to account for both the modal differences expected across different data collection modes and the sampling bias produced by the mobile modes.

These pilots demonstrated that good results can be achieved from a sample as small as n=150 surveys collected face-to-face. However, having a detailed and reliable reference data source, such as a recent nationally representative face-to-face survey or census, is critical for successful post-stratification. The core strength of the mixed modal approach lies in it still producing reliable estimates of the indicator values at sub-group levels of analysis. Once this approach was used, indicator values calculated fell inside the 95% confidence interval of the benchmark indicator in question.

The use of this methodology provides opportunities for more sustainable indicator collection and tracking.

3. The Topic of the Pilot Surveys

With the potential of technology to extend the geographic reach of financial institutions' existing services at relatively low cost, digital financial services have in recent time become the focus of most developing countries.

The Bill and Melinda Gates Foundation (BMGF) Financial Service for the Poor Program has shown that digitally connecting poor people to financial service providers not only drives financial inclusion, but also benefits poor people at a cost within their means[2].

The implementation of innovations aimed at driving digital financial services recognises that different countries will require different strategies and tailored interventions. However, progress with regard to digital financial inclusion could be measured with a standard set of indicators across different countries. To achieve this objective, The Financial Inclusion Insights Survey (FII) was developed and has been implemented since 2013.

The Financial Inclusion Insights Survey (FII)

FII was a research program funded by the Bill & Melinda Gates Foundation, in order to build a meaningful knowledge base about how the financial landscape is changing across eight countries in Africa and Asia. These are: Bangladesh, India, Indonesia, Kenya, Nigeria, Pakistan, Tanzania, and Uganda. This information will guide policy interventions and identify pathways for the poor to gain the financial tools they need to improve their economic circumstances. FII reported to BMGF annually on several key indicators of digital financial services uptake and use. These included:

- Mobile phone access;
- Mobile phone ownership;
- Ever used:
  - Mobile money;
  - Any formal financial service;
  - Any digital financial service.
- Account ownership of:
  - Mobile money;
  - Any formal financial service;
  - Any digital financial service.
- Past three month usage of:
  - Mobile money;
  - Any formal financial service;
  - Any digital financial service.
- Digital account owners who have ever made an advanced transaction.

Note: Some of these indicators were tracked by classifications across below $2.50 poverty line[3], rural, women, and a combination of these three.

We decided to focus on digital financial inclusion indicators for the pilot surveys, not only because of the current focus on digital financial inclusion in the markets that were identified for the pilot survey. Also because the FII survey had recently been conducted in these markets and, therefore, offered a unique opportunity to use the findings as a benchmark with which to compare the pilot survey results.

2 Bill Melinda Gates Foundation, 2013

3 Using the poverty probability index (PPI) developed by Innovations for Poverty Action (IPA)

4. Objectives

Examples could be: low-income adults without access to mobile phones and adults with low levels of literacy.

Key objectives of the pilot surveys were:

- To determine whether short-form surveys can be used to populate the Financial Services for the Poor (FSP) Digital Financial Services (DFS) indicators and produce a nationally representative result in eight core markets. These are: Kenya, Nigeria, Tanzania, Uganda, Bangladesh, India, Indonesia, and Pakistan
- To determine the optimal mix of different data collection methods and the most effective statistical modelling approach. This must be the one that can be applied to the data to produce reliable FII indicators in the different markets at a lower cost than that of the FII survey.

5. Data Collection Methodologies

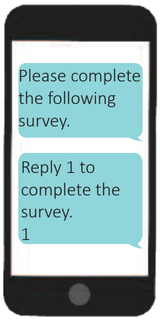
Sampling for the FII surveys is done in a way that collects data nationally representative of adults 15 years or older. As a point of departure for the pilot surveys, it was assumed that a single short-form mobile-based data collection mode would result in certain segments of the population being excluded. Examples could be: low-income adults without access to mobile phones and adults with low levels of literacy. The pilot data would not, therefore, be nationally representative and comparable to the FII data. To overcome this problem, it was decided to test a mixed-mode approach for data collection. Although not necessarily the cheapest approach, it was a trade-off between price and data quality.

In choosing the data collection approach for the pilots, the aim was to determine:

1. The optimal mix of data collection modes, ie the most cost-effective combination of modes to implement and still produce nationally representative data.
2. The optimal sample size for the mixed-modal approach, ie the sample size that would give the desired level of precision and reliability of the data.
3. The optimal survey design, ie would it be necessary to do a detailed upfront design that includes all modes, or would it be possible to start with a single, low-cost mode, assess the achieved sample, and then fill the gaps?
4. Whether data collected through mobile-based modes represent the mobile universe, ie everyone who has access to a mobile phone, in spite a non-probability sampling approach.

The mechanics of different data collection modes

SMS surveys are mobile-based and use the short message system (SMS) of the phone. Potential participants are invited to participate in the survey via an SMS invitation. A positive response to the invitation triggers the survey to begin. The respondent has to self-complete the questions that follow. Each response triggers the next question, until the survey is complete. Similar to a text message conversation.



Mobile web surveys use the internet through mobile phones. Invitations to participate are typically sent out via SMS, email, or as a pop-up, either in the in-the-web browser or in its applications. The invite contains a link to the survey and clicking the link results in the survey opening in the web browser. The survey is then self-administered.

Face-to-face in-person interviews (FTF) are interviewer-administered surveys. For financial inclusion surveys, interviewers typically visit the homes of potential participants. Once the interviewer has gained consent to conduct the survey, it is administered. Responses are captured either on paper (PAPI) or on an electronic device, such as a tablet or laptop (CATI)

The benefits of each mode of data-collection

SMS self-complete surveys and **Mobile web self-complete surveys**

- Affordable;
- Quick;
- Removes inteviewer effect;
- Minimises social desirability bias;
- Easier to reach people who are harder to reach using traditional methods e.g. people living in gated communities;
- Lower cognitive burden on respondent;
- Less time demanding on respondent;
- Easy and affordable to incentivise.

Interviewer administered computer aided telephonic interviews (CATI)

- Affordable;
- Allows for lower literacy levels;
- Greater flexibility in survey length then mobile surveys.

In-person face to face interviews

- Allows for collection of a wide range of data;
- Widest reach.

The limitations of each mode of data-collection

SMS self-complete surveys

- Limited to few questions;
- 160 character limit;
- Non-probability sample

Mobile web self-complete surveys

- Limited to few questions;
- Very limited reach.

Interviewer administered computer aided telephonic interviews (CATI)

- Limited to fewer questions than face to face;
- Sampling doesn't control for skews in phone ownership;
- Human resource and overhead costs drive cost.

In-person face to face interviews

- Human resources and detailed logistical requirements drive the cost.

6. Pilot Survey Design and Sampling Approach

A traditional cross-sectional survey design was employed as this is still regarded as the gold standard for the collection of demand-side/consumer data.

Self-administered mobile-based surveys are significantly more cost-effective than interviewer-administered surveys. However, they are also more likely to exclude certain sections of a country's adult population. This is the universe for demand-side surveys, such as the FII survey.

To be included in a mobile self-administered survey, individuals must:

- Have access to a working and active mobile phone;
- Have their mobile number included on a database from which the sample is drawn;
- Be literate enough to take part in a self-completion survey;
- Respond.

