Can Remote Survey Methods Yield Nationally Representative Samples in LMICs?
A cross-national analysis of pandemic-era studies

Elliott Collins* Isabella Contreras† Savanna Henderson‡
Caroline Lamke§ Michael Rosenbaum¶ Shana S. Warren‖

October 17, 2024

*Elliott Collins is with Innovations for Poverty Action, Research Scientist (ecollins@poverty-action.org), 1701 Rhode Island Ave NW, Washington, DC 20036, USA.

†Isabella Contreras is with Innovations for Poverty Action, Research Associate (icontreras@poverty-action.org).

‡Savanna Henderson is with Innovations for Poverty Action, Program Manager (shenderson@poverty-action.org).

§Caroline Lamke is with Innovations for Poverty Action, Research Associate (clamke@poverty-action.org).

¶Michael Rosenbaum is with Innovations for Poverty Action, Research Manager (michaelr526@gmail.com), and is now an MS candidate at the University of Chicago.

‖Shana S. Warren is with Innovations for Poverty Action, Research Scientist (swarren@poverty-action.org).

Abstract

Remote survey methods are increasingly replacing face to face interviewing in low- and middle-income countries (LMICs) because they are faster and cheaper, but they may not yield representative samples suitable for estimating national or sub-national population statistics. To address this issue, we compare these remote methods to nationally representative household surveys. We examine (1) random digit dial (RDD) surveys, (2) phone follow-ups to prior face-to-face surveys, and (3) self-selected and self-administered social media surveys. A set of contemporaneous face-to-face surveys are analyzed for comparison. In all, we harmonize 31 datasets from nine countries, drawn primarily from the first two years of the COVID-19 pandemic.

We find that on average, most remote survey samples over-represent men, heads of households, and respondents who are younger, more educated, and urban. There are significant differences in the magnitude of these biases across countries and methods. Surveys drawing from nationally representative sampling frames often face higher marginal cost, but tend to recruit more representative samples, while the social media samples tend to deviate significantly from national averages. Ex-post statistical adjustment makes little difference. Most weighted averages were not substantially closer to national benchmarks.

We conclude with a discussion of when fielding surveys with significant selection might still be worthwhile, when and why we should expect sampling weights to be helpful, as well as the need for researchers to report the socio-demographic profile of recruited samples as critical context when interpreting results, rather than relying on *ex post* adjustment alone.

Statement of Significance

National surveys in low- and middle-income countries face high marginal costs and logistical challenges, in part because they rely heavily on in-person, face-to-face methods to recruit and interview respondents instead of mobile phone networks or internet platforms. Rapidly increasing access to mobile phone and internet networks, combined with the constraints imposed by the COVID-19 pandemic, has raised the question of whether researchers can now rely on these remote methods. This study provides evidence on the feasibility of recruiting nationally representative survey samples in lower income countries using phone and internet surveys. We harmonize 31 survey samples collected across nine countries to provide insights into the strengths and limitations of these approaches. The findings reveal significant differences in who gets recruited and the ability to adjust for bias by re-weighting samples. The study emphasizes the importance of starting from a credible sampling frame and the need to consider specific research objectives when constructing post-stratification weights. We conclude that recruitment via mobile networks or internet platforms may be appropriate for some objectives, but still face significant drawbacks. The study offers evidence for researchers balancing sampling bias against practical constraints and insights for practitioners seeking to interpret data from phone and internet surveys.

1. Introduction

Nationally representative surveys play a central role in social science and public health research as a way to estimate important population statistics. In high-income countries, these surveys are often conducted using "remote" survey methods, in which individuals are recruited or interviewed using phone or internet platforms rather than in person. While these methods can have several significant advantages over in-person surveys, their adoption has been much slower in low- and middle-income countries (LMICs), where relatively low phone and internet coverage raises concerns about their ability to produce nationally representative samples. During the COVID-19 pandemic however, dozens of national surveys were fielded in LMICs using remote recruitment strategies or interview modes, often simultaneously in the same country. This industry-wide push offered an opportunity to compare different surveys in the same context, with similar topics, timing, and target population. Some studies re-purposed samples with phone numbers previously recruited through face-to-face methods, including samples from recent national Living Standard Measurement Surveys (LSMS). Others relied on random digit dialing (RDD) or social media recruitment.

This paper aims to quantify the extent of these combined errors in a wide range of pandemic-era survey samples in order to broadly characterize the conditions under which remote surveys can produce nationally representative samples in LMICs. Here we define "selection bias" in a given population statistic as the difference between its true population value and the estimated sample statistic, derived from a sample in which individuals were selected with unequal probability. Concern here focuses on sampling methods in which recruitment method and interview mode systematically over- or under-represents certain groups. By this definition, we take selection bias to comprise the sum of *coverage error*, in which some members of a population are left out of a sample frame entirely with no chance of selection, and *nonresponse error*, in which the individuals in sample frame have positive but unequal probability of selection.

Since the true value of national population averages are unobserved, we rely on official, largescale nationally representative in-person surveys to provide benchmark estimates. We then take the differences between benchmark estimates and the estimates from a national survey dataset as a measure of selection bias. We compare the benchmark estimates to both weighted and unweighted averages from each sample, giving us a look both at the types of individuals recruited and the extent to which statistical adjustment renders each sample "representative".

Our analysis compares 31 survey samples from four multi-national survey projects using different survey designs, each of which has a different combination of recruitment strategy and interview mode. These projects were chosen because they ask the same questions in the same countries over roughly the same time period.

The first project is a set of nine phone surveys using RDD to recruit sample conducted by Innovations for Poverty Action (IPA) between April and November of 2020. The second project is a set of phone surveys conducted by the World Bank drawing from call lists constructed from previous face-to-face LSMS samples. The third project used self-administered online surveys with samples recruited through social media, conducted by Facebook Research and Carnegie Mellon University beginning in April 2020 and ending in mid-2022.

Finally, in order to compare these remote survey projects to a contemporaneous face-to-face survey, we include national samples from the Afrobarometer survey program, which continued to conduct face-to-face interviews where lockdown policies and public health guidelines were permitted. Despite relying on very different sampling frames and recruitment strategies, each study took the entire nation as their population of interest, validating their comparison with the same national benchmark values.

Section 2 begins with a brief review of the potential and limitations of recruiting and interviewing national samples via remote methods in LMICs. Section 3 describes the methods employed in the included samples and briefly describes the data used. Section 4 presents the results of our main specification for several key variables. We focus on differences in five demographic variables, including gender, urbanicity, age, and educational attainment, followed by rates of pre-pandemic employment. Estimates of selection bias in gender, urbanicity, and age are meant to characterize which sub-populations are most often over- or under-represented, while educational attainment and employment serve as examples of how these methods might offer biased estimates of more pressing policy-relevant statistics.

2. Remote Surveys in LMICs

Researchers in high-income countries have long taken advantage of the lower costs and efficiency of phone surveys to collect national health and demographic data for decades (Groves 2011; Kempf and Remington 2007), encouraged by the near universality of landline telephones and, more recently, mobile phones in households (ITU 2020a). Pandemic restrictions placed constraints on in-person research just as government and multi-lateral agencies began asking for data to understand the socioeconomic impacts of the pandemic on communities in LMICs, rapidly expanding interest in remote methods in these regions. Phone-based surveys have the advantage of being able to produce data quickly, at high frequency, and often at lower cost than face-to-face surveys (Dillon 2012; Henderson and Rosenbaum 2020). High-frequency data collection can provide an accurate on-the-ground picture of events and behaviors, and has proven less subject to recall bias (Dillon 2012). These benefits are pronounced in times of conflict, disaster, or disease outbreaks, such as COVID-19 or the West Africa Ebola virus outbreak (Himelein and Kastelic 2015; Himelein et al. 2015). Researchers in LMICs must weigh these benefits against the expected level of coverage and non-response bias, which is likely to vary across populations and research areas.

Mobile phone access: Implications for survey coverage

Access to mobile phones, including networks and devices, has increased rapidly in LMICs, increasing opportunities for remote surveying (Silver et al. 2019), but available data indicates coverage, ownership, and usage vary within and between countries. Phone (2G network) coverage has reached 95 percent in some regions, while broadband (3G network) coverage has increased from 75 percent in 2015 to 90 percent in 2019 (GSM Association 2021a).[1] Despite this growth, coverage differences remain between rural and urban areas (Bahia and Delaporte 2020). As of 2021, evidence from the International Telecommunication Union (ITU) indicates that 100 percent of populations in urban areas of Africa receive 2G coverage compared to 84 percent of rural African populations. 3G coverage is yet more limited in rural Africa at 64 percent compared to 98 percent in urban Africa (ITU 2020b).

For each country included in this study, Figure 1 presents the reach of the mobile network operator with the widest 2G (phone) network coverage in pink, and the widest 3G (internet) network in yellow. These gaps are even more pronounced for 3G coverage (see the lower maps in Figure 1). Individuals living outside coverage zones likely systematically differ from others in the country on a range of socio-demographics indicators.

The ITU collects annual data on mobile phone subscriptions (i.e. SIM cards) per 100 people, finding that mobile subscriptions in developing countries rose from 68 per 100 people in 2010 to 103 in 2019 (2020b). Subscription data paired with coverage provides a general idea of access and ownership. For example, 97 percent of the Ethiopian population is covered by a mobile network, though there are only 53 mobile phone subscriptions per 100 people (ITU 2021). Table 1 presents subscription and ownership data for countries in our sample.

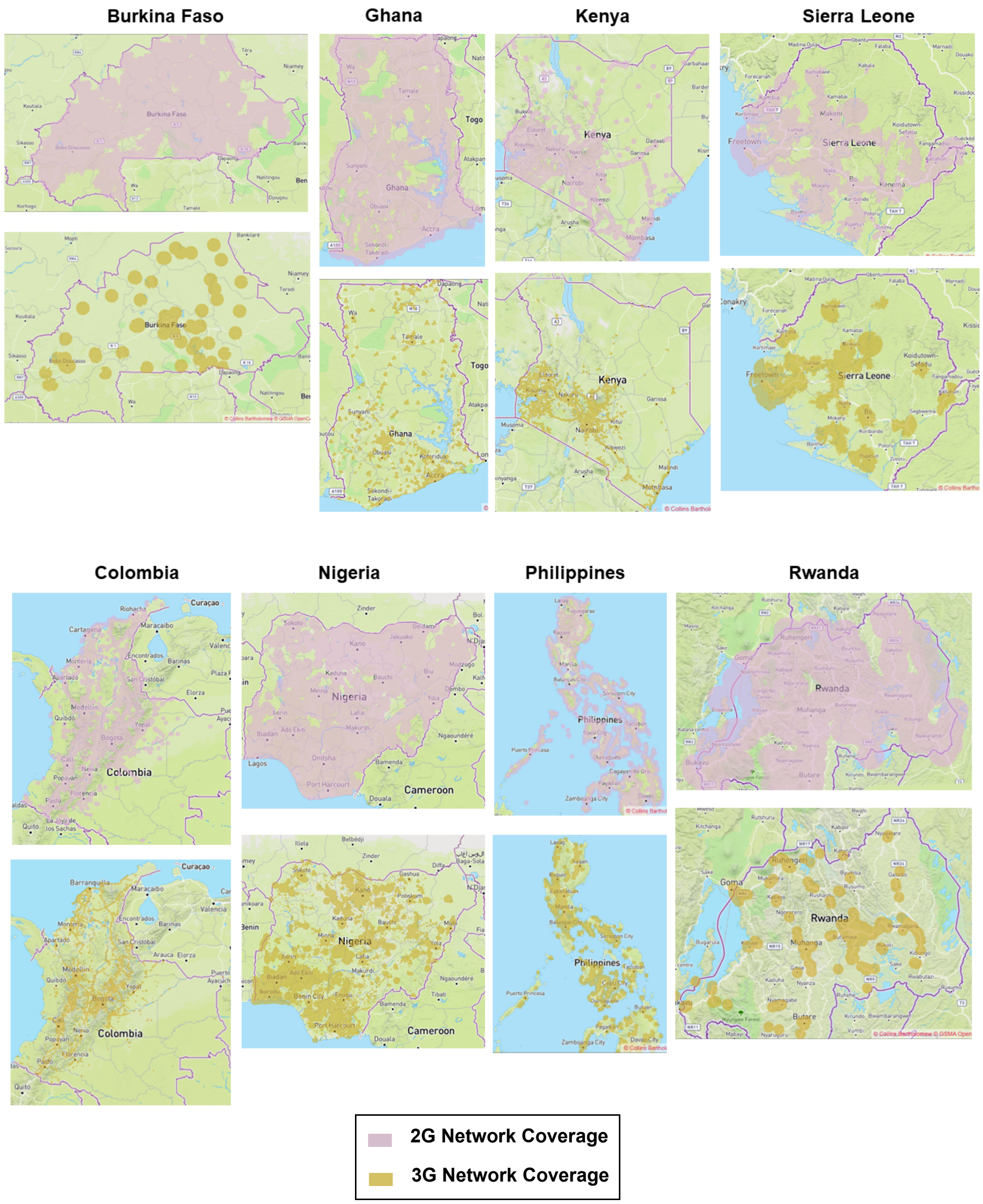
Although it has been decreasing in recent years, the gender divide in cell phone ownership adds to potential coverage errors (Rowntree and Shanahan 2020). Women are on average eight percent less likely to own a mobile phone than men[2], and use a smaller range of mobile phone services, such as SMS and internet-based uses, than men (2020).

Finally, gaps between mobile broadband coverage and actual usage induce further selection bias

[1] 3G coverage provides high-speed (broadband), quality and reliable network access to voice, data and video communications.

[2] Based on data from over 16,000 respondents from 15 LMICs. Countries included by region: Africa: Algeria, Kenya, Mozambique, Nigeria, Senegal, South Africa, Uganda; Asia: Bangladesh, India, Indonesia, Myanmar, Pakistan; Latin America: Brazil, Guatemala, Mexico.

Figure 1: Cellular (2G) and mobile internet (3G) coverage, by country



Notes: Coverage maps from GSMA, https://www.gsma.com/coverage/. Each map shows the coverage area of the most wide-reaching mobile network operator (MNO). Specific MNOs and year of data shown in each map: Burkina Faso, Telecel/ Planor afrique (2G), 2014 and Orange (3G), 2021; Colombia, Claro, 2022; Ghana, MTN (2G), 2014 and Tigo/Millicom (3G), 2019; Kenya, Safaricom, 2022; Nigeria, Bharti Airtel (2G), 2014 and MTN (3G), 2022; Philippines, Globe Telecom, 2010; Rwanda, MTN, 2022; Sierra Leone, Africell (Lintel), 2018.

Table 1: Mobile Phone Coverage and Ownership in Study Countries

Country	Population covered by a mobile-cellular network, 2021 (%)	Individuals who own a mobile phone (%)	SIM cards per 100 people, 2021
Burkina Faso	93	52 (2019)	112
Colombia	100	76 (2021)	146
Ghana	97	-	123
Kenya	97	47 (2019)	123
Mexico	97	71 (2015)	99
Nigeria	93	41 (2017)	91
Philippines	99	79 (2019)	143
Rwanda	100	-	81
Sierra Leone	93	-	98

in remote surveys that rely on internet access. As many as 47 percent of the population in LMICs do not use the internet, despite living within the footprint of a broadband network. Limited literacy and digital skills, and affordability of devices, data plans and other fees are key obstacles to internet use. The median cost of an internet-enabled phone as a share of monthly GDP per capita was 20 percent in 2019 for LMICs and the median mobile broadband connection costs 8.6 and 2.3 percent of GNI per capita in low- and lower middle-income countries, compared to only 0.4 percent in high-income countries. (GSM Association 2021b).

Response rates

While remote methods are frequently less expensive than face-to-face surveys, they face consistently lower response rates in LMICs. For example, the Demographic and Health Survey (DHS) program, a well-executed face to face survey, reports average response rates over 90 percent (Corsi et al. 2012). In contrast, phone surveys conducted in 18 LMICs between 2014 and 2018 achieved response rates ranging from 4.5 percent to 56 percent, depending on mode and sample type (Henderson and Rosenbaum 2020).

Among remote interview modes, surveys using CATI report the highest response rates (2020).

Live interviewers are often cost-effective compared to self-administered methods despite the added costs. Self-administered modes such as IVR and SMS may have lower response rates compared to CATI due to respondents' difficulties with the technologies, such as manipulating the keypad, issues with providing responses in the requested format, illiteracy, or the lack of refusal conversion, which is only available with live interviewers.

Sample selection

Researchers have found incomplete coverage, sampling bias, and response bias are significant obstacles in most LMICs. (Ballivian et al. 2015; Gibson et al. 2019; Lau et al. 2019; Leo et al. 2015; Pariyo et al. 2019; Velthausz et al. 2016). Nearly half of all mobile towers in sub-Saharan Africa experience at least six hours of power outage per day, with some towers being completely disconnected from the national electrical grid (GSM Association 2020). Lack of electricity, phone sharing, and multiple SIM ownership may all introduce coverage bias among individuals who still technically have access to mobile service. For example, respondents who have limited access to phones or electricity may technically be "covered", but may choose not to respond due to weak signal or limited battery life. The closely linked nature of these factors means that response bias (resulting from respondents choosing not to begin or complete a survey) cannot be consistently disentangled from coverage bias (resulting from the inability to access a respondent by phone). Instead of trying to empirically disentangle these factors, this paper aims to measure the overall magnitude of these combined biases.

