Survey Sampling in the Global South Using Facebook Advertisements

Leah R. Rosenzweig*ⓡ[1], Parrish Bergquist[2], Katherine Hoffmann Pham[3], Francesco Rampazzo[4,5], and Matto Mildenberger[6]

[1]Graduate School of Business, Stanford University
[2]McCourt School of Public Policy, Georgetown University
[3]Stern School of Business, New York University
[4]Centre for Population Change, University of Southampton
[5]Max Planck Institute for Demographic Research
[6]Department of Political Science, University of California Santa Barbara

July 28, 2020

Abstract

Survey research in the Global South traditionally requires large budgets and lengthy fieldwork, for which researchers hire local enumerators to conduct face-to-face surveys with respondents. However, much of the world's population is now digitally accessible, offering an opportunity for researchers with limited budgets and those seeking to study settings where in-person contact is impossible, such as natural disasters, violent conflicts, and pandemics. In this paper, we evaluate whether Facebook advertising can be used to cost-effectively generate representative survey samples in the Global South. We introduce a framework for evaluating quality in Facebook survey samples, highlighting key trade-offs for researchers considering the platform. We then quota-sample respondents in two countries: Mexico (n=5,168) and Kenya (n=1,452) to evaluate how well these samples perform on a diverse set of survey indicators related to both internal and external validity. We find that while the Facebook platform can quickly and cheaply recruit respondents, these samples tend to be more male, more educated, and more urban than the overall national populations. Applying post-stratification weighting after oversampling key demographic variables ameliorates, but does not fully overcome, these initial sample imbalances. Our analysis demonstrates the considerable potential of Facebook advertisements to cost-effectively conduct research in diverse global settings.

*ⓡOrder of the authors was randomized.

1 Introduction

Survey research in the Global South traditionally requires large budgets and lengthy fieldwork, for which researchers hire local enumerators to conduct face-to-face surveys with respondents. Historically, there have been few alternatives to in-person recruitment due to poor electricity coverage and limits to phone and internet connectivity. However, today much of the world's population is digitally accessible; an estimated 67% of the world's population has a mobile phone subscription, and approximately 49% of the world's population has access to mobile internet (GSMA, 2020; Rotondi et al., 2020). This growing connectivity represents an opportunity for researchers with limited budgets and those seeking to study settings where the traditional resource-intensive, in-person contact model of research is impossible, such as natural disasters, violent conflicts, and pandemics. Now more than ever, as in-person research has come to a stand-still due to the COVID-19 crisis, social science researchers require creative ways to recruit respondents remotely. This paper describes one such method of recruiting online samples that can complement traditional on-the-ground fieldwork and data collection.

This paper explores the opportunity that Facebook's advertising platform presents to recruit respondents online in the Global South. According to Facebook, the company had 2.45 billion monthly active users in September 2019 (Facebook Inc., 2019), nearly a third of the global population. For this same month, the company reported an average of 1.62 billion daily active users (Facebook Inc., 2019), about 20% of all living humans. Facebook advertising thus offers researchers unparalleled access to nationally, culturally, and demographically diverse populations around the world. Furthermore, Facebook offers efficiency in terms of time and money compared to other modes of online survey recruitment. It also enables micro-targeting, allowing scholars to recruit samples from specific, difficult-to-reach sub-populations such as would-be first-time voters or residents of a particular locality.

A recent literature has begun to explore and evaluate Facebook as a platform for recruiting survey samples. Facebook has been used to target respondents from demographic groups with specific characteristics or behaviors, including young smokers (Ramo and Prochaska, 2012; Ramo et al., 2014; Kapp et al., 2013); political activists in Thailand and Germany (Jäger, 2017); Polish migrants in Austria, Ireland, Switzerland, and the UK (Pötzschke and Braun, 2017); and primary voters in local elections (Hirano et al., 2015). Facebook has also been used to construct a panel of survey respondents from East Africa (Rosenzweig and Zhou, 2019), to recruit experimental subjects in Brazil (Samuels and Zucco Jr, 2014), to deliver and test political advertisements (Broockman and Green, 2014; Ryan, 2012), and to conduct surveys during the COVID-19 pandemic (Grow et al., 2020; Bicalho et al., 2020).

To date, many of these studies emphasize the utility of Facebook advertising for recruiting specific target populations that are of theoretical interest to a given research question. However, as Boas et al. (2018) demonstrate in India and the United States, Facebook's high penetration rate and the diversity of its user base make the platform an attractive opportunity for recruiting nationally representative samples. Zhang et al. (2018) use targeted Facebook advertising to cost-effectively recover approximately nationally representative estimates of public opinion in the United States. In this paper, we ask: is this kind of quota-based sampling method also successful in the Global South?

As with more traditional forms of survey research, Zhang et al. suggest that researchers must weigh the advantages offered by Facebook's platform against limitations that could compromise the quality of samples recruited through Facebook. The quality of Facebook samples, similar to samples recruited by any survey mode, can be evaluated on the basis of both internal and external validity considerations. Relevant questions about internal validity include: Are people paying

attention? Do they answer survey questions in ways that reflect their real views? Are experimental interventions actually delivered and received? Relevant questions about external validity include: is the sample representative of the target populations of interest? Can re-weighting satisfactorily correct imbalances due to selection bias?

In this paper, we extend Zhang et al.'s work in two primary directions. First, we introduce a framework for evaluating quality in Facebook survey samples, highlighting key tradeoffs in order to guide researchers who are considering running surveys on the platform. Second, we provide data from two countries, one in Latin America and the other in sub-Saharan Africa, to evaluate the representativeness of Facebook-recruited samples in the Global South. Due to lower levels of internet access, literacy, and Facebook marketing investment in these regions, it is not clear that findings from US-based studies will generalize to other national contexts. We evaluate this question empirically.

We assess cost-quality tradeoffs involved in using Facebook to quota-sample respondents in two countries in the Global South: Mexico (n=5,168) and Kenya (n=1,452). We evaluate how well these samples perform on quality indicators related to both internal and external validity. Specifically, we compare samples recruited through Facebook to the population of Facebook users in these countries as a whole, to nationally representative in-person survey samples, and to the national adult (18 years and over) population as measured in each country's national census. We evaluate representativeness along two different respondent dimensions — demographics and opinions — both before and after the application of weights designed to correct imbalances in the sample. We also describe some of the practical considerations involved in Facebook sampling. We conclude with a discussion of the tradeoffs between the quality of Facebook and other samples, and the costs associated with each sampling mode.