A mixed-modal data collection survey design offers the opportunity to overcome the limitations of mobile-based self-administered surveys.

A traditional cross-sectional survey design was employed as this is still regarded as the gold standard for the collection of demand-side/consumer data. The innovation made to the standard financial inclusion surveys and the FII-type design, was that data was collected using a combination of mobile-based modes: namely SMS, mobile web, and mCATI.

To be included in a mobile self-administered survey, individuals must:

- Have access to a working and active mobile phone;
- Have their mobile number included on a database from which the sample is drawn;
- Be literate enough to take part in a self-completion survey;
- Respond.

A mixed-modal data collection survey design offers the opportunity to overcome the limitations of mobile-based self-administered surveys.

A traditional cross-sectional survey design was employed as this is still regarded as the gold standard for the collection of demand-side/consumer data. The innovation made to the standard financial inclusion surveys and the FII-type design, was that data was collected using a combination of mobile-based modes: namely SMS, mobile web, and mCATI.

6.1 SMS Surveys

In terms of mobile-based surveys, SMS surveys form the lowest common denominator in terms of market familiarity and access. In several ways, however, they differ from the traditional face-to- face method of data collection.

With SMS surveys, the interview is limited in terms of the number of questions that can be asked (approximately 20 questions). Each question is limited to 160 characters, and the interview is fully self-completed (ie there is no interviewer present to clarify any issues of comprehension). These issues had significant questionnaire design implications for the pilot surveys. As questions could not simply be taken from the FII questionnaire, the questionnaire had to be redesigned. This necessitated pre-testing of the new questionnaire. This was done not only to ensure respondent comprehension but also to ensure that the new questions still measured the FII indicators validly and could, therefore, be compared with the data from the FII survey.

These issues had significant questionnaire design implications for the pilot surveys.

In addition, the census-based sampling frame (usually provided by the national statistics office of a country) that is used for nationally representative face-to-face financial inclusion surveys could not be used. Nor could area-based stratified probability sampling methods be employed for the SMS surveys. A new sampling frame had, therefore, to be considered. The reason for this was that any mobile-based sampling frame would skew the sample and would, at most, be representative of the mobile universe rather than of the adult population. These skews needed to be corrected statistically. It was, therefore, also necessary to develop a statistical modelling approach to account for the skews introduced by the survey design.

In order to obtain the sampling frame, the services of a sample provider were employed for the SMS surveys. To serve as the sample frame, database sample providers obtain and combine mobile subscriber databases from the different local mobile network operators (MNO) in a country.

To ensure demographic and geographic representativeness, proportional quota sampling was used for the SMS sample, using age, gender, residential area (ie province/state), and urbanisation.

The subscriber database provided information on whether subscribers had participated in at least one mobile survey, be it SMS, mobile web, or CATI, before (active subscribers) or not (inactive subscribers). In attempting to increase the response rate, this variable was also introduced for quota sampling but in this case a 50/50 quota approach was used, except for in Pakistan. In Pakistan, there was a significant skew towards the subscribers in the database as mobile surveys were fairly new in this market. The vast majority of respondents for the survey, therefore, came from the inactive database.

After dividing the database into cells based on the quota variables, simple random sampling was done within each cell and an invitation to participate in the survey was sent via SMS.

The limitations of this sampling approach included:

- Not all MNOs co-operated equally in all markets;
- Subscriber's profiling data not being available for the entire sample frame;
- Low (and slow) response rates.

6.2 Mobile Web Surveys

Mobile web surveys are even more limited in reach than SMS surveys are. They introduce even more deviations from nationally representative data. They not only require potential respondents to have access to a smart phone, but also to have the capability to access the internet and enough data to complete the survey on the internet. This escalates the need for a statistical model to correct these skews after data collection.

Similar to SMS surveys, mobile web surveys challenge questionnaire design, in that the questions have to be short because of the limiting screen size of the phone.

Similar to SMS surveys, mobile web surveys challenge questionnaire design, in that the questions have to be short because of the limiting screen size of the phone. The ease with which a respondent can opt out of a web-based survey further necessitates limiting the survey to a few engaging questions.

For the purpose of the pilot surveys, mobile web surveys were implemented in markets where the sample provider was not able to provide the SMS sample frame. This was for reasons such as MNOs being unwilling or reluctant to share their subscriber databases. In these markets, third-party partners with verified smart phone user samples, were used to facilitate mobile web surveys.

Third-party partners tapped into existing panels, eg smart phone users who have opted to be surveyed, and also sent invitations to smart phone users who were not part of a panel (inactive subscribers). Quality checks were employed to ensure that third-party partners sent adequate numbers of invitation messages to facilitate mirroring the sample methodology used in the SMS survey, in terms of the spilt of the active/inactive sample.

6.3 Random Digit Dialling (RDD) mCATI Surveys

RDD is a method used for selecting respondents by generating phone numbers at random; there is no sample frame. In theory, these numbers represent the entire universe of mobile numbers in a market. This provided the opportunity to test whether this approach produced good representation, without the application of quotas. Furthermore, with MNOs being reluctant or unwilling to provide their complete subscriber databases, it was difficult to assess how well sample frames provided by database providers represented the mobile universe. The use of RDD as a sampling method offered a means of overcoming this limitation.

The inclusion of mCATI surveys also provided the opportunity to address some of the sample skews introduced by the self-administered survey methodologies. Examples are, including respondents with very low levels of literacy who, by nature, would be not able to participate in a self-administered survey.

The process of sample generation for RDD involved four key steps:

1. All the information required to implement RDD, such as country dialling codes and MNO-specific prefixes, was obtained and verified a week prior to the launch of each survey.

2. A systematic number generation syntax was used to back-fill the remaining digits to complete the full length of the phone numbers. Duplicates were removed to ensure complete randomness when the sample was generated.

3. A platform validating the generated mobile phone numbers in the randomly derived sample was used to identify whether the number was real and attributed to a live subscriber.

4. Internal CATI managers randomly cycled through the sample for survey implementation.

In spite of its advantages, the RDD approach was not used for the selection of the SMS survey sample as it had inherent limitations for these surveys. MNOs allowed respondents to respond to the survey at no cost and facilitated the payment of incentives to participants by means of airtime. To ensure this, survey participants had to form part of the MNO's database, any numbers generated and surveyed outside of the existing database would have been at a cost to the respondent so they would not have received the incentive.

Country	Mobile self-complete (SMS/Web)	Mobile interviewer administered (RDD CATI)
Kenya	1 559 SMS	1 081
Nigeria	3 701 SMS	3 049
Tanzania	1 429 SMS	1 089
Uganda	1 362 SMS	1 000
Bangladesh	1 588 Web	1 534
India	3 208 Web	3 010
Indonesia	2 103 Web	2 052
Pakistan	2 286 SMS	1 006

Table 6. Achieved Sample Size per Country

Gender pilot's box

The gender pilot is an extension of the DFS pilot. The study was implemented in several ways that differed from the DFS:

- There were no existing indicators identified to test the method.
- i2i convened a sequential series of expert engagements to identify useful gender-relevant indicators. To test the suitability of the SMS-led mixed-mode data collection approach. This process led to a narrowing in on testing how well SMS-limited questions can cover a diverse set of headline indicators.