3. Data and Methods

We consider the representativeness of survey samples constructed from four sources: LSMS call lists (from the World Bank's HFPS), RDD (from IPA's Research for Effective Covid Response (RECOVR) project), a social media platform (from Facebook), and face-to-face interviews (from Afrobarometer). In all, the paper includes data from 31 surveys implemented across nine countries from late 2019 through mid 2022 using a range of sampling and interview modes, as summarized in Table 2, all of which aimed for national representation. Facebook surveys had by far the largest sample sizes in the countries considered here, reaching over 1.93 million individual responses in total. Descriptions of each survey project, including a summary of how post-stratification weights were calculated for each, can be found in the Data Description appendix.

Table 2: Survey Project Summary

Sample	**Sampling**	**Interview Mode**	**Timing**	**Sample Size**
Burkina Faso				
Afrobarometer	Household	Face-to-face	Q4 2019	1,200
Facebook	Social media	Internet	Q2 2020 - Q3 2021	15,328
RECOVR	RDD	Phone (CATI)	Q2 - Q4 2020	1,371
World Bank	Call list	Phone (CATI)	Q2 2020	6,592
Colombia				
Facebook	Social media	Internet	Q2 2020 - Q3 2021	968,431
RECOVR	RDD	Phone (CATI)	Q2 - Q4 2020	1,505
World Bank	RDD	Phone (CATI)	Q2 2020	998
Ghana				
Afrobarometer	Household	Face-to-face	Q2 2022	2,369
Facebook	Social media	Internet	Q2 2020 - Q3 2021	60,584
RECOVR	RDD	Phone (CATI)	Q2 2020	1,637
Kenya				
Afrobarometer	Household	Face-to-face	Q4 2021	2,400
Facebook	Social media	Internet	Q2 2020 - Q3 2021	238,564
RECOVR	RDD	Phone (CATI)	Q3 - Q4 2020	794
World Bank	Call list	Phone (CATI)	Q2 - Q3 2020	4,060
Mexico City				

Sample	Sampling	Interview Mode	Timing	Sample Size
Facebook	Social media	Internet	Q2 2020 - Q3 2021	388,875
RECOVR	RDD	Phone (CATI)	Q2 2020	1,335
Nigeria				
Afrobarometer	Household	Face-to-face	Q1-2020	1,599
Facebook	Social media	Internet	Q2, 2020 - Q3 2021	173,332
RECOVR	RDD	Phone (CATI)	Q4 2020 - Q1 2021	1,968
World Bank	Call list	Phone (CATI)	Q2 2020	6,205
Philippines				
Facebook	Social media	Internet	Q2 2020 - Q3 2021	653,879
RECOVR	RDD	Phone (CATI)	Q2 - Q3 2020	1,389
Rwanda				
RECOVR	RDD	Phone (CATI)	Q2 - Q4 2020	1,489
Sierra Leone				
Afrobarometer	Household	Face-to-face	Q1 2022	1,200
RECOVR	RDD	Phone (CATI)	Q2 - Q4 2020	1,284

Benchmark datasets

The primary parameter of interest in this analysis is the difference between sample averages in any given survey and the true population mean for a given variable. Since the true population mean is unobserved, we instead use as benchmarks the most recent official in-person national household survey. These surveys were used to report official statistics including poverty and employment rates and were conducted using random household selection stratified by sub-national region with large sample sizes.

Table 3 describes which nationally representative survey datasets were used to calculate benchmark values in each country. This table highlights an important feature of this analysis, namely that the most recent surveys were collected between two and five years prior to recruitment of the pandemic-era surveys, with an average of three years across all countries. To avoid conflating mode effects with real-world changes in social and economic conditions, our analysis is restricted to questions that are either unlikely to have been affected by the pandemic or retrospective questions about pre-pandemic factors. Even so, the benchmark values do not provide us with "ground truth", but rather show us the best estimates available.

Harmonization of key demographic variables

To allow for comparisons across samples, we harmonized key variables across all 31 datasets. Where variables are either unavailable or were collected in a way that does not allow for harmonization, we omit the samples from that particular regression. Where necessary, we re-coded more granular or continuous variables to enable harmonization, e.g. by collapsing continuous age values to match samples where only age ranges were collected. Harmonized household-level variables include household size, whether the household includes any school-age children, and urbanicity. Individual-level variables include respondent age, relationship to the household head, educational attainment, and employment status.

For respondent age, most datasets included a continuous age variable with the exception of the social media questionnaire, which provided five age ranges, the three oldest of which were combined to ensure a large enough sample size within each group for each sample. As a result, we group respondents into age ranges of 18-34, 35-54, and 55 and up. Household size is determined using a household roster listing all family members for the face-to-face and LSMS call list surveys, while RDD and social media surveys asked for the number directly (i.e. "How many people, including yourself, live in your current household?").

We construct a binary urbanicity indicator based on UN Habitat's 2020 list of metropolitan areas, defined as urban agglomerations with a population exceeding 300,000 (Habitat 2020). Few samples explicitly coded households as being in urban or rural areas. Instead, we list a household as being in an urban region if the lowest available administrative region (harmonized within country across samples) contains one of the urban agglomerations listed by the UN. Samples are only removed from the analysis if the sub-national location data is either not available or not sufficiently granular.

Educational attainment is coded following UNESCO's 2011 International Standard Classification of Education (ISCED-11), which is designed to facilitate comparability across countries. Following the ISCED classifications, we harmonize educational attainment into four broad categories: no formal education, basic/ primary, intermediate/secondary, and advanced.[3]

Respondents' employment status was measured differently across each survey project, so we harmonize a broadly-defined measure across study samples. The employment rate in a population is frequently of immediate policy importance and often needs to be measured on short notice, making the suitability of these remote methods for measuring pre-pandemic employment particularly useful in assessing remote methods. We define employment as having been engaged in any remunerative activity, including formal employment, wage labor, agriculture (for sale or own production), or self-employment.[4] Response windows varied across questionnaires from one to four weeks, with employment status treated equally regardless of recall period. To improve cross-sample comparability, we focus on pre-pandemic employment status in the first quarter of 2020 rather than the employment status during any COVID-19 lockdowns. As noted above, benchmark estimates are coded in the same way for the most recent available period.

Empirical approach

Our main analysis compares the sample averages for a range of variables to those of a benchmark survey. The primary parameter of interest in this analysis is the difference between a survey's

[3]Individuals who have not completed primary school are classified in the first category. The basic category includes those who completed at least a primary education, while the intermediate category indicates completion of (upper) secondary education, and the advanced category includes those who have completed any graduate or post-graduate degree.

[4]Self-employment and farming are included regardless of whether they have received revenue recently.

estimate and the true (unobserved) value of the population mean for selected variables. As discussed above, we start with the most recent official nationally representative face-to-face household survey in each country as our benchmark. To these, we append the harmonized data across all countries, C, and all survey modes, S, and define $I(cs)$ to be 1 if an observation is in country $c \in C$ and sample $s \in S$ and zero otherwise. This allows us to estimate a single unified model for each variable of interest, Y:

$$Y_{ics} = \sum_{c \in C} \beta_c + \sum_{s \in S} \beta_{cs} I(cs) + \epsilon_{ics} \tag{1}$$

This yields the benchmark average for Y_c in each country, $\widehat{\beta}_c$, and the estimated sampling bias for each survey, $\widehat{\beta_{cs}}$. In this case, $\beta_{cs} = 0$ would indicate that the average for that sample in that country matches the benchmark. While our primary focus is on these country-sample differences, it is also helpful to aggregate these across countries to understand the average estimate for each survey mode. To do so, we estimate a set of survey mode fixed effects, $\hat{\beta}_s$, without differentiating by country.

$$Y_{ics} = \sum_{c \in C} \beta_c + \sum_{s \in S} \beta_s I(cs) + \epsilon_{ics} \tag{2}$$

Left unadjusted, this cross-country specification would arbitrarily overweight the samples with more observations. For example, the RDD sample in Ghana contains roughly twice as many respondents as in Kenya, but when estimating the average bias across all RDD samples, we would take the two samples as more or less equally informative. We thus adjust for sample size by normalizing the weights within each sample to sum to one. This leaves $\widehat{\beta}_s$ equal to the simple average of all the coefficients $\widehat{\beta_{cs}}$.

Comparing weighted and unweighted regressions

Each survey sample was recruited with the expectation that the probability of completing a survey would be unequal across households, both as a result of stratification and coverage and non-response bias. Survey weights aim to adjust for these differences, and so the weighted estimates of β_{cs} represent our primary estimate for how closely each sample matches our benchmark estimate. For variables that were explicitly used in the calculation of survey weights, we should expect that by construction, the weighted average should come very close to the official estimate. This expectation may not hold however if the functional form of the variable used is different. For example, when the age of respondents is used, sampling weights may yield different estimates even of closely related statistics like the population share over 65 years. Finally, variables that are not directly related to the vector of survey weights, like employment or school enrollment, will have unbiased weighted averages only insofar as two assumptions are met: the sampling weights themselves must be correlated with the probability that a household would be included in the survey sample; and the weights must also be correlated with the variable of interest. Insofar as these two assumptions hold, the weighted average values of Y should be closer to the benchmark and $\widehat{\beta_{cs}}$ should be closer to zero in the weighted version of equation 1.

To distinguish the role of unbiased sampling from that of *ex post* statistical adjustment, our main figures present the coefficients $\widehat{\beta_{cs}}$ from both weighted and unweighted versions of equation 1. The unweighted estimates offer insight into what sorts of households were less likely to be recruited in the first place, while the weighted estimates tell the extent to which this bias can be corrected. Importantly, the sampling designs differ markedly across each of the four multi-national survey projects, and the methods for calculating sampling weights were chosen by researchers accordingly. Rather than harmonizing and thereby potentially mis-specifying these methods across survey modes, we use the household and individual weights calculated by the original research teams. As such, we cannot rule out the possibility that some other, more sophisticated adjustment approach might have yielded different results. Details on the specific statistical method and demographic variables used when constructing weights for each sample can be found in Table A1 in the supplementary appendix.

Table 3: Nationally Representative (Benchmark) Surveys- Year and Sample Size

Country	Sample Size	Survey	Producer	Year
Burkina Faso	45,612	Enquete Multisectorielle Continue / Living Standards Measurement Survey (LSMS)	Institut National de la Statistique et de la Demographie (INSD)	2018-9
Colombia	816,994	Gran Encuesta Integrada de Hogares (GEIH)	Departamento Administrativo Nacional de Estadistica (DANE)	2019
Ghana	31,374	Ghana Living Standards Survey (GLSS 7) / Living Standards Measurement Survey (LSMS)	Ghana Statistical Service	2016-7
Kenya	45,877	Integrated Household Budget Survey (KIHBS)	Kenya National Bureau of Statistics (KNBS)	2015-6
Mexico City*	5,618	Encuesta Nacional de Ingresos y Gastos (ENIGH)	Instituto Nacional de Estadistica, Geografía e Informatica (INEGI)	2018
Nigeria	57,838	General Household Survey-Panel / Living Standards Measurement Survey (LSMS)	National Bureau of Statistics, Nigeria (NBS)	2018-9
Philippines	41,544	Philippines Family Income and Expenditure Survey (FIES)	Philippine Statistics Authority	2015
Rwanda	33,419	Integrated Household Living conditions Survey (EICV5)	National Institute of Statistics of Rwanda (NISR)	2016
Sierra Leone	21,270	Integrated Household Survey (IHS)	Statistics Sierra Leone	2018

4. Results

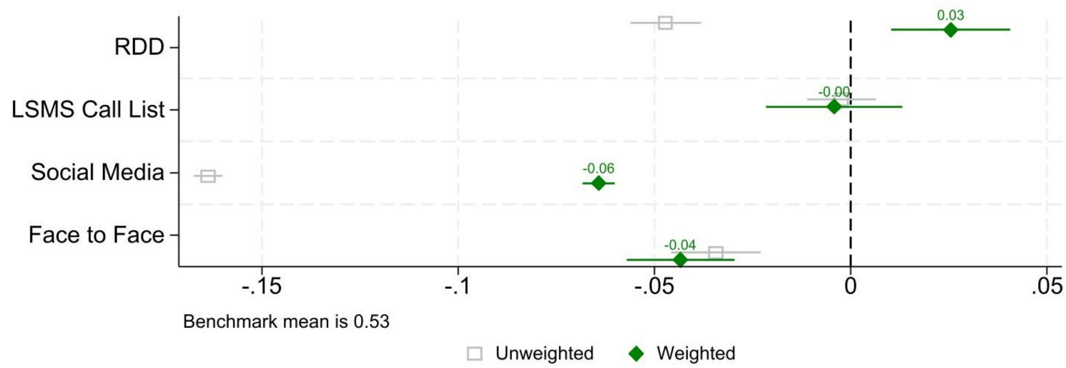
Appendix figures A1 to A11 and Tables A2 to A26 present the coefficients from equation 1 for each variable of interest. In each case, a pooled regression labeled "Pooled" displays the average of these differences within survey mode across each available country. These pooled results are shown in line in Figures 2 to 12 and are summarized in full in Table 4. We use the pooled results to discuss broad findings about each survey method, but focus heavily on country-level results to emphasize the role of region and context when evaluating different approaches. In each figure, the unweighted coefficient for each sample is presented in gray alongside the weighted coefficient in color to show how sampling weights affect the representativeness of each sample. The coefficients themselves represent the difference between a sample and the country's benchmark survey, while the actual benchmark value for each variable in each country is provided in the legend below each figure.

Respondent demographics

Beginning with the gender balance of each sample, shown in Figure A1, RDD and social media studies significantly over-sampled men in every African sample. Weighting is able to adjust for this bias in the RDD surveys, but not the social media samples. Meanwhile, surveys using nationally representative sampling frames (Afrobarometer and the LSMS Call List samples) consistently recruited balanced samples. Note that the Afrobarometer targets gender parity, essentially assuming that half of the population was female when in fact women consistently represent slightly more than half of most national populations. As such, small amounts of bias were essentially included by construction.

Turning to youth representation, RDD and social media samples disproportionately recruited respondents between 18 and 34 years old by 15pp and 6pp, respectively. The LSMS call list and face-to-face Afrobarometer samples came closer to proportional recruitment and report youth prevalence to within 2pp of the national benchmarks. Meanwhile, adults over 55 were dramat-

Figure 2: Representation of women in comparison to benchmark estimates

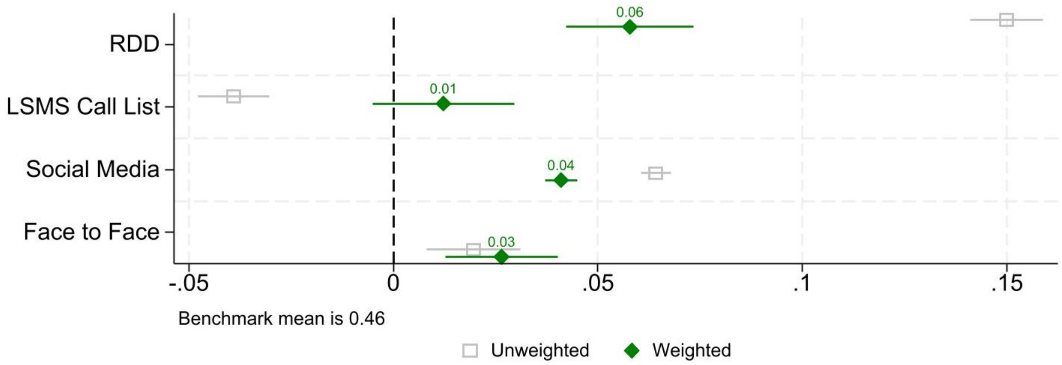


Estimates are centered around the average benchmark sample mean. Weights are normalized to sum to one within each sample. Grey boxes represent unweighted coefficients.

ically under-sampled in the RDD and social media samples, often by more than 50 percent relative to the actual population average, with smaller but statistically significant biases for Afrobarometer surveys in West Africa. Figures A2 and A3 highlight that weighted estimates are closer to the benchmarks in most (though not all) cases, but remain biased with much wider confidence intervals.