In evaluating the demographic makeup of our samples, we find that while the Facebook platform enables researchers to quickly and cheaply recruit respondents, these samples tend to be more male, more educated, and more urban than the national populations in Mexico and Kenya. Micro-targeting using Facebook advertisements allowed us to conduct quota-based sampling and thereby obtain a more diverse sample with greater coverage across key demographic variables. However, the data underlying Facebook's micro-targeting capability are not entirely accurate. Applying post-stratification weights on the basis of demographics — age, sex, education, and geography — helps to ameliorate, but does not fully overcome, initial imbalances in the sample.

In terms of political attitudes and engagement, our samples exhibit some similarities to, and some divergence from, the respective national populations. Mexican Facebook respondents demonstrate greater concern for climate change. In Kenya, respondents' political attitudes reflect those of the nation but the Facebook sample appears to be more politically engaged and active than the population as a whole. In Kenya, we also show that we are able to replicate a canonical behavioral survey experiment using our Facebook respondents. Taken together, these findings suggest that the value of Facebook samples will depend on the type of research question asked and the target population of interest. Facebook samples should be assessed for representativeness along multiple dimensions, and against various population benchmarks.

2 Assessing quality in public opinion samples

2.1 Defining quality

We begin by developing a definition of survey quality which builds upon a growing body of research assessing cost-quality tradeoffs in public opinion research. Scholars have assessed the quality of samples recruited through Amazon's Mechanical Turk (MTurk) (Berinsky et al., 2012; Huff and

Tingley, 2015), Prime Panels (Litman et al., 2017), Lucid Fulcrum Exchange (Coppock and McClellan, 2019), Google Consumer Surveys (Santoso et al., 2016), and Facebook Advertising (Kosinski et al., 2015; Jäger, 2017; Boas et al., 2018; Zhang et al., 2018). Building on this work, we assess quality using the canonical paradigm of internal and external validity.

Internal and external validity are inextricably linked in our setting, because some internal validity measures are precursors for external validity. For example, we describe the extent to which Facebook accurately identifies individuals in particular geographic and demographic strata as a problem of internal validity. However, this accuracy is of interest to us because we in turn use these strata to recruit and reweight respondents to produce an externally valid, nationally representative sample.

2.2 Internal Validity

Internal validity refers to whether a study accurately captures the phenomena it aims to measure. In our setting, the core internal validity questions are: Is the survey able to target respondents with the characteristics of interest? Do the survey responses actually come from the population targeted by the survey? Are the responses thoughtful and accurate?

To answer these questions, we first assess whether we successfully recruit individuals whose Facebook-reported characteristics match the criteria in our sampling strata, such as gender, location, and age. Next, we ask whether Facebook correctly identifies these respondents. To do this, we compare individuals' self-reported demographic and geographic characteristics with the characteristics targeted by the Facebook ad which recruited each individual.[1] Finally, other characteristics like respondent attentiveness have implications for the validity of a survey's descriptive findings. We thus assess how our Facebook sample performs on survey completion times and standard measures of respondent attention.

2.3 External validity

External validity refers to whether a study's conclusions can be generalized to a broader population of interest. In our case, key questions include: Is the survey sample drawn using Facebook ads representative of the Facebook population? Is the survey sample drawn using Facebook ads representative of the national adult populations in our target countries?

In order for responses collected via Facebook to generalize to broader populations, three assumptions should be met: positivity, ignorability, and independence. The first two of these are drawn from Zhang et al. (2018), while the third is introduced here. All researchers choosing a survey mode face tradeoffs, and most online or phone surveys, as well as many in-person surveys, suffer from violations of these assumptions. Our goal is to clearly document these tradeoffs in the context of the Facebook platform.

2.3.1 Positivity

In order to construct externally valid inferences by reweighting our survey responses, we must first assume that, for each combination of strata and observable characteristics used in the weighting, there is a non-zero probability that the associated respondents will take the survey. Borrowing from the causal inference framework, we refer to this assumption as the *positivity* assumption.

[1] As we discuss in further detail below, we have these individual-level, Facebook-reported geographic and demographic data because we know from which of our targeted ads each individual was recruited. The platform targets individuals based on a match between a) the demographic and geographic characteristics defined in our advertising strata and b) the individuals' demographic characteristics and geographic location as measured by Facebook.