We covered twenty gender-relevant indicators, including:

- Informal Saving Usage status classification;
- Own bank account;
- P12M money taken from account without permission;
- Set own major savings goal status classification
- Influence on choice of own goal;

Gender pilot's box continued

- Decide on use of land;
- Decide on number of children;
- Work for pay;
- Decide own spending;
- Family allow to work for money;
- Spouse income earner classification;
- Respondent allows spouse to work for money.

The indicators were then translated into SMS-compliant questions.
This study was implemented in four markets – three markets in East Africa: Kenya, Tanzania and Uganda, as well as Pakistan.

Cognitive interviewing was conducted in all four markets simultaneously, to ensure that the question wording was optimised for all contexts, in both English and the main local language. This was the basis used for translation into other languages, all while ensuring the 160 character limit was observed.

Interviews were conducted across SMS and face-to-face modes in all four markets, as well as CATI in Pakistan. Final samples achieved were:

Country	SMS	CAPI	CATI
Kenya	3 700	3,000	
Pakistan	1,008	3,638	3,404
Tanzania	3,050	3,026	
Uganda	3,154	3,053	

Table ??. Total sample size for each survey

To emphasise, in contrast to the DFS pilot, face-to-face data was collected as part of this pilot's work. In contrast to the DFS pilot, common underlying questions were used across the different data modes.

Furthermore, there is no accepted model on which to calibrate the MRP reference values for the indicators collected. The weighted indicator values collected in the CAPI portion of the study were assumed to be accurate and were used to assess the performance of the calibrated MRP model under various CAPI sub-sampling simulations.

The same stratification variables used in the DFS study were implemented, and relationship status was added as an additional stratification variable. This was seen as relevant to the domain of gender-relevant indicators.

The SMS survey had, by its nature, to take a far more direct approach.

7. Questionnaire Design

Because of the limitations of the SMS data collection method, a different approach to asking the questions to derive the indicators had to be taken.

A single version of the shortened SMS/Web and mCATI indicator questionnaires was piloted across all markets, with only slight variations based on market context. For example, relevant regions and institutional examples.

Because of the limitations of the SMS data collection method, a different approach to asking the questions to derive the indicators had to be taken. SMS questions can only use 160 characters at most, while the survey was limited to approximately 20 questions. On the other hand, the face-to-face FII survey relied on the construction of indicators based on the responses to a number of questions. The SMS survey had, by its nature, to take a far more direct approach.

Once the questionnaire had been designed, cognitive testing[4] was conducted to gain insight into respondents' understanding of the questions. This also aimed to determine whether the more direct approach to the indicator question produced findings that could be compared to that of the FII survey.

The questionnaire comprised 32 questions. On average, each respondent answered 27 of these (Appendix A) and was structured as follows:

Classification questions:

- Year of birth;
- Gender;
- Region;
- Urbanicity.

Digital Financial services indicators:

- Banking usage and registration, ever used and active usage;
- Mobile money usage, registration and active use;
- Microfinance usage, registration, active use and service offering;
- Use of advanced financial services.

Poverty probability indicators (PPI):

- Varied household asset and consumption measures by market or regions within markets.

Post-stratification[5] variables:

- Access to electricity;
- Highest level of education achieved;
- Reading literacy;
- Phone ownership and tenure.

[4] Bridging the gap between researchers and respondents (https://indd.adobe.com/view/02fd6747-8796-46d0-b9bf-e4ee17bd1984)

[5] See paragraph 8

Innovations for Poverty Action (IPA) assisted the i2i team in designing a short-form, mobile-ready PPI individualised for each country. The need for a short-form PPI was crucial to this work, as the DFS indicators are disaggregated by poverty levels. Unfortunately, data to develop the PPI was not available in Bangladesh and Indonesia and we, therefore, had to proceed with the pilots in these markets without the PPI measurement.

Overall the short-form PPI worked well, although there were some unexpected challenges. Two key problems arose:

1. Most notably, the PPI for India was designed to be unique per **state** in order to be more context sensitive. This split measurement led to the questionnaire being long and cumbersome. It also did not take into account the sample size per state, which was, in some cases, too small for meaningful analysis. For future studies, it is recommended that **regions,** rather than **states** be used for PPI development in India. This should not only allow for context specificity but should also provide sample sizes to facilitate further analysis.

2. In Kenya, the PPI questions included a question on the kind of floor material used for housing. This question had two response options – 'natural' floors was awarded a zero score and the option for 'other' floor types was awarded a score of 21. No respondents in either the SMS or the CATI surveys chose the option for 'other' floor types. This meant that all respondents received a score of zero in terms of this question. This was a significant deviation from the national dataset used to build the PPI, where 46.5% of respondents indicated that they had 'other' floor types. It is, therefore, recommended that this be further investigated for future PPI measurement in Kenya.

It should also be noted that there were challenges in obtaining a reliable measure of the urban/rural classification in self-report surveys. This is because the claim of being in an urban or rural area is based on the respondents' understanding and interpretation of these terms. Face-to-face surveys typically use location data and determine urbanicity based on the definitions of the national statistics office of a country, as opposed to relying on respondent claims.

It should also be noted that there were challenges in obtaining a reliable measure of the urban/rural classification in self-report surveys.

[6] https://www.povertyindex.org/about-ppi

[6] Kenyan Demographic and Health Survey

Gender pilot's questionnaire design box

The questionnaire was significantly longer than the DFS one and comprised approximately 50 questions, (Appendix B). It was structured as follows:

Classification questions:

- Year of birth;
- Gender;
- Region;
- Urbanicity;
- Reading level;
- Relationship status;
- Ownership and control over phone.

Gender-relevant indicators:

- Independent phone use;
- Mobile money use;
- Informal financial use;
- Bank account use;
- Bank account permission requirement;
- Money taken out of account without permission;
- Major savings goal decision-maker;
- Extent of influence on choice of major savings goal;
- Land ownership;
- Control over land use;
- Number of children decision-maker;
- Earnings payment frequency;
- Extent of influence on how own earnings are spent;
- Permission to work;
- Possibility to come up with 1/20th GNI for sudden need;
- Spouse earnings frequency;
- Permission for spouse to work;
- Relationship to main household earner;
- Migrant worker in immediate family;
- Poverty probability indicators (PPI);
- Varied household asset and consumption measures by market or regions within markets.

Post-stratification variables:

- Main language spoken;
- Highest level of education achieved.

8. Implementing the Pilots

GeoPoll was contracted to conduct the pilot surveys in the different markets. Table 8 provides a summary of implementation dates.

Country	SMS/Mobile web (MW) survey	mCATI survey
Bangladesh	15 April – 19 May 2019 (MW)	14 April – 9 May 2019
India	10 January – 30 January 2019 (MW)	18 January – 18 February 2019
Indonesia	22 April – 21 May 2019 (MW)	24 April – 9 May 2019
Kenya	26 November – 11 December 2018 (SMS)	3 December – 18 December 2018
Nigeria	15 November – 18 December 2018 (SMS)	17 November – 5 December 2018
Pakistan	27 March – 10 May 2019 (SMS)	21 March – 13 April 2019
Tanzania	16 November – 16 December 2018 (SMS)	20 November – 13 December 2018
Uganda	21 November – 5 December 2018 (SMS)	23 November – 5 December 2018

Table 8. Pilot Survey Implementation Dates

9. Modelling and Poststratification

Multilevel regression with poststratification (MRP) is widely used across the social sciences to address the skews in non-representative data.