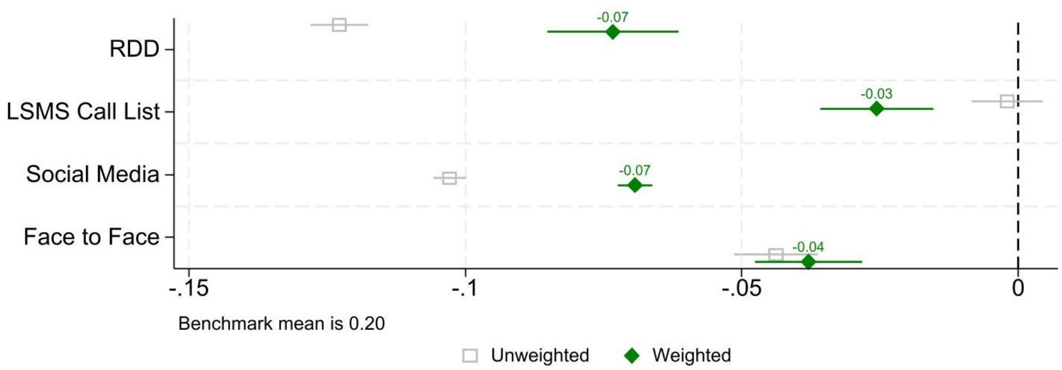
As with most LMICs, the countries considered in this analysis tend to have relatively young populations when compared to wealthier nations. While older adults represent a much smaller share of the national population in these countries than in wealthier countries, they are often of central importance for policy and public health researchers. The COVID-19 pandemic in particular highlighted the importance of being able to contact and study a representative group of older individuals. This under-representation of older adults was also consistent in face-to-face samples. As with gender, age was explicitly included as a target variable when constructing sampling weights for the RDD and social media surveys, yet re-weighting did not generally address the imbalance in presentation of this small, but critical population subgroup. This highlights that even when a group like the elderly is heavily under-represented in the unweighted sample, the problem isn't necessarily solved even by including a closely related variable (average respondent age) when calculating weights.

Figure 3: Representation of adults aged 18-34



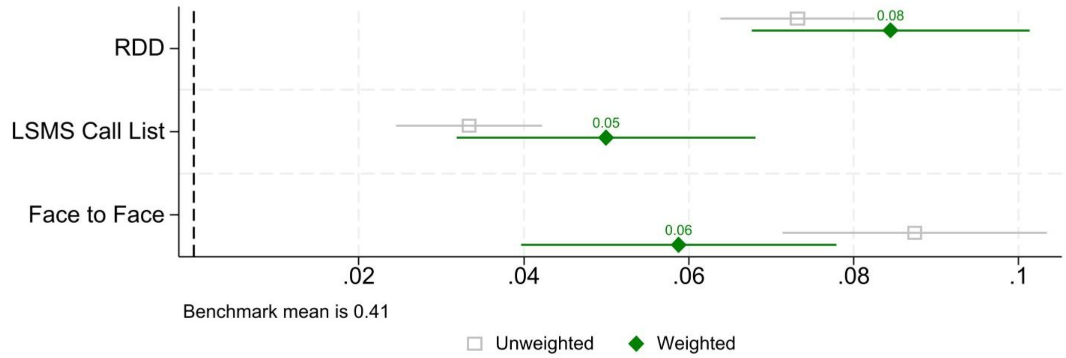
Estimates are centered around the average benchmark sample mean. Weights are normalized to sum to one within each sample. Grey boxes represent unweighted coefficients.

Figure 4: Representation of adults age 55 or higher



Estimates are centered around the average benchmark sample mean. Weights are normalized to sum to one within each sample. Grey boxes represent unweighted coefficients.

Figure 5: Representation of heads of household



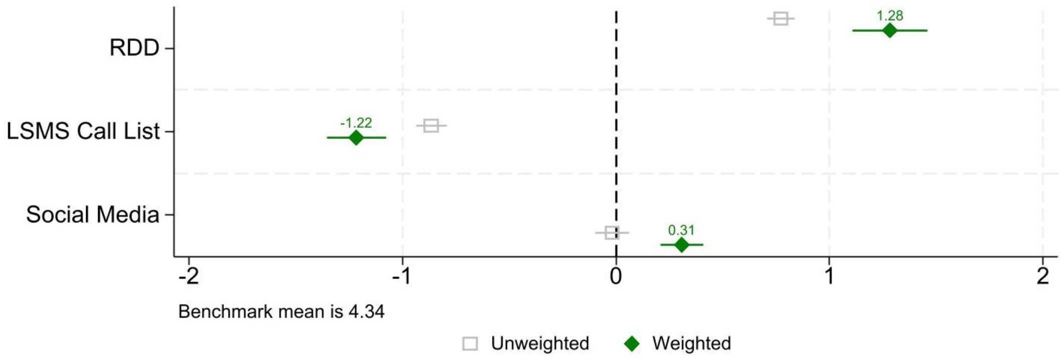
Estimates are centered around the average benchmark sample mean. Weights are normalized to sum to one within each sample. Grey boxes represent unweighted coefficients.

RDD, call list, and face-to-face surveys all tended to over-recruit household heads on average, albeit with significant differences across and within countries. Looking at Figure A4, we see out of the fifteen samples considered, sampling weights either had no effect or moved the estimates further from their benchmarks in nine. Household heads would be about as prevalent in these samples as in the general population if phone ownership, access, and propensity to respond were evenly distributed across the adult population. Instead, out of nine RDD samples, seven over-represent household heads. All three face-to-face samples over-represent the head of the household. Having the status of household head is negatively correlated with key variables used to construct weights, including gender. This may explain why statistical adjustment significantly increased the variance of the estimated sample means and were as likely to shift point estimates away from the benchmark values as towards them.

The three remote survey methods see different results for household size, with RDD overestimating household size by an average of 0.77 persons on average, LSMS call lists underestimating by 1.22 persons, and social media surveys overestimating by 0.31 persons. [5] In five of the seven countries with social media samples, the Facebook survey came closer to accurately representing household size than IPA's RDD or the LSMS call list surveys. As with the household head vari-

[5] As discussed in Section 3 The face-to-face (Afrobarometer) samples did not include total household size, and are therefore excluded from this analysis.

Figure 6: Mean household size



Estimates are centered around the average benchmark sample mean. Weights are normalized to sum to one within each sample. Grey boxes represent unweighted coefficients.

able, Figure A5 shows that weighting did little to improve on the unweighted averages in any of the three survey modes.

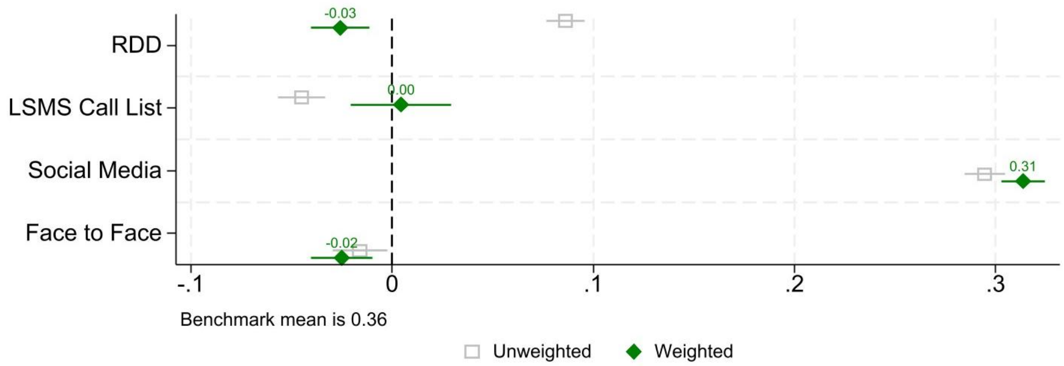
Separating respondents into urban and rural regions, we see that samples RDD and social media samples disproportionately recruit from urban areas on average, often by more than 100 percent of the benchmark estimates. Meanwhile, face-to-face surveys and call list surveys respectively come within 2 and 5pp of the benchmark estimates on average. Statistical adjustment is effective in the RDD samples, but not in the two social media samples for which urbanicity could be established. These over-represent urban respondents by a factor of 72 percent in Kenya and 130 percent in Burkina Faso.

Education

Turning to educational attainment, respondents were categorized as having completed less than basic, basic, intermediate, or advanced levels of formal education[6]. Beginning with the groups listed as having either less than or only a basic formal education, we again see a dramatic difference between RDD and social media surveys on one hand and face-to-face or LSMS call list survey on the other. While the five Afrobarometer surveys and one Kenyan call-list survey do

[6]See Section 3.

Figure 7: Residence in urban area

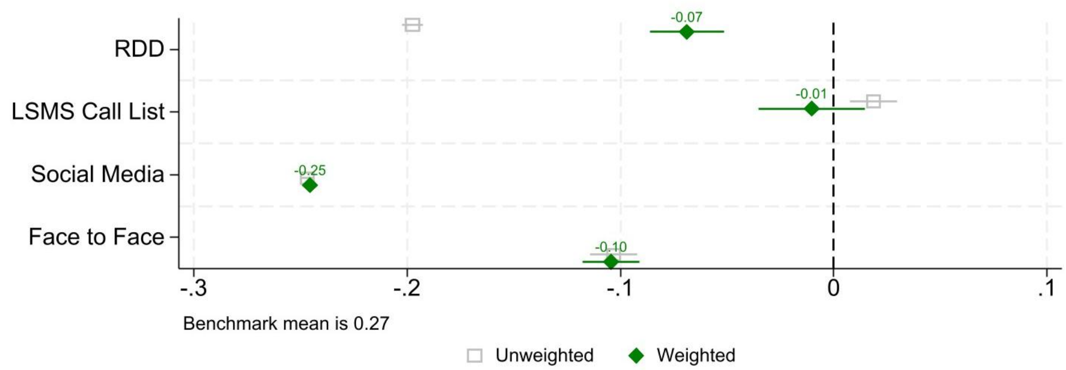


Estimates are centered around the average benchmark sample mean. Weights are normalized to sum to one within each sample. Grey boxes represent unweighted coefficients.

under-represent these respondents (especially in Nigeria), they were all able to recruit a non-trivial number of respondents in this group. This stands in stark contrast to the Facebook surveys, which almost entirely failed to recruit respondents at the basic or lower levels. The regression and summary statistic figures from Table A16 to A19 show that while this group represents the majority of every African nation (and the Philippines), they did not represent more than 5 percent of the sample in any Facebook survey. While less pronounced, the RDD samples also dramatically under-recruited from these low-education groups, especially those with no formal education. However, weighting adjusts for these errors reasonably well in the RDD samples, under-representing the low-education group by 7pp and over-representing the basic education group by 7pp. By contrast, sampling weights have strikingly little influence on the social media or face-to-face estimates.

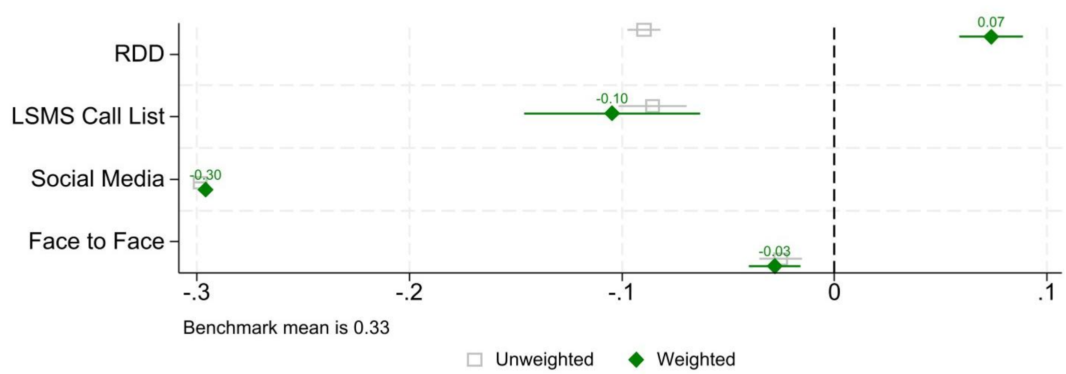
As expected, given the under-representation of respondents who have not completed a secondary education, we observe on average a corresponding *over*-representation of respondents who have completed a secondary education across all survey modes. Furthermore, respondents who have completed a post-secondary diploma or more are over-represented in nearly every case for RDD surveys and in social media surveys by an average of nearly 50 percentage points. In stark contrast, face-to-face and LSMS call list samples come close to the benchmark for higher education.

Figure 8: Respondent did not complete any level of formal education



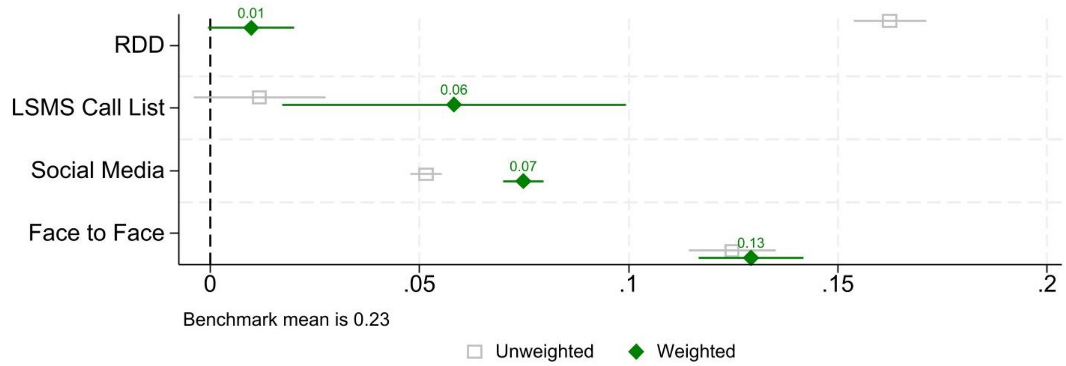
Estimates are centered around the average benchmark sample mean. Weights are normalized to sum to one within each sample. Grey boxes represent unweighted coefficients.

Figure 9: Respondent completed basic level formal education



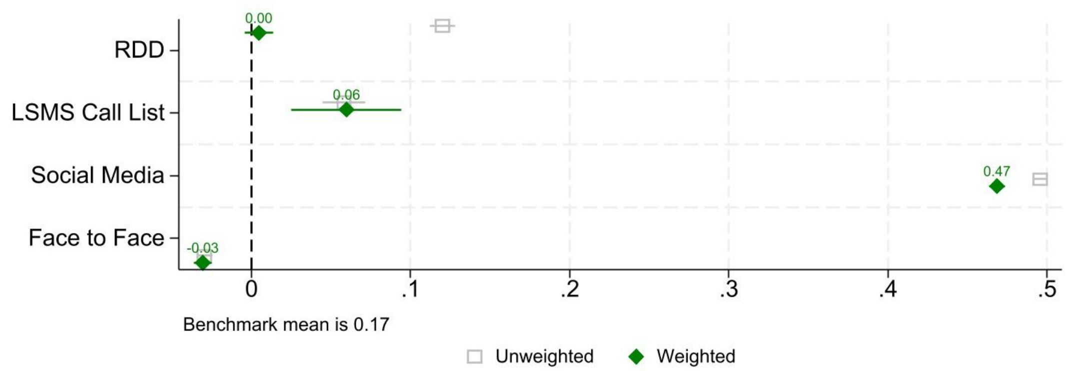
Estimates are centered around the average benchmark sample mean. Weights are normalized to sum to one within each sample. Grey boxes represent unweighted coefficients.

Figure 10: Respondent completed intermediate level formal education



Estimates are centered around the average benchmark sample mean. Weights are normalized to sum to one within each sample. Grey boxes represent unweighted coefficients.

Figure 11: Respondent completed advanced level formal education



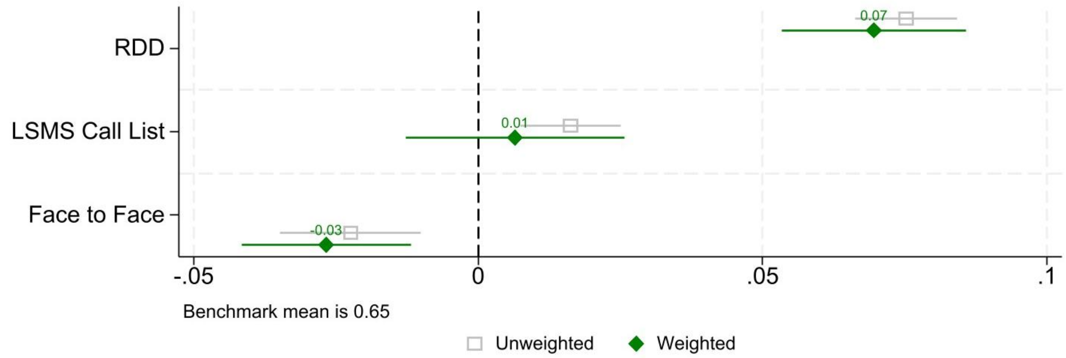
Estimates are centered around the average benchmark sample mean. Weights are normalized to sum to one within each sample. Grey boxes represent unweighted coefficients.

As with lower levels of education, sample weights appear to bring RDD samples into line with the benchmark and are ineffective everywhere else, as shown in Figure A10.

Employment

Turning finally to pre-pandemic employment, we see that respondents who are employed or self-employed prior to March 2020 are over-represented on average in RDD, but not in the LSMS call list samples and by fairly small amounts in all but one of the West African face-to-face surveys,

Figure 12: Respondents who were employed prior to March 2020



Estimates are centered around the average benchmark sample mean. Weights are normalized to sum to one within each sample. Grey boxes represent unweighted coefficients.

namely Sierra Leone.[7] This is broadly consistent with the results for education and urbanicity, two factors that are highly correlated with pre-pandemic employment in these surveys. This is arguably the best test among all variables considered of a broader question posed by this analysis, namely whether these survey methods are suitable for recovering not only averages for demographic features, but also for policy-relevant socio-economic factors. Weighting does not effectively recover anticipated population means in any of the modalities or samples, and primarily result in increasing the variance of the estimates.

[7]The social media surveys did not include sufficient employment data.