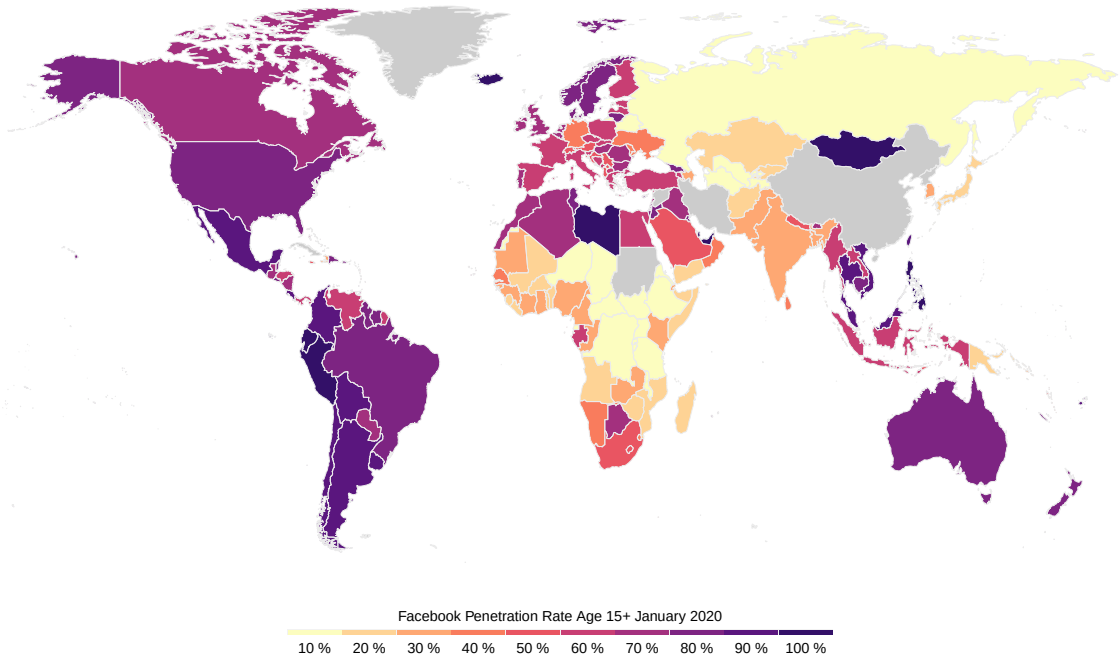

Figure 1: Facebook penetration rates among users age 15 years and older around the world. The rate is computed as of January 2020 with Facebook and United Nations data (United Nations, 2019; Ševčíková, 2020). The Facebook data is courtesy of palotti$_{m}onitoring_{2}$020.

Formally, define R_i as an indicator of whether member i of the population takes the survey; $\mathbf{D}_i$ as the vector of respondent i's characteristics that researchers used to construct strata for quota sampling; and $\mathbf{X}_i$ as the vector of respondent i's characteristics that are not used to construct strata but that are measured in the survey and used for weighting. Then, the positivity assumption states that:

Assumption 1 *Positivity*

$$\Pr(R_i = 1 \mid \mathbf{D}_i = \mathbf{d}_i,\ \mathbf{X}_i = \mathbf{x}_i) \quad > \quad 0, \qquad \forall \quad \mathbf{d}_i \in Supp(\mathbf{D}_i),\ \mathbf{x}_i \in Supp(\mathbf{X}_i) \tag{1}$$

A violation of the positivity assumption would occur if there are certain sample strata that Facebook is unable to fill. A violation of the positivity assumption would also arise if there are some groups of individuals who are not on Facebook at all, making it impossible to recruit them into our sample. In practice, for efforts to collect nationally representative data, this assumption is almost always violated since Facebook's penetration is less than 100 percent. The likelihood of violations varies across countries, since internet and Facebook use differ by country (see Figure 1).[2]

2.3.2 Ignorability

Second, we must assume that conditional on strata and observed respondent characteristics, respondents who take the survey are, in expectation, identical to those who did not take the survey.

[2]Facebook penetration rates should be less than 100%. The data shown in Figure 1 are biased due to 1) undercounting of national population estimates and 2) fake and duplicated Facebook accounts. Facebook suggests that in 2019, 11% of accounts on its platform were duplicates. This percentage is higher in South East Asia (U.S. Securities and Exchange Commission, 2019).

Again following the causal inference framework, we refer to this assumption as *ignorability*. This is equivalent to assuming that those not included in the sample are Missing at Random (MAR).

Formally, define Y_i as person i's response to a survey question. Then, the ignorability assumption states:

Assumption 2 *Ignorability*

$$Y_i \perp\!\!\!\perp R_i \;\mid\; \mathbf{D}_i = \mathbf{d}_i, \quad \mathbf{X}_i = \mathbf{x}_i, \quad \forall \;\; \mathbf{d}_i \in Supp(\mathbf{D}_i), \quad \mathbf{x}_i \in Supp(\mathbf{X}_i) \tag{2}$$

A violation of the ignorability assumption would occur if Facebook respondents from within a stratum are significantly different from those in the stratum who do not respond. This challenge arises frequently in surveys, since those who select into responding often have more time, or are more pro-social, than those who refuse the survey. Particularly in the Global South, we might worry that wealthier individuals are more likely to take the survey if they have easier access to the internet or are better able to navigate the digital survey interface. Survey incentives might shape such biases: while offering survey incentives may encourage greater participation from low-income users who pay to access the internet, incentives may motivate these low-income users more than others. We provided such incentives for respondents in Kenya, but not in Mexico; future work could directly compare samples recruited in the same country with and without incentives.

Researchers using Facebook for sample recruitment face one additional concern that is particular to Facebook: the ad platform's targeting algorithm may exacerbate selection bias. Subject to a set of targeting constraints, Facebook deliberately serves ads to those who are most likely to click and to respond. Our demographic and geographic quotas help to reduce this bias. Still, they are unlikely to fully erase the problem, since the internal workings of Facebook's targeting algorithm are unknown and Facebook may make targeting decisions using variables that we did not include in the stratification design. Encouragingly, the quotas for each of our ad targeting strata were generally too small for Facebook's algorithm to fully optimize ad delivery.[3] However, for researchers running large-scale surveys over a longer time horizon, it is worth noting that the algorithm may find a progressively less representative sample within a stratum over time.

2.3.3 Corollary: Independence

Finally, within each stratum, we assume that each respondent is independently sampled from the broader population of individuals in the stratum. This means that the response rates of two individuals within a stratum should not be correlated with each other, relative to the broader population of the stratum. Similarly, response rates from individuals in one stratum should not be correlated with the response rates of individuals in another stratum. More formally:

Assumption 3 *Independence*

$$R_i \;\mid\; \mathbf{D}_i = \mathbf{d}_i,\; \mathbf{X}_i = \mathbf{x}_i \qquad \perp\!\!\!\perp \qquad R_j \;\mid\; \mathbf{D}_j = \mathbf{d}_j,\; \mathbf{X}_j = \mathbf{x}_j, \tag{3}$$

$$\forall \;\; \mathbf{d}_i \in Supp(\mathbf{D}_i),\; \mathbf{d}_j \in Supp(\mathbf{D}_j), \quad \mathbf{x}_i \in Supp(\mathbf{X}_i),\; \mathbf{x}_j \in Supp(\mathbf{X}_j)$$

Most survey re-weighting approaches assume an independent and identically distributed (i.i.d.) random sample from each stratum of interest, but on Facebook this may not be the case due to link sharing among users. Individuals may thus come into a survey by invitation or recruitment within

[3] Facebook typically reported that the algorithm was still in the training phase, suggesting that its targeting algorithm was not functioning optimally.

a smaller social network. For example, women aged 25-30 might share the link with other friends in their age cohort (leading to non-independent response rates within the stratum), or they may share the link with husbands aged 25-30 (leading to correlated response rates across strata). Sharing may be particularly likely when there are generous rewards for survey completion. If appropriate safeguards are not in place, we might also be concerned that the same individuals attempt to complete the survey repeatedly in an effort to secure multiple units of compensation.[4]

Note that this independence violation can be seen as a specific violation of the ignorability assumption. Network-recruited respondents share an unobserved common characteristic (their location in the network) which is correlated with their representation in the sample.