When FTF financial inclusion surveys are done, census-based sample frames are used. After these surveys have been implemented, any skews in the data are addressed by weighting the data. Respondent-level weights are calculated based on the sample frame to ensure that the weighted sample mirrors the sample frame in terms of the distributions. The weighting variables typically include demographic variables, such as age, gender, geographic, and urban/rural distribution.

This type of weighting cannot be done in the absence of a national sample frame. Therefore, the skews in the pilot data introduced by the sampling methodologies needed to be addressed with statistical modelling. For this purpose, multilevel regression modelling with poststratification was used.

Multilevel regression with poststratification (MRP) is widely used across the social sciences to address the skews in non-representative data. In layman's terms MRP involves[9]:

- Using stratification variables, dividing the data into 'cells' (eg rural-based females in the 30-35 year age group would be one cell if we consider urbanisation, gender, and age group as stratification variables). A key assumption of MRP is that, within each cell, the data in the sample can be considered to be representative of the target population. However, with non-representative data, some of the cells have no, or very few, observations and this assumption does not hold

- To overcome this problem, regression analysis is done on the total sample to determine the relationship between each of the stratification variables and the outcome. For purposes of the pilots, this would be the digital financial inclusion indicator variables. Based on the findings of the regression analysis, it is possible statistically estimate the outcome for the cells with few or no observations

- Poststratification involves estimating the outcome variable, based on the weighted averages of the cells in the target population

9.1 Stratification Variables

The goal with choosing stratification variables was to include all potential confounding variables. These are all the variables known to affect the outcome of the survey, as well as variables that could skew the sample.

For the pilots, confounding variables included gender, age, PPI, geographical region, and level of urbanisation. Because of the nature of the survey methodology, and based on a literature review that revealed that mobile-based surveys tended to be skewed towards respondents with a higher level of education and access to electricity, reading literacy, highest level of education level and access to electricity were added for post-stratification.

[9] For a full description of the MRP methodology reference

9.2 Calibration with Representative Data

As mentioned earlier, a key assumption of MRP is that within each cell, the data in the sample can be considered to be representative of the target population. This assumption, however, did not hold true for the pilot samples. Respondents to the pilot surveys were a very specific group. They:

(a) needed to have access to a mobile phone;

(b) had to be able to read the questions in the language administered; and

(c) had to be willing to take the time to respond to the survey.

The post-stratification data could, therefore, at most only be representative of the mobile universe, rather than of the total adult population of a country.

One way to correct this bias was to incorporate a small amount of representative data into the sample, allowing the model to calibrate the results to the adult population. For purposes of the pilots, this was done by including small samples of data from the county-specific FII survey (5% and 10% of responses from the sample, randomly selected) and repeat the MRP process. This made it possible to estimate the expected rate of digital financial inclusion for the adult population of a country.

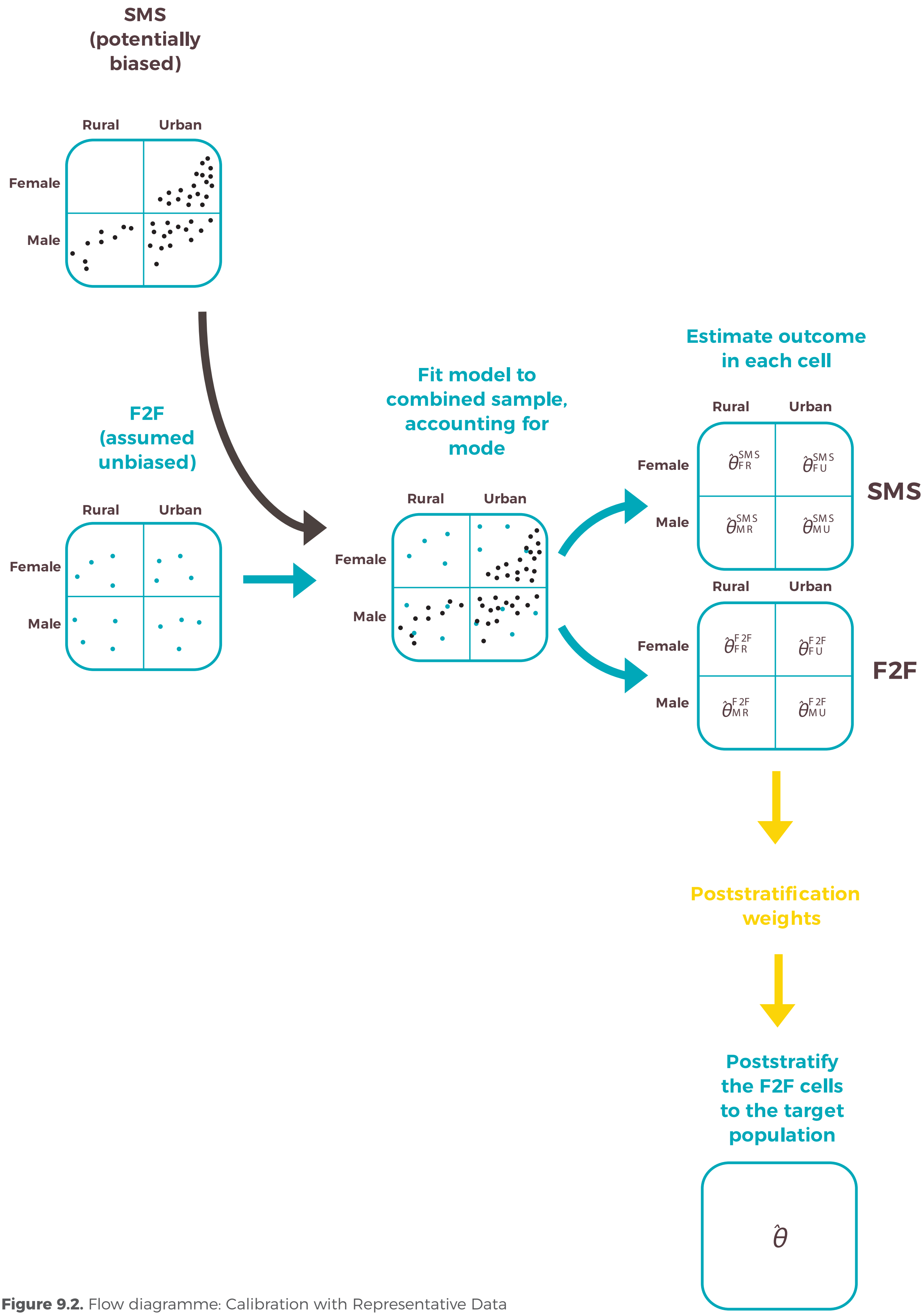


Figure 9.2. Flow diagramme: Calibration with Representative Data

9.3 Implementing MRP

MRP was done for each of the country datasets, using six different 'data mixes':

- SMS survey data only;
- SMS data and 5% randomly selected responses from the FII survey;
- SMS data and 10% randomly selected responses from the FII survey;
- SMS data and mCATI survey data;
- SMS data and mCATI data and 5% randomly selected responses from the FII survey;
- SMS data and mCATI data and 5% randomly selected responses from the FII survey.