Table 4: Covariate Balance by Outcome (Pooled across countries)

Outcome	Benchmark	Weight	RDD	Call List	Social Media	Face-to Face
Female respondents (%)	0.53	Unweighted	-0.05*	-0.00	-0.16*	-0.03*
		Weighted	0.03*	-0.00	-0.06*	-0.04*
Aged 18 to 34 (%)	0.46	Unweighted	0.15*	-0.04*	0.06*	0.02*
		Weighted	0.06*	0.01	0.04*	0.03*
Aged 35 to 54 (%)	0.34	Unweighted	-0.03*	0.04*	0.04*	0.02*
		Weighted	0.02†	0.01	0.03*	0.01
Aged over 55 (%)	0.20	Unweighted	-0.12*	-0.00	-0.10*	-0.04*
		Weighted	-0.07*	-0.03*	-0.07*	-0.04*
Household Heads (%)	0.41	Unweighted	0.07*	0.03*		0.09*
		Weighted	0.08*	0.05*		0.06*
Household size	4.34	Unweighted	0.77*	-0.87*	-0.02	
		Weighted	1.28*	-1.22*	0.31*	
In a major city (%)	0.36	Unweighted	0.09*	-0.05*	0.29*	-0.02†
		Weighted	-0.03*	0.00	0.31*	-0.02*
Education: > Basic (%)	0.27	Unweighted	-0.20*	0.02*	-0.25*	-0.10*
		Weighted	-0.07*	-0.01	-0.25*	-0.10*
Education: Basic (%)	0.33	Unweighted	-0.09*	-0.09*	-0.30*	-0.03*
		Weighted	0.07*	-0.10*	-0.30*	-0.03*
Education: Intermediate (%)	0.23	Unweighted	0.16*	0.01	0.05*	0.12*
		Weighted	0.01	0.06*	0.07*	0.13*
Education: Advanced (%)	0.17	Unweighted	0.12*	0.06*	0.50*	-0.03*
		Weighted	0.00	0.06*	0.47*	-0.03*
Employed respondents (%)	0.65	Unweighted	0.08*	0.02*		-0.02*
		Weighted	0.07*	0.01		-0.03*

† p<0.05, * p<0.01

5. Discussion

This paper harmonizes 31 separate survey samples collected in nine countries using five separate methodologies, each of which had a stated goal of national representation. Sampling bias in a given project is driven by a wide variety of factors, including coverage and non-response bias. In identifying consistent patterns across contexts, we aim to draw generalizations about which of these is likely to "work" for a given research agenda. To inform research design decisions in a new context, these results must be weighed alongside the cost and difficulty of each approach, as well as careful consideration as to what degree of survey error is likely to be "good enough".

We consistently see an important distinction between Random Digit Dialing (RDD) and social media samples on one hand and face-to-face samples and call lists using face-to-face samples on the other. The main common feature between the call list and in-person samples is that they both start with a nationally representative sampling frame, whereas RDD and Facebook surveys must draw from the subset of active phone and social media users. This limitation seems to lead to significantly biased estimates even for straightforward demographic features. Social media surveys see especially dramatic bias towards educated and urban respondents, calling into question the sorts of questions that approach is able to answer. The RDD and call list surveys both employ CATI with a live interviewer and face the same limitations in mobile phone coverage. Insofar as this holds mode effects and coverage error constant, we might conclude that the differences between them are driven primarily by differences in non-response error.

Looking beyond bias in the initially recruited samples, the analysis then asks to what extent this bias can be corrected for with post-stratification weights. In most cases, we find that the RDD samples are more successful in re-weighting to bring sample means closer to the benchmark estimates compared to the other survey modes. Educational attainment serves as the clearest example. Both RDD and social media samples under-recruited low-education respondents, but the RDD samples were able to largely adjust for this while weighting had virtually no effect on the social media estimates. Nonetheless, every country and methodology saw many cases in which weighting appeared unhelpful, significantly increasing variance without bringing sample means closer to the benchmarks.

The mixed results on sampling weights highlight some basic and well known features of sample adjustment methods that are vital to bear in mind when designing a questionnaire and sampling strategy. First, heavily under-represented groups cannot be adequately represented in a weighted analysis if they make up only a very small portion of the initial sample. Heavily biased recruitment will inevitably lead to biased estimates of population statistics, especially for modestly sized samples. Second, sampling weights will only improve estimates of a population mean if the variables used to construct the weights are highly correlated with the outcome of interest *and* the selection probabilities in the population. While implicit in many descriptive results across the social science literature, researchers cannot simply assume this is the case without justification or evidence. Instead, this analysis should serve as a reminder that we cannot simply assume that generic household or individual weights will address concerns about sample bias, and that researchers should design weights specifically with respect to their outcomes of interest.

Nonetheless, data from remote surveys can still be informative. When such surveys use sampling frames that are known to be representative of the community in question, we often find little evidence of selection bias. Among surveys without nationally representative sampling frames, statistical adjustment can often bring estimates in line with official estimates when the target variables and weighting methods used are appropriate for the context. Given the limitations of remote methods in building representative samples, one important step that we recommend is for researchers to clearly present unweighted summary statistics for their samples, with comparison to high quality population estimates, before presenting any other study findings. This allows readers to understand both the sampling frame and who was interviewed, and to contextualize results accordingly, which may often be far more informative than any statistical adjustment done "behind the scenes". Finally, researchers seeking to identify appropriate research designs should assess the geographic reach and general accessibility of a given survey mode within their population of interest.

Bibliography

Bahia, K, and A Delaporte. 2020. Connected Society: The State of Mobile Internet Connectivity 2020 GSMA.

Ballivian, Amparo, João Pedro Azevedo, Will Durbin, J Rios, J Godoy, and C Borisova. 2015. "Using mobile phones for high-frequency data collection." *Mobile Research Methods* 21.

Blackmon, William, Rafe Mazer, and Shana Warren. 2021. "Kenya Consumer Protection in Digital Finance Survey." *Innovation for Poverty Action*, 2021.

Corsi, Daniel J, Melissa Neuman, Jocelyn E Finlay, and SV Subramanian. 2012. "Demographic and health surveys: a profile." *International journal of epidemiology* 41 (6): 1602–1613.

Dillon, Brian. 2012. "Using mobile phones to collect panel data in developing countries." *Journal of international development* 24 (4): 518–527.

Gibson, Dustin G, Adaeze C Wosu, George William Pariyo, Saifuddin Ahmed, Joseph Ali, Alain B Labrique, Iqbal Ansary Khan, Elizeus Rutebemberwa, Meerjady Sabrina Flora, and Adnan A Hyder. 2019. "Effect of airtime incentives on response and cooperation rates in non-communicable disease interactive voice response surveys: randomised controlled trials in Bangladesh and Uganda." *BMJ global health* 4 (5): e001604.

Groves, Robert M. 2011. "Three eras of survey research." *Public opinion quarterly* 75 (5): 861–871.

GSM Association. 2020. "Renewable Energy for Mobile Towers: Opportunities for low- and middle-income countries, 2020." *URL: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2020/09/Clean*$_{T}ech_{R}eport_{RW}ebSingles.2.pdf[accessed2023-05-19]$, 2020.

———. 2021a. "The state of mobile internet connectivity 2020." *URL: https://www.gsma.com/r/wp-content/uploads/2020/09/GSMA-State-of-Mobile-Internet-Connectivity-Report-2020.pdf [accessed 2021-05-18]*, 2021.

GSM Association. 2021b. "The state of mobile internet connectivity 2021." *URL: https://www.gsma.com/r/wp-content/uploads/2021/09/The-State-of-Mobile-Internet-Connectivity-Report-2021.pdf [accessed 2021-05-18]*, 2021.

Habitat, UN. 2020. "World Cities Report 2020: The value of sustainable urbanization." *Nairobi, Kenya*, 2020.

Henderson, Savanna, and Michael Rosenbaum. 2020. "Remote surveying in a pandemic: research synthesis." *Innovation for Poverty Action*, 2020.

Himelein, Kristen, and Jonathan G Kastelic. 2015. "The socio-economic impacts of Ebola in Liberia," 2015.

Himelein, Kristen, Mauro Testaverde, Abubakarr Turay, and Samuel Turay. 2015. "The Socio-Economic Impacts of Ebola in Sierra Leone: Results from a High Frequency Cell Phone Survey, Round 3," 2015.

ITU. 2020a. Manual for measuring ICT access and use by households and individuals.

———. 2020b. World Telecommunication/ICT Indicators Database.

———. 2021. World Telecommunication/ICT Indicators Database.

Kastelic, Kristen Himelein, Stephanie Eckman, Jonathan G Kastelic, Kevin Robert Mcgee, et al. 2020. "High frequency mobile phone surveys of households to assess the impacts of COVID-19 (Vol. 2): guidelines on sampling design," 2020.

Kempf, Angela M, and Patrick L Remington. 2007. "New challenges for telephone survey research in the twenty-first century." *Annu. Rev. Public Health* 28:113–126.

Lau, Charles Q, Alexandra Cronberg, Leenisha Marks, and Ashley Amaya. 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.

Leo, Benjamin, Robert Morello, Jonathan Mellon, Tiago Peixoto, and Stephen T Davenport. 2015. "Do mobile phone surveys work in poor countries?" *Center for Global Development Working Paper*, no. 398.

Pariyo, George W, Abigail R Greenleaf, Dustin G Gibson, Joseph Ali, Hannah Selig, Alain B Labrique, Gulam Muhammed Al Kibria, Iqbal Ansary Khan, Honorati Masanja, Meerjady Sabrina Flora, et al. 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 (4): e0214450.

Rowntree, Oliver, and Matthew Shanahan. 2020. "Connected women: the mobile gender gap report 2020." *London, England: GSM Association*, 2020.

Silver, Laura, Aaron Smith, Courtney Johnson, Kyle Taylor, Jingjing Jiang, Monica Anderson, and Lee Rainie. 2019. "Mobile connectivity in emerging economies." *Pew Research Center* 7.

Velthausz, D, R Donco, H Skelly, and M Eichleay. 2016. "Mozambique mobile access and usage study: computer-assisted telephone interview (CATI) survey results." *United States Agency for International Development, Retrieved on February* 25:2021.

Appendix

Data Description

Random Digit Dial samples: RECOVR & related IPA surveys

Starting in April 2020, IPA initiated the RECOVR initiative, which aimed to provide government and multinational organizations with timely descriptive statistics about the social and economic impacts of the pandemic, while also providing a common set of indicators across several countries in Africa, Asia, and Latin America. The RECOVR studies in this analysis consist of phone surveys with RDD-generated samples with questionnaire design, mode, and protocol held relatively constant. First-round sample sizes ranged from 800 to 1,500 per country.

During mid-late 2020, IPA also conducted phone surveys with RDD sampling on the gendered impacts of COVID-19 in Nigeria and the pandemic's effect on consumer financial protection in Kenya. Kenya's RDD sample was the only sample in this study not explicitly aiming for national representation. Instead, it was designed to be representative of adult users of digital financial services, and so filtered respondents before conducting interviews. Since the estimated 23 million users of digital financial services in Kenya represent over 85% of adults in Kenya and the majority of adults in every major demographic group (Blackmon, Mazer, and Warren 2021), we continue to compare this group to the overall adult population in other samples. In each RDD sample, there was at least one round of surveys sometime between April and December 2020, after the outbreak of COVID-19 and the initial implementation of government lockdowns and/or other social distancing policies. As a consequence, most studies were conducted during periods of lockdowns and mobility restrictions. Respondents in IPA's surveys were contacted from a random sample of valid phone numbers, pre-pulsed or 'screened' by a survey sampling firm to identify active numbers and stratified proportionate to network operator market share.[8] This procedure avoids lost time dialing inactive numbers while producing a statistically representative set

[8] Prefixes are typically assigned to specific mobile network operators.

of active mobile phone numbers in a country.[9]

Like all RDD surveys, the IPA Nigeria and Kenya studies were conducted with an individual adult associated with a phone number who provided data on him- or herself as well as other household members during 20-40 minute surveys. A limited set of demographic questions included gender, age, educational attainment, and some information on household size and composition.

For all surveys, post-stratification weights were constructed using inverse probability weighting and raking adjustment using the national benchmark samples as the source for population weights. For each country, households were separated into bins based on a set of household characteristics. The bins separated households by variables chosen based on the major sources of imbalance of concern, which include educational level, age, gender and region. Weights were then calculated as the relative frequency of RDD households in that bin over that of respondent households in the national samples. These were then adjusted using raking adjustment to ensure that within each country, sample averages of the weighting variables match the estimated population averages. Weighting variables in this case included gender, urbanicity, three age categories (18 to 30, 31 to 60 and over 61 years old) and a binary indicator for secondary education completion.

Call list samples: World Bank High Frequency Phone Surveys

We refer to phone survey samples drawn from initial face-to-face surveys as "call list" surveys. In these cases, recruitment of the initial sample frame was done in person and researchers have some information about each individual. Relative to RDD surveys, this allows researchers to attempt to minimize and adjust for differential nonresponse. In the analysis that follows, we bring in a specific type of call list sample drawn from high quality face-to-face samples to represent the best case of call list sampling for phone surveys. During the early months of the pandemic, the World Bank decided to implement phone panel surveys drawn from recent nationally repre-

[9]The RDD sampling frame was drawn based on the mobile network operator (MNO) market share where numbers were allocated by provider. This approach aimed to make the sample representative of mobile phone numbers in the country. See Sample Solutions' RDD white paper at https://sample.solutions/about/library/.

sentative face-to-face samples that included phone numbers. For the HFPS, the World Bank's LSMS partners randomly sampled households to be contacted, accounting for non-response rates, to meet target sample sizes in each of its nine countries. Colombia is the only HFPS that did not implement a call list sampling design but instead used an RDD design.

HFPS weights were constructed starting with the base weights from each country's respective nationally representative survey which the study sample was drawn from. From there, each country study produced a variation of this weighting, accounting for selection probability, attrition based on demographics, and post-stratification, to create both panel and cross-sectional weights. For more information on the approach used, see Kastelic et al. (2020).

Social media samples: Facebook COVID-19 Trends and Impact Survey (CTIS)

Facebook Research and Carnegie Mellon University implemented their COVID-19 Trends and Impact Survey (CTIS) beginning in April 2020, and ending in mid-2022. The sampling frame consisted of users over the age of 18 in the Facebook Active User Base (FAUB) who used Facebook in one of the supported locales and survey languages. The CTIS surveys were daily repeated cross-sections. The Facebook app recruited a new sample of adult users to take the survey each day via an invitation at the top of their Facebook News Feed. Sampled users were selected via stratified random sampling by sub-national administrative boundaries, and non-respondents were invited to take the survey again in either a few weeks or months, depending on the population density of their area. In low density regions, eligible users were recruited monthly, while in high density regions, they were recruited every two to six months. There are no unique user identifiers in the data; thus, data from sampled users who completed the survey multiple times cannot be linked longitudinally. Once respondents were recruited, Facebook constructed weights aiming at representation of the entire adult population in a country or territory, including people not covered by the FAUB. For countries and territories outside the US, Facebook applied raking over age and gender using benchmarks obtained from the United Nations (UN) Population Division 2019 World Population Projections, and first administrative level regions using benchmarks constructed from publicly available population density maps. Finally, weights were trimmed for outliers that were less than 0.03 or more than 10 times the national average weight, with the sum of weights re-scaled to the population size from the UN Population Projections.

Face-to-face samples: Afrobarometer public opinion surveys

Here, we include Afrobarometer surveys from Burkina Faso, Ghana, Nigeria, Sierra Leone, and Kenya as a point of comparison with our three sets of remote surveys. Three of these were pre-pandemic surveys, with the data from Sierra Leone, Nigeria, and Burkina Faso collected in March 2020, January 2020, and December 2019 respectively. The Kenya and Ghana surveys were collected well after the end of lockdowns and other restrictive policies in November 2021 and April 2022, respectively.

Afrobarometer has been collecting public opinion data on politics and economics across the continent since 1999, with surveys held every 2-3 years. Afrobarometer sampling frames are designed to be a representative cross-section of all voting-age citizens in a given country, sampling proportionate the most recent national census data. While the sample sizes are smaller and questionnaires shorter than the LSMS studies we use as primary benchmarks, the sampling and recruitment methods were broadly similar and allow for much more deliberate sampling than is available in an RDD or social media study. Combined with the high response rates of in-person household surveys, we might expect estimates from these studies to be as close as is feasible to nationally representative for a survey of this size. Given the timing of these surveys, we will have more insight into whether or not underlying parameters changed since LSMS surveys were conducted, and the role of remote methods in representativeness issues.

Afrobarometer surveys first employed geographic stratification to select enumeration areas. Within these areas starting points and households were randomly selected. Enumerators then randomly selected among all adults, alternating between men and women to ensure gender balance. Afrobarometer typically clusters eight interviews per enumeration area, and its standard sample national sizes of 1200 or 2400 depending (primarily) on country population therefore contain 150 or 300 enumeration areas, allocated across strata proportionate to population.[10] Afrobarometer is a public opinion survey at the individual level, and does not include complete household rosters.

Figures & Tables

NOTE: In each table & figure below, Colombia I RDD refers to IPA's study. Colombia II RDD refers to the World Bank High Frequency Phone Survey.