3 Procedures and Data Collection

3.1 Case selection

We selected Mexico and Kenya as typical cases for Facebook sampling in the Global South for several reasons. First, Facebook penetration in both countries is about average for their respective regions. Similar to other African countries, a minority (25%) of the Kenyan population over 15 years old uses Facebook. Comparisons to Facebook penetration in South Africa (45%), Nigeria (22%), and Tanzania (13%) show that Kenya is neither a best nor a worst case. Facebook penetration in Latin America is much higher and Facebook usage in Mexico (87%) is also similar to other countries in its region, such as Brazil (72%) and Argentina (83%).[5] Second, both countries have average levels of mobile phone use for their respective regions. Countries on the African continent vary widely in the number of cellular subscriptions per capita, with a maximum of 1.84, minimum of 0.27, and average of 0.88 in 2018. Kenya falls in the center of this range, with 0.96 cellular subscriptions per capita in 2018 (World Bank, 2020). Likewise, countries in Latin America had an average of 1.1 cellular subscription per capita in 2018. In Mexico, there are 0.95 cellular subscriptions per capita (World Bank, 2020). Finally, adult literacy is about 82% and 95% in Kenya and Mexico, respectively, meaning that a majority of citizens would be able to read and potentially self administer an online survey (World Bank, 2020).

3.2 Quota Sampling

To recruit respondents in Kenya and Mexico, we use a quota sampling design inspired by well-respected, in-person representative surveys in these countries (Latin American Public Opinion Project and Afrobarometer). Our geographic targeting approach was modeled on the geographic stratification approaches used by these surveys. Within each geographic stratum, we designed target cells based on the demographic characteristics used in our benchmark in-person surveys: gender and (in Mexico) age. We attempted to correct observed or expected imbalances by targeting additional respondents within underrepresented education (in Mexico and Kenya) and age (in Kenya) categories.

Facebook allows for two different types of geographic targeting. First, researchers can directly target audiences by providing an address or a latitude and longitude of interest, as well as a radius defining the catchment area. Second, researchers can use Facebook's predefined geographic entities,

[4] To address these concerns, we only paid survey incentives to unique phone numbers, unless respondents provided a reason why the same number was used twice. We also changed the survey link frequently, so that link sharing was only possible in the short term.

[5] We calculated these figures based on the number of users per country that Facebook reports (Araujo et al., 2017) and the population figures for individuals age 15 years and older in each country available from the UN (United Nations, 2019; Ševčíková, 2020).

which typically consist of neighborhoods, cities, or national administrative units (e.g. states). In Mexico, we targeted respondents according to these predefined geographic entities. In Kenya, we targeted primarily based on latitude and longitude, although we supplemented this sample with respondents from older and low-education groups recruited at the province level. We are therefore able to examine the viability of both approaches to geographic targeting. Below, we summarize our sampling strategies; Section 2 of the Supplementary Information includes further details about the approach taken in both countries, along with an extended discussion of the constraints and opportunities embedded within Facebook's advertising platform.

Table 1: Comparison of the data collected in Mexico and Kenya

	Mexico	**Kenya**
Demographic targeting	Gender, age, (education)*	Gender, (age, education)*
Geographic targeting	Administrative unit	Grouped Afrobarometer clusters
Field dates	August 17-18, 2019	September 21-29, 2019
Incentives	No	Yes ($\sim$ \$0.50)
Comparison	2015 Census 2019 Americas Barometer	2019 Census 2016 AfroBarometer
Question types	Demographics, political party affiliation, climate change	Demographics, mobility, political attitudes, social media use, fertility, household assets, behavioral experiment
N. questions	23	60
N. quota sampling cells	128	66
N. respondents	5,168	1,452

** The parentheses denote additional strata added to correct for observed/expected sample imbalances.*

3.2.1 Mexico: sampling by administrative unit

To draw a nationally representative sample of Mexican residents, we targeted Facebook users by age, gender, and geographic location. In order to gauge baseline interest in taking the survey without compensation, we did not compensate respondents for their participation in the short survey. Our geographic sampling protocol mirrors the procedure used by Latin American Public Opinion Project (2017) (LAPOP) to sample within small, medium, and large municipalities.[6] Within these geographic strata, we further stratified respondents based on gender and age categories. After this initial data collection, we found that our sample underrepresented Mexicans whose highest level of education completed was high school or less. To correct this imbalance, we conducted a second round of data collection in which we targeted respondents in each of our previously constructed geographic-demographic strata who had no more than a high school education. After collecting this low-education sample and dropping individuals with self-reported ages under 18 (n=165); those who did not report their age, gender, geographic location, or education level; and one individual without a recorded stratum,[7] our sample contained 5,168 individuals.

[6] LAPOP collects a sample that is stratified by four geographic regions, size of municipality (100,000+ inhabitants; 25,000-100,000 inhabitants, and less than 25,000 inhabitants), and urban and rural areas within municipalities.

[7] This likely occurred because the user inadvertently modified the Facebook Ad url which contained quota-related data.

3.2.2 Kenya: sampling by geolocation

In Kenya, we targeted ads according to gender (male, female) and geography. Respondents were compensated with 50 Kenyan Shillings' ($\sim$ \$0.50) worth of airtime sent to their phones.[8] The Facebook ad clearly stated that this was the incentive for participation. We used a geographic quota sampling approach to mimic the Afrobarometer sampling strategy. We began by obtaining a list of 227 site locations from the 2016 Afrobarometer survey and identified those with more than 1,000 monthly active Facebook users for both genders. We clustered these site locations into 25 groups and targeted audiences within 12 miles of the centroid of each cluster. Within these clusters, we stratified respondents by gender. Anticipating that we would have a hard time reaching less educated and older respondents, we also created two ads in each of the eight provinces (excluding the capital of each province)[9] of Kenya to target users 32 years and older, and users with an "unspecified" education level.[10] After removing respondents who completed the survey in less than five minutes, respondents who did not complete the survey, duplicate entries from respondents who took the survey more than once, and those who did not report an age, we were left with a sample of 1,452 respondents.