Country	SMS/Web	CATI	Complete FII	FII (10%)	FII (5%)
Uganda	1 362	1 125	3 001	300	150
Tanzania	1 372	1 080	3 060	306	153
Kenya	1 559	1 080	3 129	303	156
Nigeria	3 701	3 021	6 042	604	302
India	3 207	3 008	48 027	4 803	2 402
Indonesia	2 103	2 047	6 060	606	303
Pakistan	2 286	1 005	6 000	600	300
Bangladesh	1 588	1 533	6 000	600	300

Table 9.3: Sample sizes per country

10. Results Summary

These were those residing in rural areas, women, and respondents with self-claimed poor literacy levels.

As expected, different modes of data collection produced samples that In general differed from the FII survey distribution in terms of respondent demographics. Table 10, which summarises this analysis for Uganda, serves as an example. The data also illustrated that mobile-based data collection methods produce samples that differ significantly from the mobile universe that was extracted from the FII databases. However, in all markets the mobile-based data collection methods resulted in reaching, although not in large numbers nor representative, populations that were originally expected to be unreachable. These were those residing in rural areas, women, and respondents with self-claimed poor literacy levels. Using mobile-based data collection methods, therefore, necessitated applying statistical modelling (such as the MRP approach) to achieve more representative samples.

	Uganda surveys			
	FII	**Mobile access universe**	**SMS**	**mCATI**
Age distribution				
15-24	32%	30%	33%	32%
25-34	24%	26%	36%	33%
35-44	16%	17%	14%	17%
45-54	13%	14%	9%	14%
55 +	15%	13%	7%	4%
Gender distribution				
% Female adults	52%	50%	45%	52%
Rural/urban distribution				
% Rural adults	73%	71%	66%	732%
Welfare distribution				
% Adults below poverty line	57%	51%	46%	47%
Literacy distribution				
Cannot do this at all	41%	36%	3%	7%
Very badly	14%	14%	3%	17%
Somewhat badly	16%	17%	6%	21%
Good	19%	21%	52%	41%
Excellent	10%	12%	36%	14%
Values in red indicate significant differences from the FII and mobile access universe				

Table 10: Uganda demographic distributions

10.1 Determining the Optimal Mix

To determine the optimal mix of data collection modes, pilot indicator values were compared with the FII indicator values for the different countries.

Figure 10.1 serves as an example to illustrate the comparisons done for the different countries regarding the FII indicators. The percentage of adults in a country actively using mobile money services was used.

The 95% confidence interval (Table 10.1) for the FII indicator values for the different countries was used to compare the pilot survey findings with those of the FII survey. This was regarded as the gold standard in terms of the estimation of the indicator values in a country. The following conclusions could then be reached:

- Using a small face-to-face sample from the FII survey, along with the SMS sample and applying a statistical modelling technique, such as MRP, produced the best estimation of indicator values. These are values that fell within the confidence interval for the FII indicator in question and this was found to be true across countries and indicators.

- Mixing all three modes (ie SMS, mCATI and face-to-face) was found to not to be necessary. The inclusion of a sample of face-to-face interviews was adequate for the MRP modelling to produce representative results.

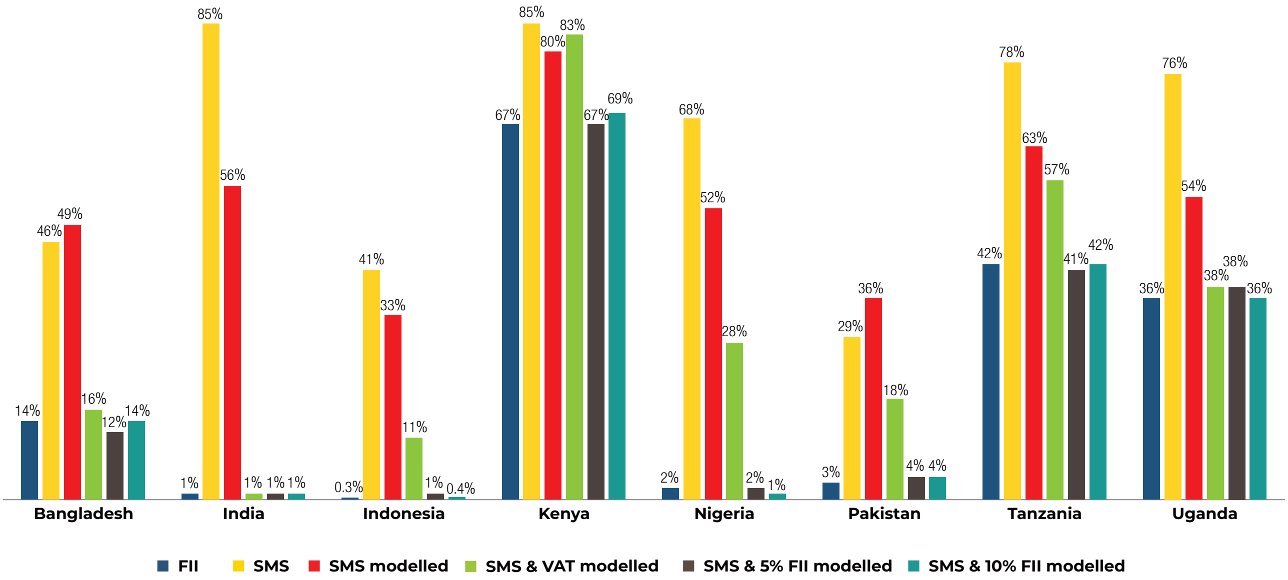


Figure 10.1: Active Mobile Money Usage

Country	Confidence Interval	
	SMS/Web	CATI
Bangladesh	13%	15%
India	1%	1%
Indonesia	0%	1%
Kenya	66%	69%
Nigeria	1%	3%
Pakistan	2%	4%
Tanzania	40%	44%
Uganda	34%	38%

Table 10.1: Confidence intervals by country

10.2 Mixed Modal vs. a Smaller FII Sample

Sub-group analysis was done with regard to the FII indicators. This was to determine whether it was necessary to use a mixed modal approach, or whether a smaller sample than the nationally representative FII sample of face-to-face interviews, could still produce accurate estimates of indicator values.

Figures 10.2 and 10.3 illustrate the types of comparisons which were done for the different countries. Respondent's **gender** were used as sub-groups and, once again, the **percentage of adults actively using mobile money services** as indicator.

Using the 95% confidence interval, (Table 10.2) for the FII indicator values per sub-group per country, the following conclusions could be reached:

- The FII survey conducted per country was nationally representative. One could assume that it likely represented the minimum number of face-to-face interviews to achieve this level of representativeness in terms of the available budgets. It was, therefore, not surprising to find that, once smaller samples of the data (ie either a 5% or a 10% sample of the FII survey) were used to estimate the indicator values for sub-groups, these were not reliable in estimating indicator values. Indicator values calculated for these samples fell outside the 95% confidence interval of the FII indicator in question.