Table A1: Weighting method by Survey mode

Survey mode	Country	Weighting method	Variables in weighting vector
RDD	Burkina Faso Colombia Ghana Kenya Nigeria Rwanda Sierra Leone	Inverse Probability Weighting and Raking	Age, gender, region, urban indicator, highest educational level achieved
	Colombia II	Raking	Age, gender, region, telephone coverage
	Mexico City	Inverse Probability Weighting and Raking	Age, gender, urban indicator, highest educational level achieved
	Philippines	Inverse Probability Weighting and Raking	Region, urban indicator, highest educational level achieved

Continued on next page

[10]See https://www.afrobarometer.org/surveys-and-methods/sampling/ for details on Afrobarometer methodology.

Table A1: Weighting method by Survey mode (continued)

Survey mode	Country	Weighting method	Variables in weighting vector
Call List	Burkina Faso	Propensity Score Weighting	Age, gender, urbanicity, district, education, marital status, school attendance, labor force participation, household consumption and assets, household size, and dwelling characteristics
	Kenya	Inverse Probability Weighting on mobile phone ownership and Propensity Score Weighting	Urbanicity, county, gender and age group of household head, and mobile phone ownership
	Mexico City	Raking	Age, gender, region, telephone coverage
	Nigeria	Inverse Probability Weighting and Propensity Score Weighting	State, sector (urban/rural), household size, per capita consumption expenditure, household head gender and education, and household ownership of a mobile phone
Face to Face	Burkina Faso Ghana Kenya Nigeria Sierra Leone	Inverse Probability Weighting	Household size, region or province, sector (urban/rural), gender, and enumeration area

Continued on next page

Table A1: Weighting method by Survey mode (continued)

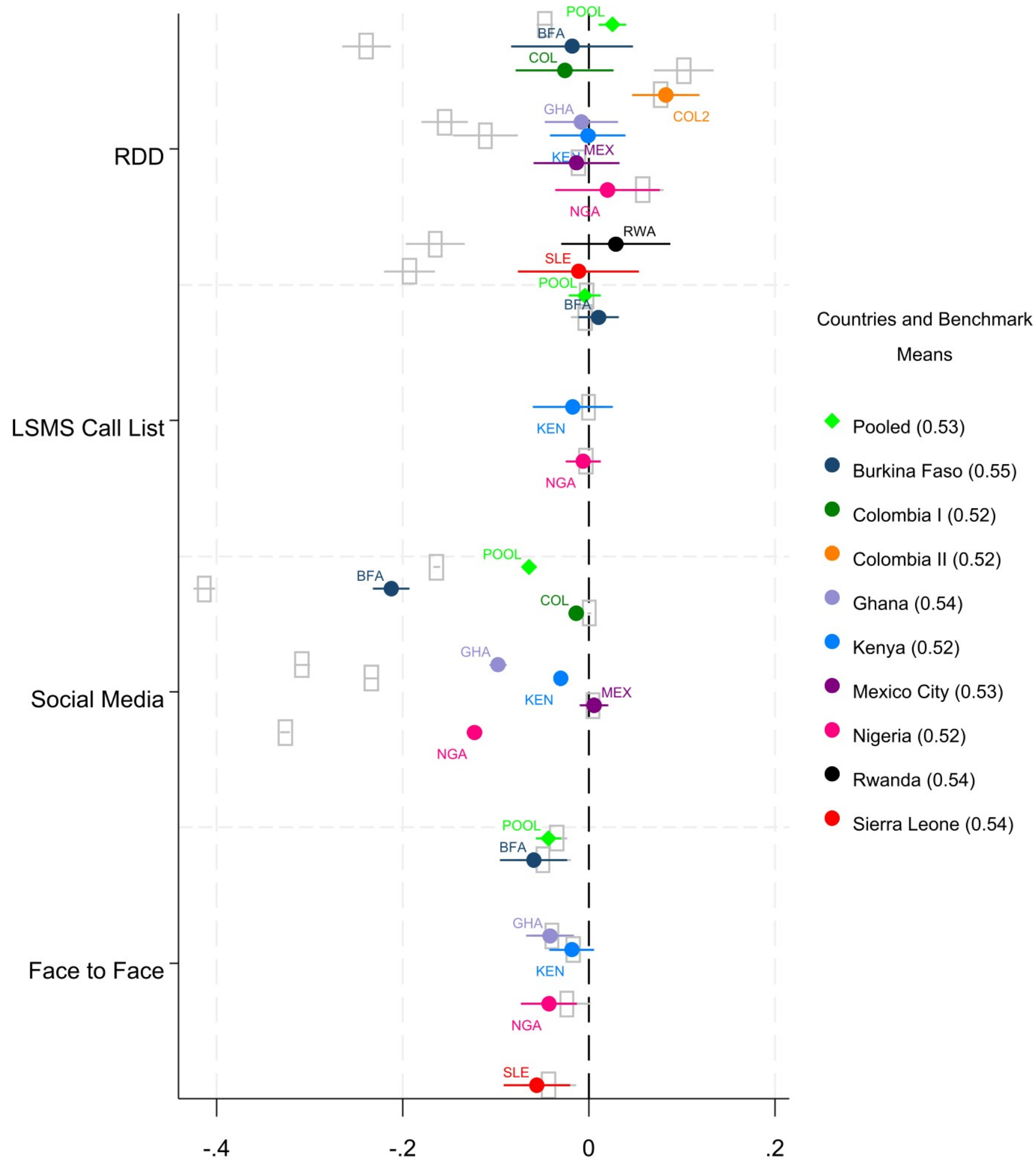
Survey mode	Country	Weighting method	Variables in weighting vector
Facebook	Burkina Faso Colombia Ghana Kenya Mexico City Nigeria	Inverse Probability Score Weighting and Raking	Age, gender, geographical variables, and other demographic attributes that correlate to survey outcomes

Table A2: Covariate Balance in the % of Female Respondents

	Benchmark	RDD	RDD Weighted	Call List	Call List Weighted	Social Media	Social Media Weighted	Face to Face	Face to Face Weighted
All Countries		–0.05*	0.03*	–0.00	–0.00	–0.16*	–0.06*	–0.03*	–0.04*
Burkina Faso	0.55	–0.24*	–0.02	–0.00	0.01	–0.41*	–0.21*	–0.05*	–0.06*
Colombia I	0.52	0.10*	–0.03			0.00	–0.01*		
Colombia II	0.52	0.08*	0.08*						
Ghana	0.54	–0.15*	–0.01			–0.31*	–0.10*	–0.04*	–0.04*
Kenya	0.52	–0.11*	–0.00	–0.00	–0.02	–0.23*	–0.03*	–0.02	–0.02
Mexico City	0.53	–0.01	–0.01			0.00	0.01		
Nigeria	0.52	0.06*	0.02	–0.00	–0.01	–0.33*	–0.12*	–0.02	–0.04*
Rwanda	0.54	–0.16*	0.03						
Sierra Leone	0.54	–0.19*	–0.01					–0.04*	–0.06*

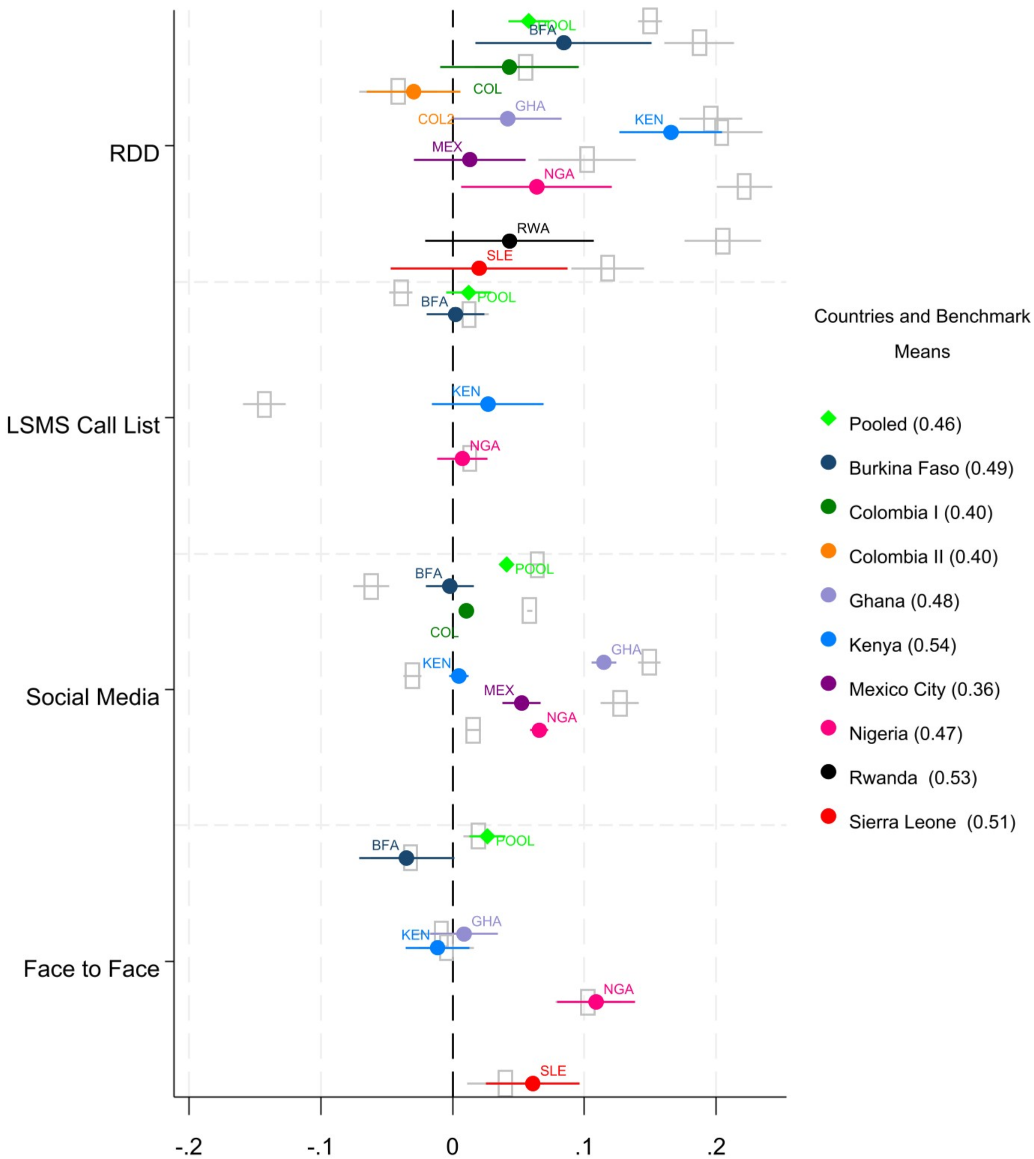
† p<0.05, * p<0.01.

Figure A1: Representation of women in comparison to benchmark estimates, by method and country



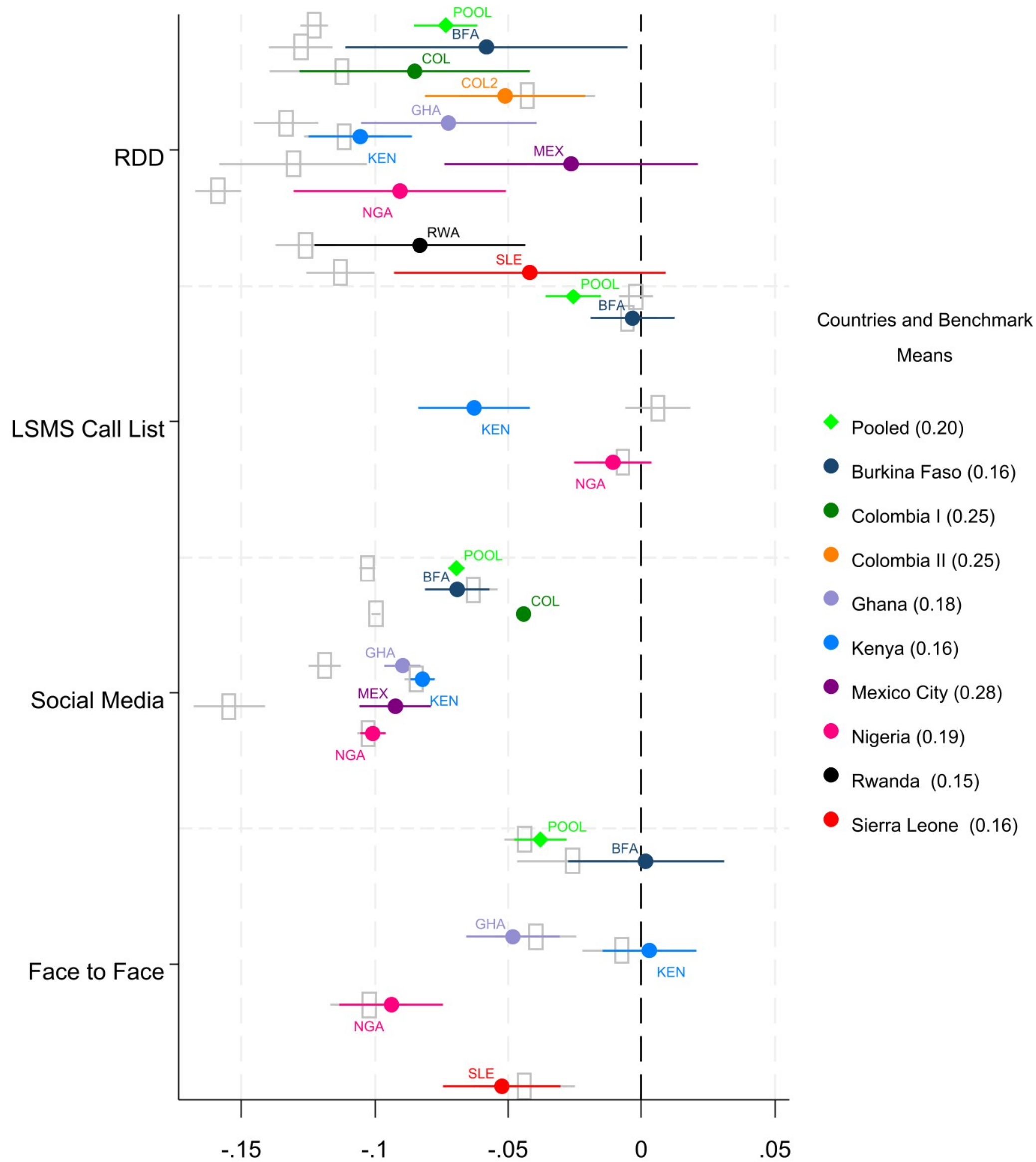
Notes: "Female" is a binary variable, coded as missing if respondents listed 'Other' for gender (Facebook survey only). Estimates are centered around country benchmark sample means (shown in parentheses). Grey boxes represent the corresponding unweighted sample means for each coefficient. RDD surveys conducted by IPA, Call lists by World Bank, Social Media surveys jointly by Facebook Research, University of Maryland and Carnegie Mellon University, and Face-to-Face surveys by Afrobarometer.

Figure A2: Representation of adults aged 18-34, by method and country



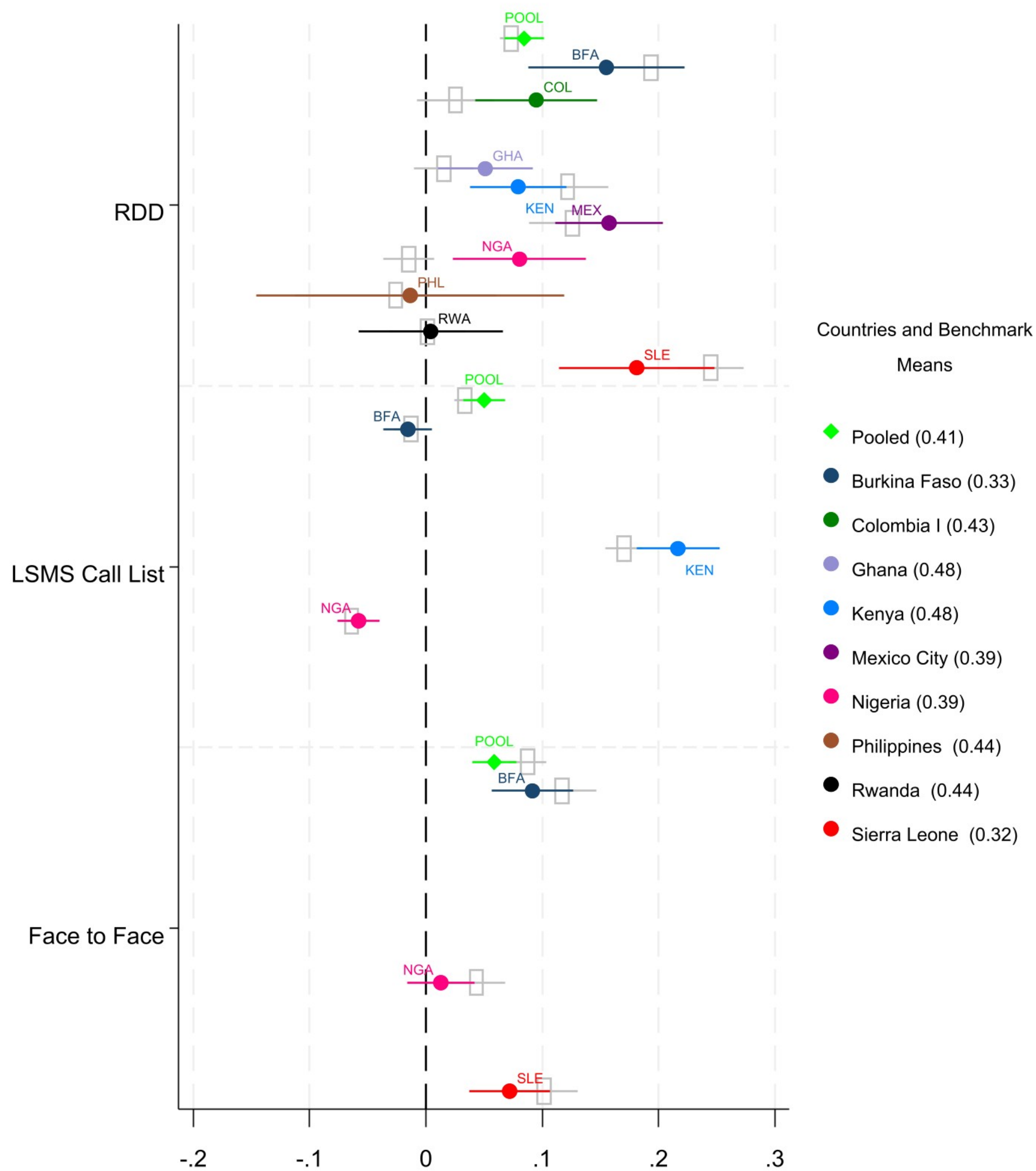
Notes: Estimates are centered around country benchmark sample means (shown in parentheses). Grey boxes represent the corresponding unweighted sample means for each coefficient. RDD surveys conducted by IPA, Call lists by World Bank, Social Media surveys jointly by Facebook Research, University of Maryland and Carnegie Mellon University, and Face-to-Face surveys by the Afrobarometer.