3.3 Survey Instruments

To direct respondents to the surveys, we created Facebook pages representing our survey campaigns, and placed ads from these pages to target people within the sampling strata described. After clicking on the Facebook ad, respondents were sent to a survey hosted on Qualtrics. For Kenyan respondents, the first survey question asked respondents to choose from one of five possible languages (English, Kiswahili, Kikuyu, Luo, and Somali) in which to take the survey. For Mexican respondents the survey was administered in Spanish. Upon completing the survey, respondents were directed to a thank-you page with an embedded Facebook "Pixel" which allowed Facebook to identify the users who clicked on the ad and completed the survey.

We designed both the Mexico and the Kenya surveys to collect information on demographics and attitudes that we could compare to each country's census and to other nationally representative surveys fielded around the same time. Full copies of our surveys are presented in the Supplementary Information (Section 4). The survey used in Mexico contained 14 substantive questions and 9 demographic questions. It replicated certain questions from the Mexican Census and the LAPOP survey, fielded in early 2019 (LAPOP, 2019). The attitude questions were largely focused on measuring beliefs about climate change. Climate change is a global problem, which will require coordination across countries to solve, including popular political support in many countries. The issue of climate change thus offers an important empirical setting in which to explore the potential for Facebook to generate inexpensive but accurate estimates of cross-national opinions, including in countries across the Global South. In order to avoid biasing the sample, our Facebook advertisements did not mention climate change, but instead invited respondents to provide their opinion on an unspecified issue.

[8] "Airtime" is mobile credit that can be used for calling or data.

[9] Kenya no longer uses provinces as the country's primary geographic unit, which is now the county. However, provinces were the administrative units available on the Facebook targeting interface.

[10] Given that the education levels specified on Facebook mimic the US system, the available targeting levels do not correspond to those in Kenya. We guessed that those who did not finish primary school (or who otherwise had little schooling) might have left their education blank or Facebook would be unable to impute their education level, and therefore would be assigned to this category. Our understanding from conversations with a Facebook marketing advisor is that Facebook estimates education using a range of data sources including a user's location, websites visited, pages liked, and information posted on the user's profile page.

The survey fielded in Kenya contained 60 questions which collected information on demographic characteristics, mobility and geographic location, political attitudes and engagement, news consumption and social media use, household assets and living conditions, and other topics. Several of the questions about demographics, political attitudes, media use, and household assets were drawn from the 2016 (Round 7) Afrobarometer survey. In addition, we included a set of questions to replicate a canonical risk-framing experiment that demonstrates the human tendency towards loss aversion (Tversky and Kahneman, 1981). We fielded our survey on Facebook immediately following the 2019 census and concurrent with the 2019 Afrobarometer in order to ensure that observed differences between our survey and nationally representative samples would not result from differences in timing.[11]

3.4 Poststratification adjustments

For both samples, we then used iterative proportional fitting, or raking, to create weights for all respondents according to the distribution of the national populations across gender, education, age cohort, and geography.[12] We created the weights using the rake function in the **survey** package in R (Lumley, 2020), which iterates to proportionally fit weights based on the marginal distributions of demographic variables of interest. In Kenya we collated the marginal distributions of Kenyans age 18 years and older in the population using the 2019 census data for the following categories: age (18-29, 30-49, 50-59, 60+), gender (male/female), education (primary or less, secondary, technical training, university or above), and geography (urban/rural) (KNBS, 2010). In Mexico we used the same age and gender categories as in Kenya, education (none, secondary or less, technical training, university or above), and geography (size of municipality within the four regions of the country) (INEGI, 2015a). In both cases, we trimmed weights by setting the minimum weight to the 5th percentile and the maximum weight to the 95th percentile.[13]

4 Internal Validity

Below, we assess three dimensions of internal validity as they apply to Facebook survey data from Kenya and Mexico. First, we ask whether it was feasible to target our respondents of interest and to recruit enough survey participants to fill each sampling cell. Second, we evaluate whether the respondents recruited into each cell reported demographic characteristics and geographic locations that matched the demographic and geographic characteristics targeted by the Facebook ad they received. Finally, conditional on having been recruited, we check whether respondents were attentive to the survey and honest in their responses.

4.1 Does Facebook allow us to target the respondents of interest?

A first-order internal validity question is whether Facebook allows us to target and fill the strata of interest. As a first check, we examine our success in targeting individuals in each geographic area of interest. In Mexico, we find a good distribution of responses. Using only self-reported residential

[11] Note: the 2019 Afrobarometer data are not yet available, so we currently compare our sample to the 2016 Afrobarometer.

[12] A comparison of weighting procedures suggests that choosing appropriate weighting variables is more important than using a more complex statistical procedure, and that raking works just as well as more complex methods to reduce bias (Mercer et al., 2018).

[13] In Mexico, 512 individuals (9.9% of the sample) had untrimmed weights falling outside this range (0.1-4.6) and were adjusted. In Kenya 67 observations (5% of the sample) had original weights outside of this range (0.7-2.5) and were adjusted.

data, in each of our 128 target cells we collected responses from between 17 and 77 individuals, with a median of 41 individuals.[14]

In Kenya, we were less successful in filling our 66 strata. The quota targets for each stratum were set according to the Afrobarometer weights associated with each stratum's geolocation, which are based on population.[15] Therefore, the number of respondents targeted per stratum ranged from 4 - 167. Ultimately, we filled 17 of our 66 target strata; the strata that fell short were missing a median of three respondents.[16] Two main factors contributed to the failure to fill some strata. First, we manually closed several of our survey strata because the advertising cost per completed survey was too high ($5 or more per respondent). Second, because of concern about viral sharing and completions of surveys that were not recorded by Facebook, we typically ended survey ads slightly before the corresponding stratum was filled.