- The strength of using the mixed-modal approach lies in it still producing reliable estimates of the indicator values at sub-group levels of analysis. Once this approach was used, indicator values calculated fell inside the 95% confidence interval of the FII indicator in question.

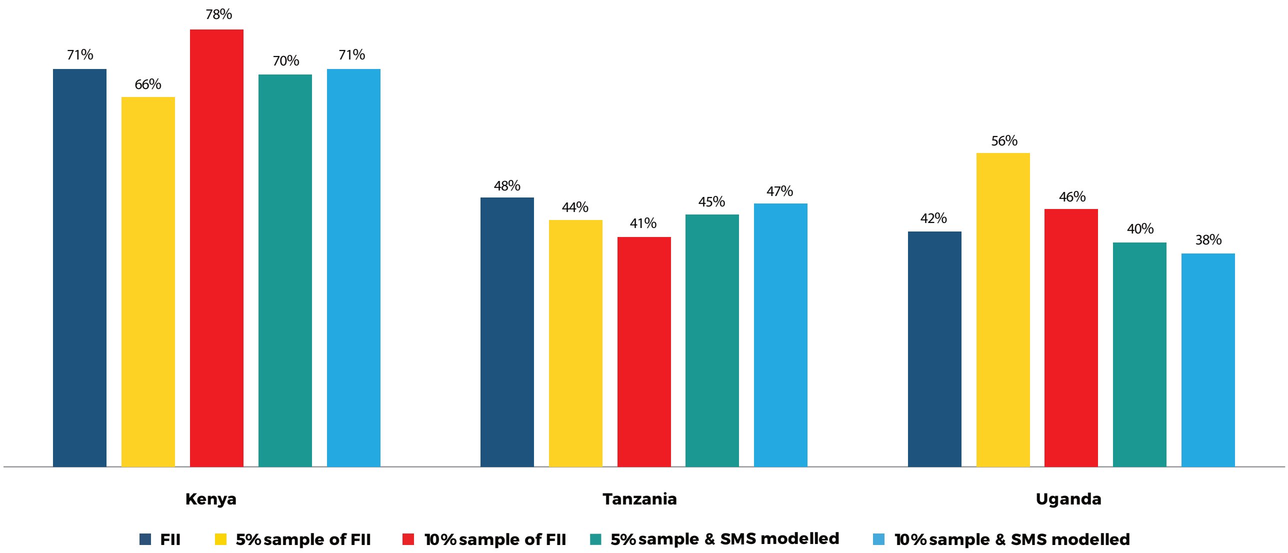


Figure 10.2: % of males actively using mobile money services

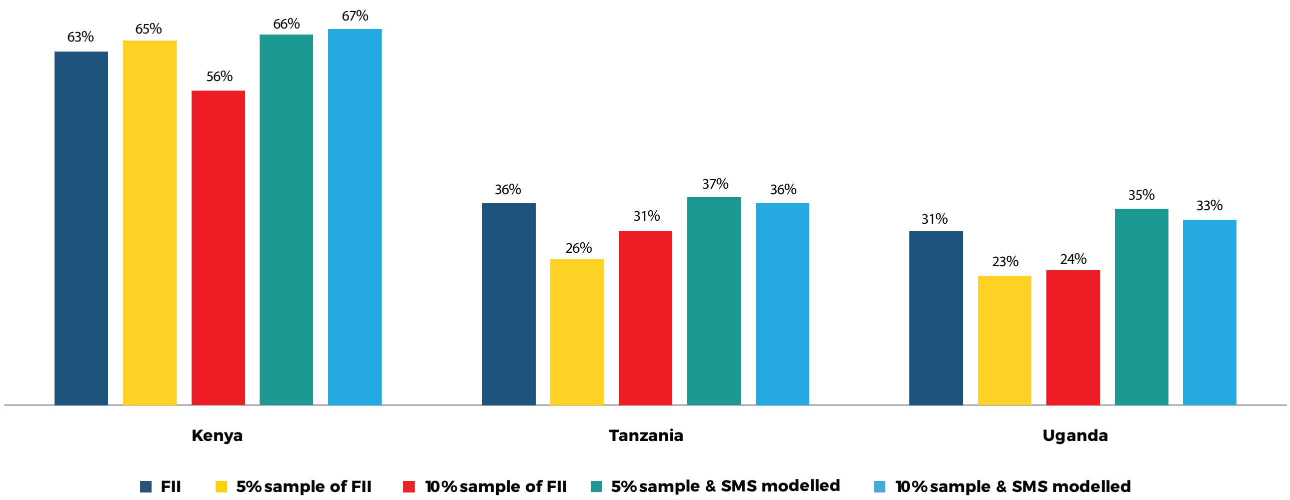


Figure 10.3: % of females actively using mobile money services

Country	Male Confidence Interval		Female Confidence Interval	
	Lower value	Upper value	Lower value	Upper value
Kenya	69%	74%	61%	66%
Tanzania	45%	51%	17%	36%
Uganda	38%	45%	28%	33%

Table 10.2: Confidence intervals by gender

Gender pilot's insights box

We had several diverse objectives with our gender pilot extension on the DFS work:

1. Apply the calibrated multi-mode data collection approach to a very different area of research that was less closely correlated with mobile phone ownership.
2. Cover a broader range of indicator construct types.
3. Further our knowledge of how to cost-effectively optimise the sample mix across the various available modes in each country for a targeted precision level.
4. Use the approach to integrate indicators reported in the DFS pilot, but not in the gender pilot.

The first two objectives were based on covering a different area of research and different types of indicators. This was more a study implementation challenge based on indicator identification and questionnaire translation. Both these objectives were successfully achieved through curating expert input. These indicators into covering and extending cognitive implementation of cognitive interviewing to each market, ensured that questions were contextually appropriate, while still meeting SMS question limitations.

We have identified the third objective as an area of research that could have significant future impact to guide sustainable data collection. This is because the data needs required to guide ongoing development and policy work continue to expand, while hard financial resource constraints, particularly amongst middle- and lower-income countries, are a continuous barrier. Development of this work on clear and specific guidelines for cost-optimised samples for data collection is, therefore, a global public good. Although this objective is currently unmet and requires further ongoing work, our experiences have shaped the appropriate parameters for future study.

As far as exploring performance of the overall calibrated MRP model in our gender pilot goes, our simulation results indicate that Calibrated MRP performs well under most scenarios. The cases where it does not appear to perform as well are correctable. For example, we found that the performance of Calibrated MRP deteriorates when the CAPI-only MRP estimate has high variance. This occurs when there are demographic groups that have little or no sample in the small CAPI dataset.

There are cases where the MRP estimate using only a single mode, performs extremely well on its own, perhaps due to chance. The second explanation for cases where single-mode MRP outperforms Calibrated MRP, occurs in cases where the MRP and raking estimates using only CAPI data differ substantially. For these cases, we really do not know whether the Calibrated MRP estimate or the single-source estimate is closer to the true rate of the indicator in the population.