Figure A3: Representation of adults age 55 or higher, by method and country



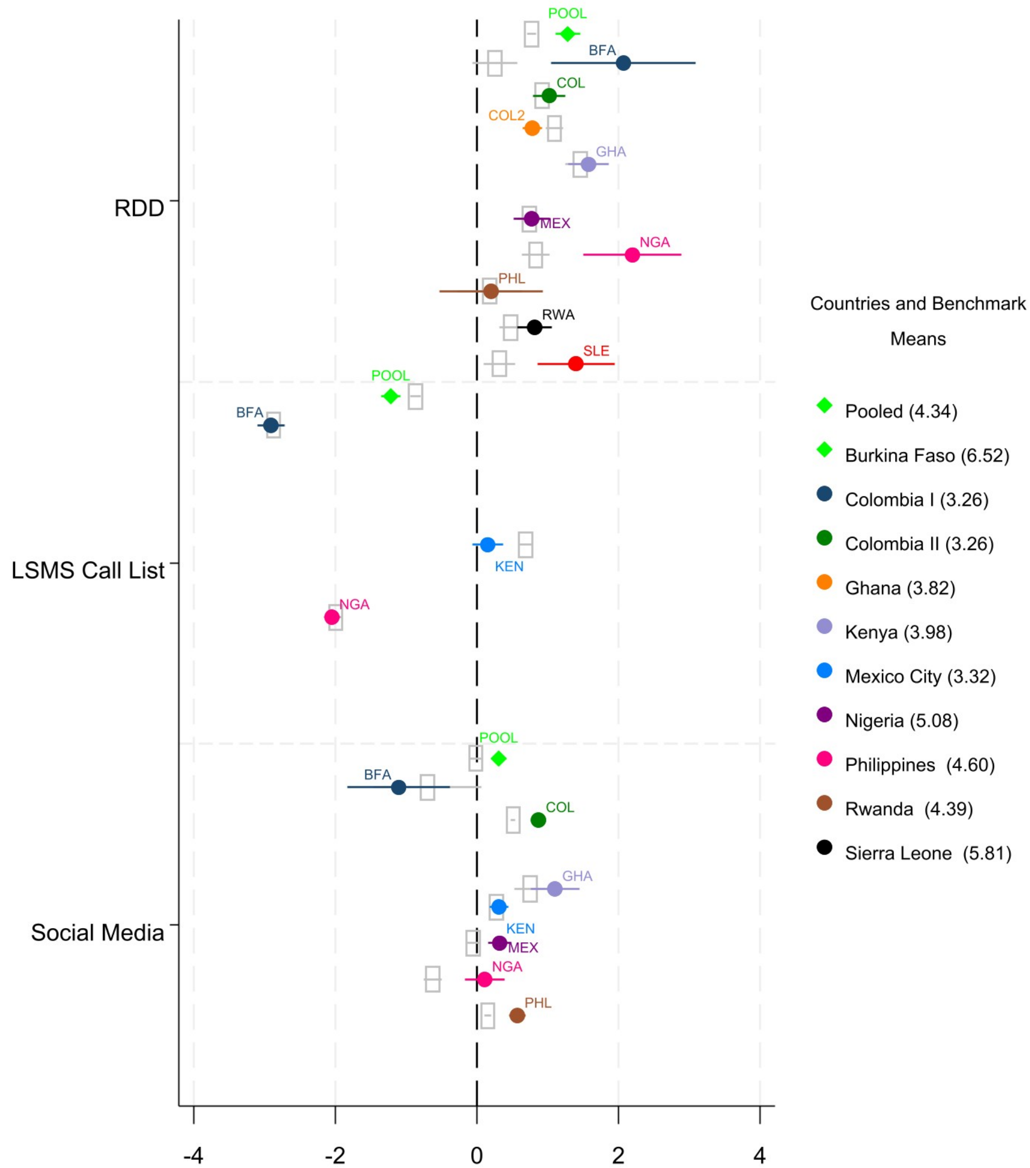
Notes: Estimates are centered around country benchmark sample means (shown in parentheses). Grey boxes represent the corresponding unweighted sample means for each coefficient. RDD surveys conducted by IPA, Call lists by World Bank, Social Media surveys jointly by Facebook Research, University of Maryland and Carnegie Mellon University, and Face-to-Face surveys by the Afrobarometer.

Figure A4: Representation of heads of household, by method and country



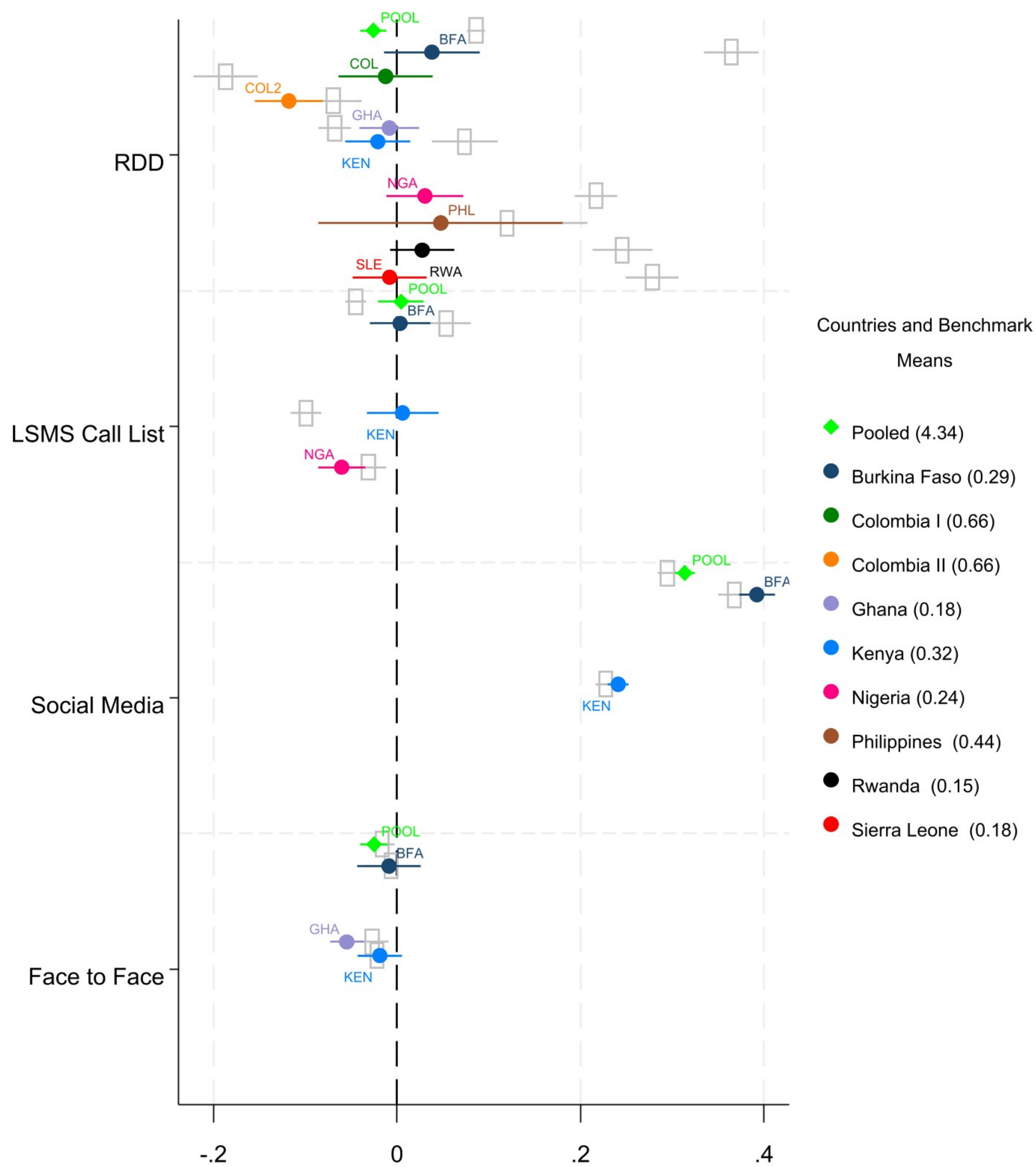
Notes: Estimates are centered around country benchmark sample means (shown in parentheses). Grey boxes represent the corresponding unweighted sample means for each coefficient. RDD surveys conducted by IPA, Call lists by World Bank, Social Media surveys jointly by Facebook Research, University of Maryland and Carnegie Mellon University, and Face-to-Face surveys by the Afrobarometer.

Figure A5: Mean household size, by method and country



Notes: Estimates are centered around country benchmark sample means (shown in parentheses). Grey boxes represent the corresponding unweighted sample means for each coefficient. RDD surveys conducted by IPA, Call lists by World Bank, Social Media surveys jointly by Facebook Research, University of Maryland and Carnegie Mellon University, and Face-to-Face surveys by the Afrobarometer.

Figure A6: Residence in urban area, by method and country

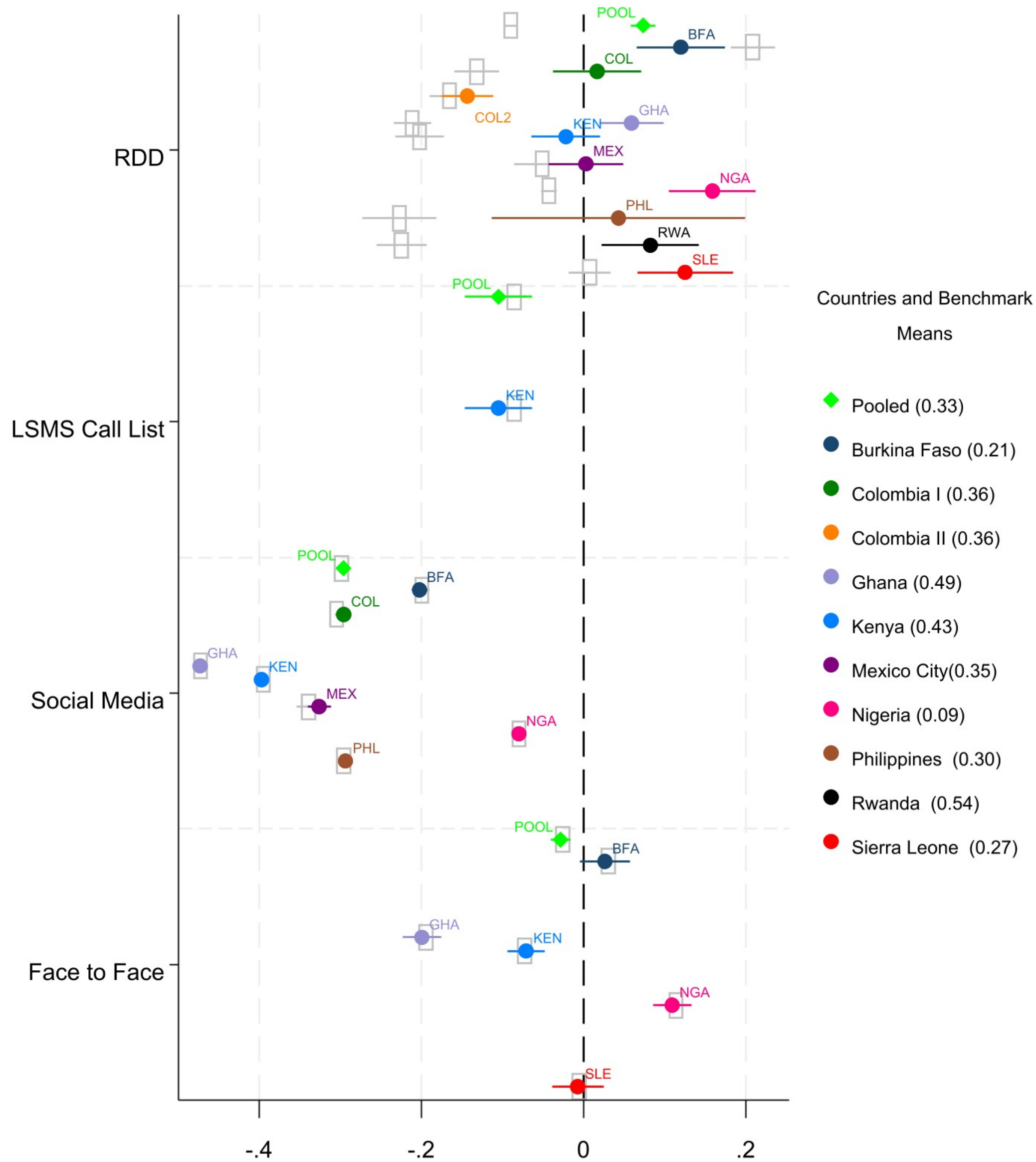


Notes: Estimates are centered around country benchmark sample means (shown in parentheses). Grey boxes represent the corresponding unweighted sample means for each coefficient. RDD surveys conducted by IPA, Call lists by World Bank, Social Media surveys jointly by Facebook Research, University of Maryland and Carnegie Mellon University, and Face-to-Face surveys by the Afrobarometer.

Figure A7: Respondent did not complete any level of formal education, by method and country

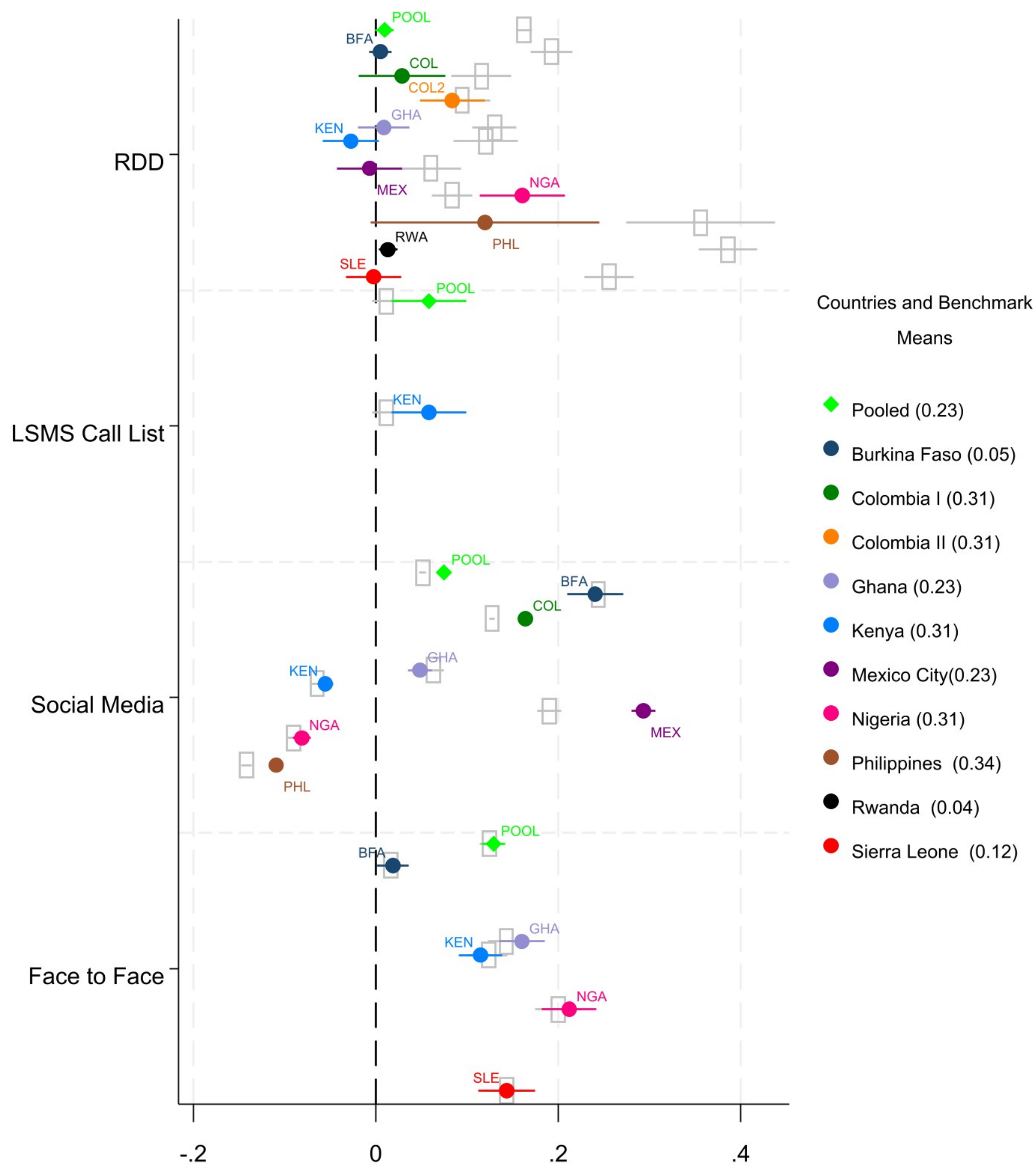
Notes: Estimates are centered around country benchmark sample means (shown in parentheses). Grey boxes represent the corresponding unweighted sample means for each coefficient. RDD surveys conducted by IPA, Call lists by World Bank, Social Media surveys jointly by Facebook Research, University of Maryland and Carnegie Mellon University, and Face-to-Face surveys by the Afrobarometer.