4.2 Does Facebook correctly identify respondents?

We next assess whether Facebook appropriately directed us to individuals who matched the criteria in each particular quota cell. Since we recorded the ad on which each respondent clicked in order to enter our survey, we can compare self-reported demographics with those used by Facebook's platform. We present these comparisons according to respondents' age, gender, highest level of education completed, and region of residence.

In the Mexico sample, 87% of respondents reported ages consistent with their Facebook advertisement strata.[17] There were no systematic patterns in which age categories were prone to mismatches. When we targeted ads to Kenyans 32 years and older, on the other hand, only 47% of respondents who reached our survey from these ads were indeed 32 years old or older. The ages of these respondents ranged from 19 to 48 years old, with a mean of 31 years.

Facebook's gender data were more accurate in both countries. In Mexico, Facebook assigned a gender that matched respondents' self-reported gender for 99% of the respondents. In Kenya, Facebook ads performed almost as well, with 90% of respondents reporting a gender identity in the survey that matched the ad that targeted them. The 10% of respondents who were recruited from an ad that was targeted toward the opposite gender might have resulted from respondents sharing ads with friends so that they could also benefit from taking the survey and receiving 50 KES in airtime. It would not be surprising that greater sharing would have occurred in the context of the incentivized survey in Kenya, compared to the non-incentivized survey in Mexico.

Geographic targeting was less accurate in both countries. In Mexico, 67% of respondents reported living in a municipality that matched their Facebook advertising target cell. Most of these errors were a function of mismatches within (rather than between) Mexican regions: 92% of respondents had matching self-reported and Facebook-advertised regions. Most Kenyan respondents were targeted geographically via clusters of Afrobarometer coordinates, which could fall across multiple provinces. Therefore, we first checked whether the province corresponding to the respondent's self-reported location matched the province of at least one of the corresponding Afrobarometer coordinates. Using this definition, we achieved a 64% match rate between the Facebook target province and the Afrobarometer province. The remainder of respondents in hard-to-reach age or education groups (n=67) were targeted at the province level. In these strata, 69% of respondent-reported

[14] 1st quartile: 33, 3rd quartile: 46.

[15] For the province-level oversampling of older and less-educated respondents, we set a quota of 5 respondents per stratum.

[16] Min: 1, max: 52.

[17] We asked age through an open-ended question. Some respondents (n=52) did not provide their age, and these individuals are not included in this calculation.

Table 2: Accuracy of Facebook targeting, as defined by the percent match between Facebook- and self-reported data

Characteristic	**Mexico**	**Kenya**
Age	87	47*
Gender	99	90
Location	67	64
Education level	30	13‡

** This reflects the proportion of respondents who were at least 32 years old, given that they responded to an ad targeting this age group specifically.*

‡ *This reflects the proportion of respondents who reported some secondary school or less, given that they responded to an ad targeting respondents with an "unspecified" level of education.*

towns of residence fell within the targeted province.

Education targeting was noisy in both countries. In our second round of data collection in Mexico, we targeted respondents with a high school degree and lower levels of education. Only 30% of the respondents who reached our survey through these ads self-reported education levels that match Facebook's categorization. In other words, Facebook miscategorized 70% of the respondents in this sample. However, 43% of the individuals in the low-education targeted group reported their highest level of education as "Bachillerato/ Profesional Técnico/ Media Superior." If we consider that Facebook may qualify these technical school degrees — which overlap with the type of schooling that U.S. residents might call "high school" — then Facebook correctly categorized 73% of the respondents in this group and, conversely, only miscategorized 27% of them. The ambiguity in Facebook's category definitions for this group illustrates one challenge of using a cross-national platform with universal targeting categories.

As described above, in Kenya we targeted Facebook users who did not specify their educational attainment in an attempt to recruit less educated Kenyans. Among the population of Facebook users 18 years and older in Kenya, Facebook reports an "unspecified" education level for 40% of people. The 31 respondents recruited by these ads did have slightly lower levels of education, on average, than the rest of our sample.[18] However, this was not an exact targeting strategy: only 13% of the sample recruited from ads targeting those with an unspecified education level self-reported having some secondary school or less, compared to 3% of respondents recruited from all other ads.

4.3 Are respondents paying attention to survey questions?

We assess a second dimension of internal validity by examining the quality of responses provided by individuals recruited through Facebook. We first assess response times, since rushing through a survey is a standard indicator of low-quality responses. In Mexico, respondents took between 2 minutes and 10 hours to complete the survey, with a median of 6 minutes. The survey consisted of 23 questions, and 5-8 minutes is a reasonable amount of time to complete a survey of this length. In Kenya, respondents took between 4 minutes and 140 hours to complete the survey, with a median time of 24 minutes. The survey consisted of 60 questions covering a broader range of topics, so it is unsurprising that average completion times were longer. Again, 20-30 minutes is a reasonable

[18] For example, 10% reported that primary school, informal school, or no school was the highest level of education they had attained, compared with 2% in the rest of our sample. Only 23% of this subsample reported completing university or post-graduate education, relative to 34% in the rest of our sample.

amount of time in which to complete a survey of this length.[19]

Another concern with online surveys is that even if respondents take time to complete the survey, they simply click through without reading and answering questions honestly. This is particularly worrisome when respondents are paid for their participation and time, as was the case in Kenya. In Kenya we included two questions in the survey to check whether respondents were paying attention.[20] The questions varied in their difficulty, and pass rates varied accordingly: 98% of respondents answered the "easy" attention check question correctly, whereas only 54% of respondents passed the more taxing attention check. Based on these findings, we recommend that researchers include multiple attention checks of varying difficulty (Berinsky et al., 2014, 2019), to ensure that the questions are testing attention and honesty rather than respondent comprehension or sophistication.

5 External Validity

To assess the external validity of our results, we compare the distribution of demographic characteristics in our Facebook samples with the national census, other nationally representative samples, and the population of Facebook users in both countries. We assess the extent to which weighting allows our samples to more closely approximate national populations. We also examine responses to attitudinal and behavioral survey questions, as well as to a canonical framing experiment.