When it comes to sample mix questions, we also found that oversampling poor rural women improved estimation, while oversampling non-mobile phone owners did not. A potential explanation for this observation is that, even though mobile phone ownership is highly associated with selection into the study, it may not be associated with outcomes. Poverty, urbanicity, and gender, on the other hand, may be associated with both outcomes and the selection process so they are more important to control for. For future studies, we recommend careful consideration of the demographic characteristics that would be associated with both outcomes and selection. Using a sampling scheme that ensures adequate representation of groups defined by these confounding variables, is also recommended.

To summarise, on the objective of cost-effective sample optimisation, the gender-relevant study has several limitations. One of these is that we do not know the true rate of each indicator in the target population and we had to estimate it (by raking applied to the CAPI dataset). This will be the case in any real-world application, unless benchmark data is available on the same indicators. For this reason, we believe future work should be done to perform a simulation in study for which the outcomes themselves are simulated and can be carefully controlled using a known data-generating process. If we have control over the data-generating process, we would know for certain what the true rate of the indicator is in the target population and would be better able to assess the performance of each estimation approach. The downside of taking this approach to simulation is that we might not be able to take all of the complexities of real-world data into account. However, we believe we have learnt enough through the course of this study to understand many of the issues in these datasets so that they can be built into the simulation. Another limitation, as previously mentioned, is that our simulation models might not have optimally modelled certain variables, such as the education variable in Kenya.

To better understand the overall observations around performance of calibrated MRP, as well as simulations using sub-samples biased towards particular population sub-groups, requires highly focused research with research design technical specificity. We have learnt from this study that illuminating sample optimisation is best served by research that has this as the sole specific research objective going forward. Based on these outcomes, suggestions on principles and specific research permutations to better inform cost-optimised multi-mode data collection include:

- Ensure that indicators selected already have very recently published estimates that are universally endorsed as accurate at all targeted levels of reporting sought.
- Identify priority sub-segments of the population required for accurate reporting to structure the research to test cost-optimised reporting modelling estimate simulations.

- Hypothesise sub-samples of the population thought to be more useful to plugging key sources of variation in accurate estimates. For example, where does the lower cost data collection mode tend to miss out on parts of the population that need a face-to-face collected components to best moderate that uni-mode modelled estimate through calibrated MRP? Is this a general mode effect that requires a balanced sample? Or is it particularly illiterate groups or rural dwellers or women, or a combination of attributes?
- Identify context-informed sampling approaches to optimally reach the priority sub-populations identified pragmatically.
- To inform cost-optimised sampling, develop accurate costing for the different sampling approaches to understand the overall cost of different sampling approaches. Further estimate the cost per targeted sub-group, given the success rate for derived success of obtaining the desired targets.
- Consider a revisiting research design across the data collection mode samples to revisit respondents with various data collection modes. To do this in order to better estimate mode effects for different sub-groups of the population and different indicator thematic areas and construct types.
- Ensure selection of post-stratification variables aligns as best as possible with broader domains appropriate to broad content focus. For example, phone ownership for DFS and relationship status for gender.

The fourth objective was identifying a technical approach that would integrate content from two separate surveys in an appropriate manner. To generate these estimates, we first had to develop a statistical framework that provides valid, efficient estimates of each indicator when one of the subgroup variables comes from a separate dataset. A requirement for the fusion procedure is that both models must be fit using the same poststratification variables. This was a relatively straightforward extension of the work and was assisted by the framework used for, and steps fulfilled, in the DFS approach.

11. Conclusion

The responsible use of this methodology provides opportunities for more sustainable indicator collection and tracking.

In conclusion, the use of a lower-cost, mixed modal sample, along with multilevel regression and post stratification does product nationally representative results. These were achieved for DFS indicators in eight core markets and, to achieve a good result a sample as small as n=150 surveys collected face-to-face, can be sufficient to produce good results. However, determining the optimal mix of data collection modes is highly context and content specific. Having existing benchmarks to compare with is a useful starting point. It is also impractical, as the aim is to collect new data more sustainably without the reliance on a large-scale face-to-face national survey with the indicators of interest. The success of this mixed modal method, therefore, requires a detailed understanding of what the mobile population in a market looks like and needs a reliable reference source for post-stratification.

Ultimately, our pilots did not extend to testing the use of mixed modes up front vs. collecting with the lowest cost mode first and then using face-to-face to top up on the gaps. We hypothesise that both will produce similar results. For the most practical solution, having a detailed understanding up front of the gaps that a mobile sample is likely to produce, using information on phone ownership from available supply and demand side sources, is likely best to have. This will allow for sample design that can allow for the limitations in the mobile sampling method and concurrent fieldwork.

The responsible use of this methodology provides opportunities for more sustainable indicator collection and tracking. Good reference data and a detailed understanding of the potential limitations to the sample that will occur using the mobile methodology is required.

References

Atkeson, L. R., Adams, A. N. & Alvarez, R. M. 2014. Nonresponse and Mode Effects in Self-and Interviewer-Administered Surveys. Political Analysis, Vol. 22, No. 3, Summer 2014, pp. 304-320.

Gelman, A., Goel, S., Rothschild, D. & Wang, W. 2020. High-Frequency Polling with Non-Representative Data. In Political Communication in Real Time: Theoretical and Applied Research Approaches. Dan Schill, Rita Kirk, Amy E. Jasperson, eds. New York, New York. Also available at: http://researchdmr.com/HighFrequency.pdf

Lau, C., Lombaard, A., Baker, M., Eyerman, J. and Thalji, L. 2018. How representative are SMS surveys in Africa? Experimental evidence from four countries. International Journal of Public Opinion Research. Oxford University Press, Routledge, 2017. Available at: http://www.stat.columbia.edu/~gelman/research/published/high_frequency_polling.pdf.

Ghitza, Y. & Gelman, A. 2013. Deep Interactions with MRP: Election Turnout and Voting Patterns among Small Electoral Subgroups. American Journal of Political Science, vol. 57, no. 3, July 2013, pp. 762–776. Available at: http://www.stat.columbia.edu/~gelman/research/published/misterp.pdf

Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, Version 2.15.0, 2016. Available at: http://mc-stan.org

Si, Y., Trangucci, R., Gabry, J. S. & Gelman, A. 2017. Bayesian hierarchical weighting adjustment and survey inference. 2017. Available at: https://arxiv.org/pdf/1707.08220.pdf

The Bill & Melinda Gates Foundation. 2013. Fighting poverty, profitably: Transforming the economics of payments to build sustainable, inclusive financial systems. Available at: https://docs.gatesfoundation.org/documents/fighting%20poverty%20profitably%20full%20report.pdf

Appendix A: DFS Questionnaire

Question in English

In what year were you born? Reply with a year like 1980.

Are you a woman or a man? Reply with 1 or 2.
1) Woman
2) Man

What province do you currently live in? Reply with the name of your province, like: Sindh.
Reply with 1 or 2:
1) My home is in a village
2) My home is in a city

What is the name of your village? Reply with the name of the village you live in

What city do you live in? Reply with the name of the city you live in.