Figure A8: Respondent completed basic level formal education, by method and country



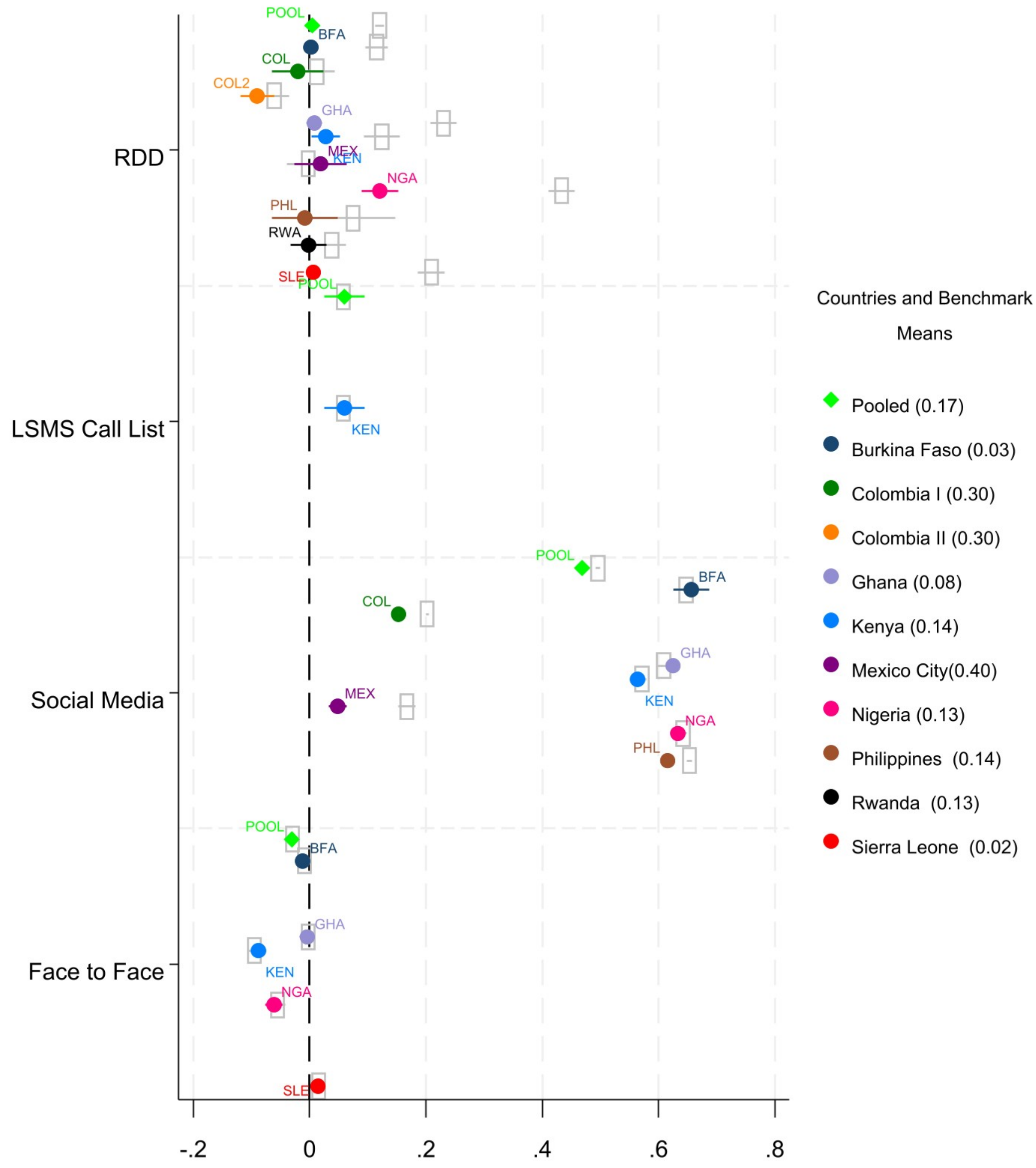
Notes: Estimates are centered around country benchmark sample means (shown in parentheses). Grey boxes represent the corresponding unweighted sample means for each coefficient. RDD surveys conducted by IPA, Call lists by World Bank, Social Media surveys jointly by Facebook Research, University of Maryland and Carnegie Mellon University, and Face-to-Face surveys by the Afrobarometer.

Figure A9: Respondent completed intermediate level formal education, by method and country



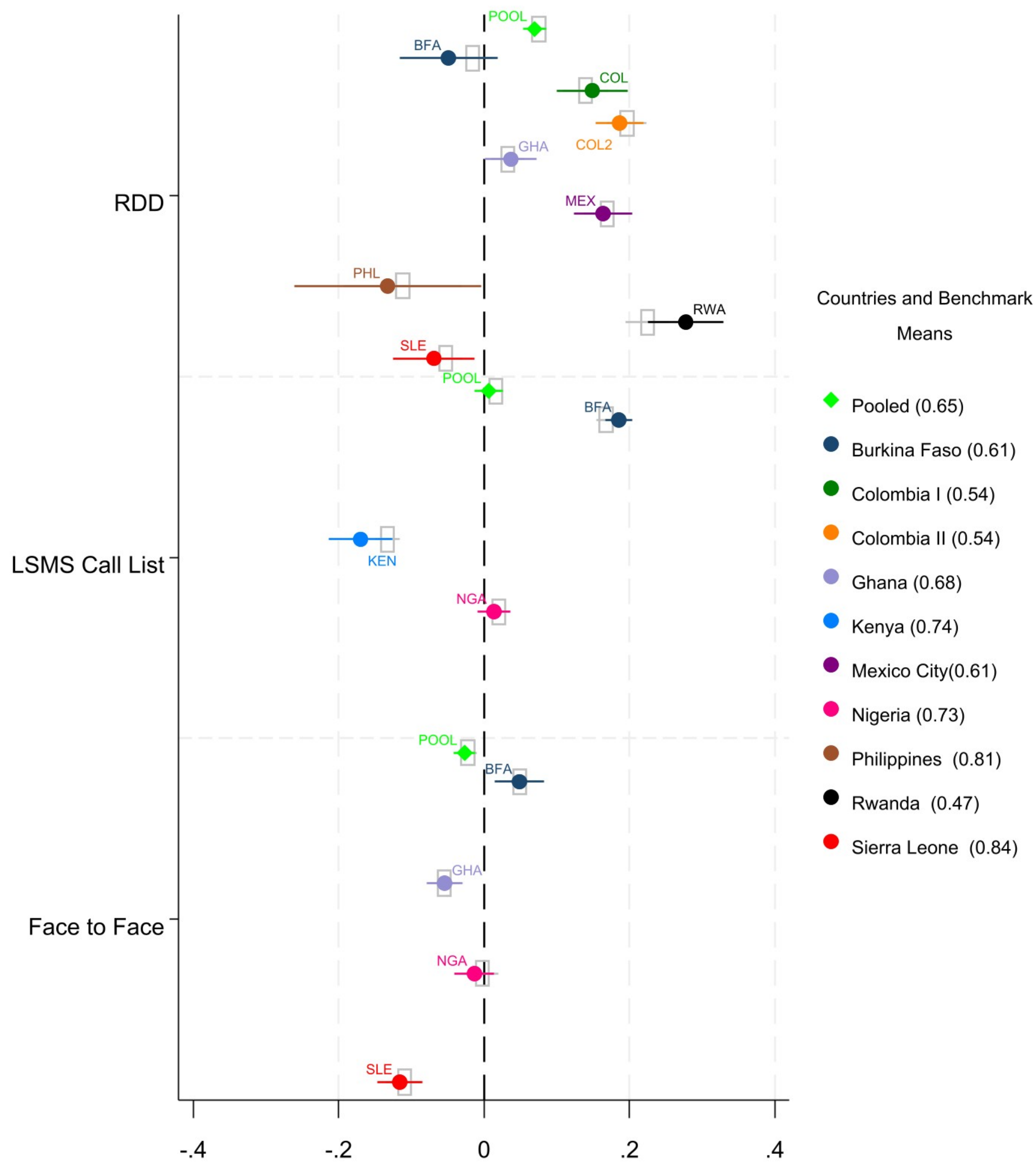
Notes: Estimates are centered around country benchmark sample means (shown in parentheses). Grey boxes represent the corresponding unweighted sample means for each coefficient. RDD surveys conducted by IPA, Call lists by World Bank, Social Media surveys jointly by Facebook Research, University of Maryland and Carnegie Mellon University, and Face-to-Face surveys by the Afrobarometer.

Figure A10: Respondent completed advanced level formal education, by method and country



Notes: Estimates are centered around country benchmark sample means (shown in parentheses). Grey boxes represent the corresponding unweighted sample means for each coefficient. RDD surveys conducted by IPA, Call lists by World Bank, Social Media surveys jointly by Facebook Research, University of Maryland and Carnegie Mellon University, and Face-to-Face surveys by the Afrobarometer.

Figure A11: Respondents who were employed prior to March 2020, by method and country



Notes: Estimates are centered around country benchmark sample means (shown in parentheses). Grey boxes represent the corresponding unweighted sample means for each coefficient. RDD surveys conducted by IPA, Call lists by World Bank, Social Media surveys jointly by Facebook Research, University of Maryland and Carnegie Mellon University, and Face-to-Face surveys by the Afrobarometer.

Table A3: Summary Statistics: % of Female Respondents

	Benchmark	RDD	LSMS Call List	Social Media	Face to Face
All Countries		0.48	0.53	0.37	0.50
Burkina Faso	0.55	0.31	0.55	0.14	0.50
Colombia I	0.52	0.63		0.52	
Colombia II	0.52	0.60			
Ghana	0.54	0.38		0.23	0.50
Kenya	0.52	0.41	0.52	0.28	0.50
Mexico City	0.53	0.53		0.54	
Nigeria	0.52	0.58	0.52	0.20	0.50
Rwanda	0.54	0.36			
Sierra Leone	0.54	0.35			0.50

Notes: Colombia I RDD refers to IPA's study. Colombia II RDD refers to the World Bank High Frequency Phone Survey.

Table A4: Covariate Balance in the % of Respondents Aged 18 to 34

	Benchmark	RDD	RDD Weighted	Call List	Call List Weighted	Social Media	Social Media Weighted	Face to Face	Face to Face Weighted
All Countries		0.15*	0.06*	–0.04*	0.01	0.06*	0.04*	0.02*	0.03*
Burkina Faso	0.49	0.19*	0.08†	0.01	0.00	–0.06*	–0.00	–0.03†	–0.03
Colombia I	0.40	0.06*	0.04			0.06*	0.01*		
Colombia II	0.40	–0.04*	–0.03						
Ghana	0.48	0.20*	0.04†			0.15*	0.11*	–0.01	0.01
Kenya	0.54	0.20*	0.17*	–0.14*	0.03	–0.03*	0.00	–0.00	–0.01
Mexico City	0.36	0.10*	0.01			0.13*	0.05*		
Nigeria	0.47	0.22*	0.06†	0.01	0.01	0.02*	0.07*	0.10*	0.11*
Rwanda	0.53	0.21*	0.04						
Sierra Leone	0.51	0.12*	0.02					0.04*	0.06*

†p<0.05, *p<0.01

Table A5: Summary Statistics: %of Respondents Aged 18 to 34

	Benchmark	RDD	LSMS Call List	Social Media	Face to Face
All Countries		0.61	0.46	0.52	0.52
Burkina Faso	0.49	0.68	0.50	0.43	0.46
Colombia I	0.40	0.45		0.46	
Colombia II	0.40	0.36			
Ghana	0.48	0.68		0.63	0.47
Kenya	0.54	0.74	0.39	0.51	0.53
Mexico City	0.36	0.47		0.49	
Nigeria	0.47	0.69	0.48	0.48	0.57
Rwanda	0.53	0.72			
Sierra Leone	0.51	0.63			0.55

Notes: Colombia I RDD refers to IPA's study. Colombia II RDD refers to the World Bank High Frequency Phone Survey.

Table A6: Covariate Balance in the % of Respondents Aged 35 to 54

	Benchmark	RDD	RDD Weighted	Call List	Call List Weighted	Social Media	Social Media Weighted	Face to Face	Face to Face Weighted
All Countries		–0.03*	0.02†	0.04*	0.01	0.04*	0.03*	0.02*	0.01
Burkina Faso	0.34	–0.06*	–0.03	–0.01	0.00	0.12*	0.07*	0.06*	0.03
Colombia I	0.35	0.06*	0.04			0.04*	0.03*		
Colombia II	0.35	0.08*	0.08*						
Ghana	0.34	–0.06*	0.03			–0.03*	–0.03*	0.05*	0.04*
Kenya	0.31	–0.09*	–0.06*	0.14*	0.04	0.11*	0.08*	0.01	0.01
Mexico City	0.36	0.03	0.01			0.03*	0.04*		
Nigeria	0.34	–0.06*	0.03	–0.01	0.00	0.09*	0.04*	–0.00	–0.02
Rwanda	0.32	–0.08*	0.04						
Sierra Leone	0.33	–0.00	0.02					0.00	–0.01

†p<0.05, *p<0.01

Table A7: Summary Statistics: % of Respondents Aged 35 to 54

	Benchmark	RDD	LSMS Call List	Social Media	Face to Face
All Countries		0.31	0.37	0.38	0.36
Burkina Faso	0.34	0.29	0.34	0.47	0.40
Colombia I	0.35	0.40		0.39	
Colombia II	0.35	0.43			
Ghana	0.34	0.27		0.30	0.38
Kenya	0.31	0.22	0.44	0.42	0.32
Mexico City	0.36	0.39		0.39	
Nigeria	0.34	0.28	0.34	0.43	0.34
Rwanda	0.32	0.25			
Sierra Leone	0.33	0.33			0.34

Notes: Colombia I RDD refers to IPA's study. Colombia II RDD refers to the World Bank High Frequency Phone Survey.

Table A8: Covariate Balance in the % of Respondents Aged over 55

	Benchmark	RDD	RDD Weighted	Call List	Call List Weighted	Social Media	Social Media Weighted	Face to Face	Face to Face Weighted
All Countries		–0.12*	–0.07*	–0.00	–0.03*	–0.10*	–0.07*	–0.04*	–0.04*
Burkina Faso	0.16	–0.13*	–0.06†	–0.01	–0.00	–0.06*	–0.07*	–0.03†	0.00
Colombia I	0.25	–0.11*	–0.09*			–0.10*	–0.04*		
Colombia II	0.25	–0.04*	–0.05*						
Ghana	0.18	–0.13*	–0.07*			–0.12*	–0.09*	–0.04*	–0.05*
Kenya	0.16	–0.11*	–0.11*	0.01	–0.06*	–0.08*	–0.08*	–0.01	0.00
Mexico City	0.28	–0.13*	–0.03			–0.15*	–0.09*		
Nigeria	0.19	–0.16*	–0.09*	–0.01	–0.01	–0.10*	–0.10*	–0.10*	–0.09*
Rwanda	0.15	–0.13*	–0.08*						
Sierra Leone	0.16	–0.11*	–0.04					–0.04*	–0.05*

†p<0.05, *p<0.01

Table A9: Summary Statistics: % of Respondents Aged over 55

	Benchmark	RDD	LSMS Call List	Social Media	Face to Face
All Countries		0.08	0.17	0.10	0.13
Burkina Faso	0.16	0.04	0.16	0.10	0.14
Colombia I	0.25	0.15		0.15	
Colombia II	0.25	0.21			
Ghana	0.18	0.05		0.07	0.14
Kenya	0.16	0.04	0.16	0.07	0.15
Mexico City	0.28	0.14		0.12	
Nigeria	0.19	0.03	0.18	0.09	0.09
Rwanda	0.15	0.03			
Sierra Leone	0.16	0.05			0.11

Notes: Colombia I RDD refers to IPA's study. Colombia II RDD refers to the World Bank High Frequency Phone Survey.