5.1 Are Facebook samples representative?

We compare data from four main sources: the national census, nationally-representative survey samples (LAPOP and Afrobarometer), the Facebook population, and our Facebook surveys. The census serves as our benchmark for whether our samples accurately represent the national population. Comparisons with other representative survey samples allow us to evaluate the comparative advantages or disadvantages of the Facebook platform for cost-effectively collecting data with the goal of approximating a nationally representative sample. Comparing our Facebook sample with the Facebook population allows us to identify the selection of platform members into the survey. We can also compare the Facebook population to the census population in order to illuminate the limitations of the platform to recruit samples that resemble the national population. Figure 2 shows these comparisons for our unweighted and weighted Facebook samples in Mexico and Kenya, respectively.

In both countries, the unweighted Facebook samples differ most from the national population with respect to education and age. The Facebook populations and Facebook survey samples reflect higher levels of education than those reported in the census or nationally representative in-person surveys (LAPOP and Afrobarometer). Specifically, the Facebook samples over-represent respondents with more than secondary or high school education, and under-represent those who completed secondary/high school education or less.

In Kenya, the Facebook survey sample is also significantly younger and more urban than the population as a whole. In contrast, the Mexican survey sample contains a greater share of respondents over 50 and a smaller share of younger respondents. This highlights the effectiveness of our stratified sampling method: while the Facebook population in Mexico *over*represents 18-29

[19] By examining the operating system respondents used to complete the survey, we estimate that 96% of respondents in Kenya took the survey on their phone. Linux and Windows operating systems were coded as computers, while Android and iPhone were coded as mobile devices.

[20] A full description of both attention checks is included in Section 2.2.1 of the Supplementary Information.

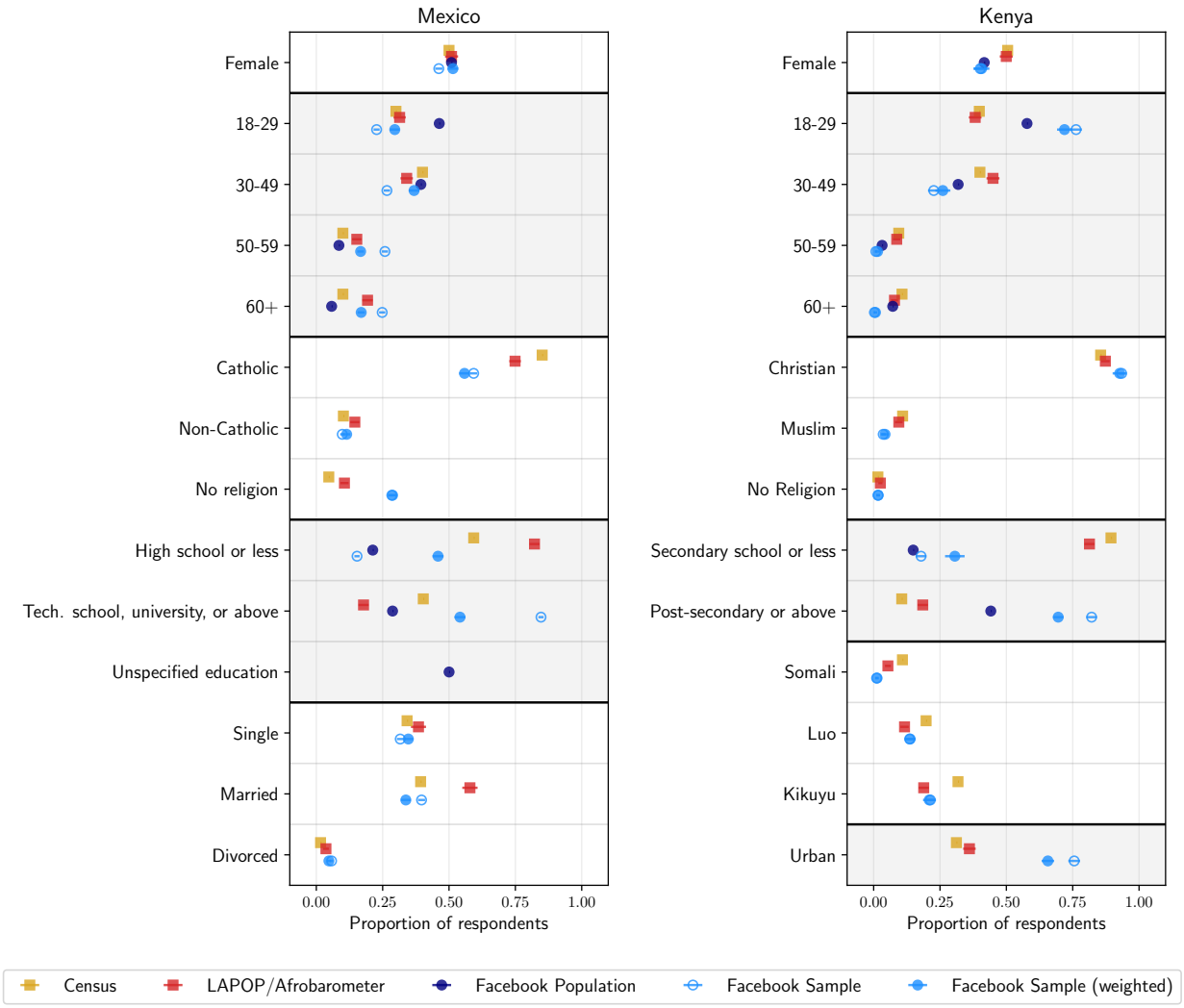

Figure 2: A comparison between the demographics reported in the national census, nationally representative survey samples (LAPOP/Afrobarometer), the Facebook population, and our Facebook sample for Kenya and Mexico. The Facebook sample is weighted using raking to match the national census on gender, education, region, and age. Mexican comparison data are from INEGI (INEGI, 2015b) and LAPOP. Kenyan comparison data are from the 2019 national census (KNBS, 2010) and the 2016 Afrobarometer (Afrobarometer, 2016).

year-olds compared with the national population, our quotas resulted in a sample that actually *under*represented this age group.

The application of weights helps to ameliorate some of these demographic imbalances. In Mexico, the weights correct the age imbalance described above. In Kenya, the age weights seem to perform somewhat worse and the weighting process does not result in better representation of respondents 50 years and older (of which there were only 16 in our sample). With respect to education, raking reduces some of the imbalance but does not completely eliminate it. Here again, a comparison with the Facebook population is illuminating. Facebook reports an "unspecified" educational status for half of the population of users in Mexico and 40% of users in Kenya. While targeted oversampling of low-education respondents helped reduce imbalance in our samples, the

surfeit of missing education information hindered our ability to fully correct it.