Reply with 1, 2 or 3:
1) I have a bank account in my own name
2) I use somebody else's bank account
3) I don't have a bank account

Have you ever used a bank? Reply with a 1 or 2.
1) Yes
2) No

Have you used this bank account in the past 90 days? Reply with a 1 or 2.
1) Yes
2) No

Do you use a PHONE to buy or pay for anything? Reply with 1, 2 or 3
1) Yes - I use my OWN phone
2) Yes - I use SOMEBODY ELSE'S phone
3) No

To buy or pay for things on your phone do you:
1) Call only
2) Use a service e.g. Telenor Easy Paisa
3) Both - call and use a service
4) Other

Do you send money on a PHONE? Reply with 1, 2 or 3.
1) Yes - I use my OWN phone
2) Yes - I use SOMEBODY ELSE'S phone
3) No

Do you receive money on a PHONE? Reply with 1, 2 or 3.
1) Yes - I use my OWN phone
2) Yes - I use SOMEBODY ELSE'S phone
3) No

Have you received or sent money or bought anything on a PHONE in the past 90 days?
Reply with 1 or 2.
1) Yes
2) No

Do you have an account with a microfinancer/cooperative? Reply with 1, 2 or 3:
1) I have an account in my name
2) I use somebody else's account
3) I don't have one

Does the microfinance / cooperative offer any of the following : savings, money transfer, investment or insurance? Reply with 1, 2 or 3
1) Yes
2) No
3) Don't know

Does the microfinancer / cooperative have: a card for a cash machine or buying things, transfers without cash, a website or phone app?
1) Yes
2) No
3) Don't know

Have you used this account at a microfinancer / cooperative in the past 90 days? Reply with a 1 or 2.
1) Yes
2) No

Accounts may be used to save or borrow money, buy things, pay bills, or get insurance. Have you ever used your account for any of these things?
1) Yes
2) No

Have you ever used your accounts to receive wages or to receive money from government? Reply with a 1 or 2.
1) Yes
2) No

Have you ever used your account to pay taxes, fines and government fees or to pay school fees? Reply with a 1 or 2.
1) Yes
2) No

Count all the adults and all the children who live together in your home. How many are there? Reply with the number.

Does your home have a gas connection?
1) Yes
2) Yes, through an extension
3) No

Does your home have an electricity connection?
1) Yes
2) Yes, through an extension
3) No

Is the drainage / sewage system in your house?
1) Underground drains
2) Covered drains
3) An open drain
4) There is no system in your house

What type of toilet do you have in your house?
1) Flush toilet, connected to public sewage, a pit or open drain
2) Dry latrine
3) No toilet

Is there a television in your home? Reply with a 1 or 2.

1) Yes
2) No

What is your highest level of education?

1) Primary school not complete
2) Primary school
3) High school/vocational
4) University / Diploma
5) Koranic / other

How well do you read in [Language of the survey]?

1) Cannot do this at all
2) Very badly
3) Somewhat badly
4) Good
5) Excellent

The phone that you are using right now, is it your own personal phone? Reply with a 1 or 2.

1) Yes
2) No

When did you get your first phone?

1) less than 3 months ago
2) between 3 and 6 months ago
3) between 6 and 12 months ago
4) longer than a year ago

How old are you? Reply with your age, like 38.

Appendix B: Gender Questionnaire

Question in English

Which language do you wish to proceed with?

1) English
2) Luganda

In what year were you born?

Are you? Reply with a 1 or 2.

1) A man
2) A woman

What district do you currently live in? Reply with the name of your district, like: Tororo.

Reply with 1 or 2. Currently, I live in:

1) A rural area
2) An urban area

How well do you read in Luganda?

1) Not at all
2) Poorly
3) Fair
4) Good
5) Excellent

Are you currently;

1) Married
2) Not married but in a relationship
3) Single - Not in a relationship

Who owns the phone you are currently using?

1) Your own phone/business/employer
2) Husbands/wives
3) Parent
4) Sibling
5) My children
6) Other man/woman

Can you use a mobile phone without anyone helping you?

1) Yes
2) Somewhat
3) No

Whose phone do you use to buy things, send money or get money?

1) My own phone
2) Share/use someone else’s phone
3) None-don’t use a phone to do financial things

Have you saved or borrowed money from savings groups like saccos or powesa in the past month?

1) Yes
2) No

Do you use a bank account?

1) Yes - my own
2) Yes - my spouse's
3) Yes - other family's
4) Yes - other
5) No

Do you use a bank account?

1) Yes - my own
2) Yes - my relative's
3) Yes - other person's
4) No

Do you need permission from someone to open a bank account?

1) Yes
2) No

In the past 12 months, was money taken from your bank, ATM, mobile phone or other account without your knowledge/permission?

1) Yes
2) No

Who decided on your current MAJOR saving goal?

1) Me
2) My Husband/wife
3) We Decided together
4) Others
5) I don't have a major savings goal

Who decided on your current MAJOR saving goal?

1) Me
2) My family/friends
3) I Decided with others
4) I have no major savings goal

When coming up with the MAJOR saving goal with others, your opinion is

1) Always accepted
2) Sometimes accepted/rejected
3) Always rejected
4) I don't consult

Do you own any land?

1) Yes - I own by myself
2) Yes - I own together with others
3) No - I don't own land

Who decides on how to use this land?

1) Me only
2) Family/friends without me
3) Decide together with others

In your relationship, who decided on the total number of children to have?

1) Me
2) Spouse
3) Decide together
4) Other
5) We do not talk about it

How often are you paid for your main job or business?

1) I don't work
2) Work unpaid
3) Daily
4) Weekly/Biweekly
5) Monthly
6) When I get work

Do you fully decide how your earnings are spent?

1) Yes fully decide on my own
2) Yes - but with help from someone else
3) No - someone else decides for me

Would your family allow you to work for money, if you got a job?

1) Yes
2) No

How possible is it for you to come up with UGX115,000 in 7 days for a sudden need?

1) Very possible
2) Somewhat possible
3) Not very possible
4) Not at all possible

How is your spouse paid for their main job/business?

1) They don't work
2) Work unpaid
3) Daily
4) Weekly/Biweekly
5) Monthly
6) When they work
7) Don't know

Would you allow your spouse to work for money if he/she got a job?

1) Yes
2) No

Who earns the most money in your house?

1) Me
2) Spouse
3) Other family member
4) Other
5) Don't know

Who earns the most money in your house?

1) Me
2) Other family member
3) Other
4) Don't know

Including you, currently how many people do you live with and eat with in your home?

1) 1
2) 2
3) 3
4) 4
5) 5
6) 6
7) 7
8) 8
9) 9
10) 10 or more

Who in your family works far away and sends money home?

1) No-one - we all live together
2) Me
3) My spouse
4) Other family

Which language do you mainly speak?

1) English
2) Luganda
3) Runyakole/Rukiga/Runyakitara
4) Lusoga
5) Iteso
6) Luo
7) Lugishu
8) Lugbara
9) Nkore
10) Alur
11) Other

What is your highest level of schooling?

1) None
2) Primary not complete
3) Primary complete
4) Secondary
5) Vocational
6) Diploma/Degree
7) Religious

Sustainable data collection for development: mixed modes and statistical modelling

Sanofi House, Second Floor,
44 on Grand Central Office Park,
2 Bond Street, Grand Central
Ext 1, Midrand

Tel: +27 11 315 9197
Fax: +27 86 518 3579
info@finmark.org.za
www.finmark.org.za