Table A10: Covariate Balance in the % of Respondents that Are Household Heads

	Benchmark	RDD	RDD Weighted	Call List	Call List Weighted	Face to Face	Face to Face Weighted
All Countries		0.07*	0.08*	0.03*	0.05*	0.09*	0.06*
Burkina Faso	0.33	0.19*	0.16*	–0.01	–0.02	0.12*	0.09*
Colombia I	0.43	0.03	0.09*				
Ghana	0.48	0.02	0.05†				
Kenya	0.48	0.12*	0.08*	0.17*	0.22*		
Mexico City	0.39	0.13*	0.16*				
Nigeria	0.39	–0.01	0.08*	–0.06*	–0.06*	0.04*	0.01
Philippines	0.44	–0.03	–0.01				
Rwanda	0.44	0.00	0.00				
Sierra Leone	0.32	0.24*	0.18*			0.10*	0.07*

† p<0.05, * p<0.01

Notes: Colombia I RDD refers to IPA's study.

Table A11: Summary Statistics: % of Respondents that Are Household Heads

	Benchmark	RDD	LSMS Call List	Face to Face
All Countries		0.48	0.43	0.43
Burkina Faso	0.33	0.52	0.32	0.45
Colombia I	0.43	0.46		
Colombia II	0.43			
Ghana	0.48	0.49		
Kenya	0.48	0.61	0.66	
Mexico City	0.39	0.49		
Nigeria	0.39	0.37	0.32	0.43
Philippines	0.44	0.39		
Rwanda	0.44	0.45		
Sierra Leone	0.32	0.57		0.42

Notes: Colombia I RDD refers to IPA's study.

Table A12: Covariate Balance in Household Size

	Benchmark	RDD	RDD Weighted	Call List	Call List Weighted	Social Media	Social Media Weighted
All Countries		0.77*	1.28*	–0.87*	–1.22*	–0.02	0.31*
Burkina Faso	6.52	0.25	2.07*	–2.87*	–2.91*	–0.70	–1.10*
Colombia I	3.26	0.92*	1.02*			0.51*	0.87*
Colombia II	3.26	1.09*	0.78*				
Ghana	3.82	1.46*	1.58*			0.75*	1.10*
Kenya	3.98			0.69*	0.15	0.28*	0.31*
Mexico City	3.32	0.74*	0.77*			–0.05	0.32*
Nigeria	5.08	0.83*	2.20*	–1.99*	–2.05*	–0.62*	0.11
Philippines	4.60	0.18	0.20			0.15*	0.57*
Rwanda	4.39	0.48*	0.82*				
Sierra Leone	5.81	0.32*	1.40*				

† p<0.05, *p<0.01

Table A13: Summary Statistics: Household Size

	Benchmark	RDD	LSMS Call List	Social Media
All Countries		5.16	4.16	4.32
Burkina Faso	6.52	6.77	3.65	5.82
Colombia I	3.26	4.12		3.77
Colombia II	3.26	4.35		
Ghana	3.82	5.29		4.57
Kenya	3.98		4.67	4.26
Mexico City	3.32	4.08		3.27
Nigeria	5.08	5.91	3.09	4.46
Philippines	4.60	4.81		4.75
Rwanda	4.39	4.93		
Sierra Leone	5.81	6.13		

Notes: Colombia I RDD refers to IPA's study. Colombia II RDD refers to the World Bank High Frequency Phone Survey.

Table A14: Covariate Balance in the % of Households that Live in an Urban Area

	Benchmark	RDD	RDD Weighted	Call List	Call List Weighted	Social Media	Social Media Weighted	Face to Face	Face to Face Weighted
All Countries		0.09*	–0.03*	–0.05*	0.00	0.29*	0.31*	–0.02†	–0.02*
Burkina Faso	0.29	0.36*	0.04	0.05*	0.00	0.37*	0.39*	–0.01	–0.01
Colombia I	0.66	–0.19*	–0.01						
Colombia II	0.66	–0.07*	–0.12*						
Ghana	0.18	–0.07*	–0.01					–0.03*	–0.05*
Kenya	0.32	0.07*	–0.02	–0.10*	0.01	0.23*	0.24*	–0.02†	–0.02
Nigeria	0.24	0.22*	0.03	–0.03*	–0.06*				
Philippines	0.44	0.12*	0.05						
Rwanda	0.15	0.25*	0.03						
Sierra Leone	0.18	0.28*	–0.01						

†p<0.05, *p<0.01

Notes: A household is considered to be residing in an urban area if it is located within the lowest available administrative region (harmonized consistently within each country across samples; region for Burkina Faso, municipality for Colombia, county for Kenya, province for the Philippines, and district for Ghana, Sierra Leone and Rwanda), which encompasses any of the urban agglomerations with a population exceeding 300,000, listed in the UN Habitat's 2020 metropolitan areas report (UN Habitat, 2020).

Table A15: Summary Statistics: % of Households that Live in a Major City

	Benchmark	RDD	LSMS Call List	Social Media	Face to Face
All Countries		0.45	0.24	0.60	0.24
Burkina Faso	0.29	0.65	0.34	0.65	0.28
Colombia I	0.66	0.50			
Colombia II	0.66	0.59			
Ghana	0.18	0.11			0.15
Kenya	0.32	0.39	0.22	0.55	0.30
Nigeria	0.24	0.46	0.21		
Philippines	0.44	0.50			
Rwanda	0.15	0.40			
Sierra Leone	0.18	0.46			

Notes: Colombia I RDD refers to IPA's study. Colombia II RDD refers to the World Bank High Frequency Phone Survey. A household is considered to be residing in an urban area if it is located within the lowest available administrative region (harmonized consistently within each country across samples; region for Burkina Faso, municipality for Colombia, county for Kenya, province for the Philippines, and district for Ghana, Sierra Leone and Rwanda), which encompasses any of the urban agglomerations with a population exceeding 300,000, listed in the UN Habitat's 2020 metropolitan areas report (Habitat 2020).

Table A16: Covariate Balance in the % of Respondents with less than a Basic Formal Education

	Benchmark	RDD	RDD Weighted	Call List	Call List Weighted	Social Media	Social Media Weighted	Face to Face	Face to Face Weighted
All Countries		–0.20*	–0.07*	0.02*	–0.01	–0.25*	–0.25*	–0.10*	–0.10*
Burkina Faso	0.70	–0.52*	–0.13*			–0.69*	–0.69*	–0.09*	–0.08*
Colombia I	0.04	0.00	–0.03*			–0.03*	–0.02*		
Colombia II	0.04	0.07*	0.10*						
Ghana	0.21	–0.15*	–0.08*			–0.20*	–0.20*	0.04*	0.03†
Kenya	0.12	–0.04*	0.02	0.02*	–0.01	–0.11*	–0.11*	0.04*	0.04*
Mexico City	0.02	–0.01	–0.02*			–0.02*	–0.02*		
Nigeria	0.48	–0.45*	–0.16*			–0.47*	–0.47*	–0.32*	–0.32*
Philippines	0.21	–0.19*	–0.15*			–0.21*	–0.20*		
Rwanda	0.29	–0.20*	–0.09*						
Sierra Leone	0.58	–0.50*	–0.13*					–0.20*	–0.20*

† p<0.05, *p<0.01

Table A17: Summary Statistics: % of Respondents with less than a Basic Formal Education

	Benchmark	RDD	LSMS Call List	Social Media	Face to Face
All Countries		0.07	0.14	0.01	0.31
Burkina Faso	0.70	0.19		0.01	0.62
Colombia I	0.04	0.05		0.02	
Colombia II	0.04	0.11			
Ghana	0.21	0.06		0.01	0.25
Kenya	0.12	0.08	0.14	0.01	0.16
Mexico City	0.02	0.01		0.00	
Nigeria	0.48	0.03		0.00	0.15
Philippines	0.21	0.01		0.00	
Rwanda	0.29	0.07			
Sierra Leone	0.58	0.08			0.39

Notes: Colombia I RDD refers to IPA's study. Colombia II RDD refers to the World Bank High Frequency Phone Survey.

Table A18: Covariate Balance in the % of Respondents with Basic Education

	Benchmark	RDD	RDD Weighted	Call List	Call List Weighted	Social Media	Social Media Weighted	Face to Face	Face to Face Weighted
All Countries		–0.09*	0.07*	–0.09*	–0.10*	–0.30*	–0.30*	–0.03*	–0.03*
Burkina Faso	0.21	0.21*	0.12*			–0.20*	–0.20*	0.03†	0.03
Colombia I	0.36	–0.13*	0.02			–0.30*	–0.30*		
Colombia II	0.36	–0.16*	–0.14*						
Ghana	0.49	–0.21*	0.06*			–0.47*	–0.47*	–0.19*	–0.20*
Kenya	0.43	–0.20*	–0.02	–0.09*	–0.10*	–0.39*	–0.40*	–0.07*	–0.07*
Mexico City	0.35	–0.05*	0.00			–0.34*	–0.33*		
Nigeria	0.09	–0.04*	0.16*			–0.08*	–0.08*	0.11*	0.11*
Philippines	0.30	–0.23*	0.04			–0.30*	–0.29*		
Rwanda	0.54	–0.22*	0.08*						
Sierra Leone	0.27	0.01	0.13*					–0.00	–0.01

† p<0.05, * p<0.01

Table A19: Summary Statistics: % of Respondents with Basic Education

	Benchmark	RDD	LSMS Call List	Social Media	Face to Face
All Countries		0.24	0.35	0.02	0.27
Burkina Faso	0.21	0.42		0.02	0.25
Colombia I	0.36	0.22		0.05	
Colombia II	0.36	0.19			
Ghana	0.49	0.28		0.02	0.29
Kenya	0.43	0.23	0.35	0.04	0.36
Mexico City	0.35	0.30		0.01	
Nigeria	0.09	0.04		0.01	0.20
Philippines	0.30	0.12		0.00	
Rwanda	0.54	0.33			
Sierra Leone	0.27	0.28			0.27

Notes: Colombia I RDD refers to IPA's study. Colombia II RDD refers to the World Bank High Frequency Phone Survey.

Table A20: Covariate Balance in the % of Respondents with Intermediate Education

	Benchmark	RDD	RDD Weighted	Call List	Call List Weighted	Social Media	Social Media Weighted	Face to Face	Face to Face Weighted
All Countries		0.16*	0.01	0.01	0.06*	0.05*	0.07*	0.12*	0.13*
Burkina Faso	0.05	0.19*	0.01			0.24*	0.24*	0.02†	0.02†
Colombia I	0.31	0.12*	0.03			0.13*	0.16*		
Colombia II	0.31	0.09*	0.08*						
Ghana	0.23	0.13*	0.01			0.06*	0.05*	0.14*	0.16*
Kenya	0.31	0.12*	–0.03	0.01	0.06*	–0.06*	–0.06*	0.12*	0.11*
Mexico City	0.23	0.06*	–0.01			0.19*	0.29*		
Nigeria	0.31	0.08*	0.16*			–0.09*	–0.08*	0.20*	0.21*
Philippines	0.34	0.36*	0.12			–0.14*	–0.11*		
Rwanda	0.04	0.39*	0.01†						
Sierra Leone	0.12	0.26*	–0.00					0.14*	0.14*

† p<0.05, *p<0.01

Table A21: Summary Statistics: % of Respondents with Intermediate Education

	Benchmark	RDD	LSMS Call List	Social Media	Face to Face
All Countries		0.39	0.32	0.30	0.33
Burkina Faso	0.05	0.24		0.30	0.07
Colombia I	0.31	0.40		0.43	
Colombia II	0.31	0.40			
Ghana	0.23	0.36		0.29	0.37
Kenya	0.31	0.43	0.32	0.24	0.43
Mexico City	0.23	0.31		0.42	
Nigeria	0.31	0.39		0.22	0.51
Philippines	0.34	0.61		0.20	
Rwanda	0.04	0.43			
Sierra Leone	0.12	0.37			0.26

Notes: Colombia I RDD refers to IPA's study. Colombia II RDD refers to the World Bank High Frequency Phone Survey.

Table A22: Covariate Balance in the % of Respondents with Advanced Education

	Benchmark	RDD	RDD Weighted	Call List	Call List Weighted	Social Media	Social Media Weighted	Face to Face	Face to Face Weighted
All Countries		0.12*	0.00	0.06*	0.06*	0.50*	0.47*	–0.03*	–0.03*
Burkina Faso	0.03	0.12*	0.00			0.65*	0.66*	–0.01†	–0.01*
Colombia I	0.30	0.01	–0.02			0.20*	0.15*		
Colombia II	0.30	–0.06*	–0.09*						
Ghana	0.08	0.23*	0.01			0.61*	0.62*	–0.00	–0.00
Kenya	0.14	0.12*	0.03†	0.06*	0.06*	0.57*	0.56*	–0.09*	–0.09*
Mexico City	0.40	–0.00	0.02			0.17*	0.05*		
Nigeria	0.13	0.43*	0.12*			0.64*	0.63*	–0.05*	–0.06*
Philippines	0.14	0.07†	–0.01			0.65*	0.62*		
Rwanda	0.13	0.04*	–0.00						
Sierra Leone	0.02	0.21*	0.01					0.02*	0.01†

† p<0.05, * p<0.01

Table A23: Summary Statistics: % of Respondents with Advanced Education

	Benchmark	RDD	LSMS Call List	Social Media	Face to Face
All Countries		0.29	0.20	0.67	0.05
Burkina Faso	0.03	0.15		0.68	0.02
Colombia I	0.30	0.33		0.50	
Colombia II	0.30	0.23			
Ghana	0.08	0.31		0.68	0.07
Kenya	0.14	0.27	0.20	0.71	0.05
Mexico City	0.40	0.38		0.56	
Nigeria	0.13	0.56		0.77	0.07
Philippines	0.14	0.26		0.80	
Rwanda	0.13	0.17			
Sierra Leone	0.02	0.23			0.04

Notes: Colombia I RDD refers to IPA's study. Colombia II RDD refers to the World Bank High Frequency Phone Survey.

Table A24: Covariate Balance in the % of Households with at least one Child Attending School (before February 2020)

	Benchmark	RDD	RDD Weighted	Call List	Call List Weighted
All Countries		–0.22*	–0.18*	0.16*	0.13*
Burkina Faso	0.66	0.05*	0.15*	0.16*	0.13*
Colombia I	0.90	–0.28*	–0.28*		
Colombia II	0.90	0.02	0.01		
Ghana	0.99	–0.33*	–0.27*		
Mexico City	0.96	–0.53*	–0.55*		
Rwanda	0.96	–0.30*	–0.20*		
Sierra Leone	0.98	–0.20*	–0.09*		

† p<0.05, * p<0.01

Table A25: Summary Statistics: % of Households with at least one Child Attending School (before February 2020)

	Benchmark	RDD	LSMS Call List
All Countries		0.68	0.87
Burkina Faso	0.66	0.71	0.81
Colombia I	0.90	0.63	
Colombia II	0.90	0.92	
Ghana	0.99	0.66	
Mexico City	0.96	0.44	
Rwanda	0.96	0.65	
Sierra Leone	0.98	0.79	

Notes: Colombia I RDD refers to IPA's study. Colombia II RDD refers to the World Bank High Frequency Phone Survey.

Table A26: Covariate Balance in the % of Employed Respondents (before February 2020)

	Benchmark	RDD	RDD Weighted	Call List	Call List Weighted	Face to Face	Face to Face Weighted
All Countries		0.08*	0.07*	0.02*	0.01	–0.02*	–0.03*
Burkina Faso	0.61	–0.02	–0.05	0.17*	0.19*	0.05*	0.05*
Colombia I	0.54	0.14*	0.15*				
Colombia II	0.54	0.20*	0.19*				
Ghana	0.68	0.03*	0.04†			–0.05*	–0.05*
Kenya	0.74			–0.13*	–0.17*		
Mexico City	0.61	0.17*	0.16*				
Nigeria	0.73			0.02*	0.01	–0.00	–0.01
Philippines	0.81	–0.11*	–0.13†				
Rwanda	0.47	0.23*	0.28*				
Sierra Leone	0.84	–0.05*	–0.07†			–0.11*	–0.12*

† p<0.05, * p<0.01

Table A27: Summary Statistics: % of Employed Respondents (before February 2020)

	Benchmark	RDD	LSMS Call List	Face to Face
All Countries		0.71	0.71	0.69
Burkina Faso	0.61	0.60	0.78	0.66
Colombia I	0.54	0.69		
Colombia II	0.54	0.74		
Ghana	0.68	0.71		0.62
Kenya	0.74		0.61	
Mexico City	0.61	0.77		
Nigeria	0.73		0.75	0.72
Philippines	0.81	0.67		
Rwanda	0.47	0.71		
Sierra Leone	0.84	0.79		0.73

Notes: Colombia I RDD refers to IPA's study. Colombia II RDD refers to the World Bank High Frequency Phone Survey.