We also examine the representativeness of our samples on characteristics that are not incorporated into the weights. In Mexico, the unweighted and weighted samples closely mirror the population on measures of marital status. They under-represent Catholics and over-represent those with no religious affiliation. The Kenyan Facebook sample closely matches the proportion of people in the population who do not have a religion, but under-represents Muslims and over-represents Christians. Looking at some of the major tribes in Kenya, for which our survey was offered in these local languages, we obtain proportions of Luo, Kikuyu, and Somali respondents that are relatively similar to the Afrobarometer. Weighting does not substantially affect the distribution of these variables, which are not incorporated into our weights, in Mexico or Kenya. This suggests a negligible correlation between age, gender, education, and geography and the other variables we examine. Notably, while both samples suffer from some remaining imbalances, these demographic imbalances are often similar to or smaller in magnitude (though different in direction in some cases) to the imbalances in the high-quality samples collected by the LAPOP and Afrobarometer in-person surveys.

5.2 Can Facebook surveys recover benchmark public opinion estimates?

We next compare responses from our Facebook samples to opinion questions we replicated from surveys fielded around the same time as our Facebook surveys.

In Mexico, we replicated the question wording and response options for a question that LAPOP asked in its 2019 survey of Mexican respondents (LAPOP, 2019). In translation, the question reads:

> *Some people think that environmental protection should be given priority over economic growth, while others think that economic growth should be given priority over environmental protection. On a scale of one to seven, where one means that the environment should be the top priority and seven means that economic growth should be, where would you position yourself?*

The results from the LAPOP and Facebook samples are presented in Figure 3. Facebook respondents were substantially more likely to answer that environmental protection should be given the highest priority, and slightly less likely to choose responses at the economic growth end of the Likert scale. When we convert the Likert responses to a continuous variable, the average response from Facebook (2.83) falls over one point closer to the environmental protection end of the scale than the LAPOP sample (4.08).

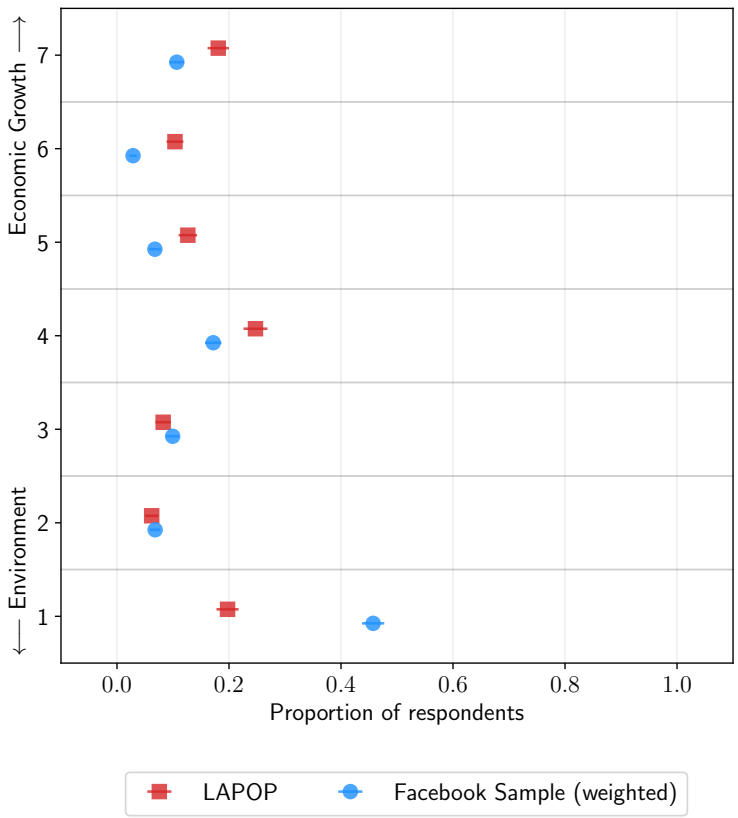

Figure 3: This figure shows responses from LAPOP and Facebook samples, to the question of whether environmental protection (1) or economic growth (7) should be given priority. Results are weighted for the Facebook sample, according to the procedure described in section 3.4, and unweighted for the LAPOP sample, consistent with the survey design documentation.

In Kenya, we similarly replicate questions from the Afrobarometer survey.[21] First, we compare perceptions of political freedom and opinions of political leaders between the Facebook and Afrobarometer samples. When asked *"In this country, how free are you to say what you think?"* 77% of Afrobarometer respondents report being "somewhat free" or "completely free" compared with 80% of Facebook respondents. Similarly, when asked *"Do you approve or disapprove of the way President Uhuru Kenyatta is handling his job as president of Kenya?"* respondents from the two samples share similar sentiments: 76% of Facebook respondents approve, compared with 77% of Afrobarometer respondents. Thus, political attitudes appear quite similar across the samples.

On the other hand, Facebook respondents are more likely to report having engaged in politics. When asked if they voted in the last election, 73% of Afrobarometer respondents said they did, compared to 81% of Facebook respondents. Interestingly, the proportion of Facebook voters is close to the government-reported turnout rate of 79% for the August 8, 2017 general election (Independent Electoral and Boundaries Commission, 2017). A similar difference is found for attending a community meeting; 64% of Afrobarometer respondents reported doing so in the past year, compared with 71% of Facebook respondents. The Facebook sample appears even more likely to take

[21] At the end of our survey we asked if respondents had been previously surveyed by any other organization. While 47% reported that they were surveyed by census enumerators, less than 1% reported being surveyed by the Afrobarometer. 16% said they were surveyed but did not recall by whom.

other types of political action (or at least to report doing so) than Afrobarometer respondents, including requesting government action, joining with others to raise an issue, contacting a government official, and participating in a protest. Facebook survey respondents in Kenya appear to be somewhat more politically engaged than the Afrobarometer sample, but one unexpected finding is that they are *less* likely to feel close to a political party.

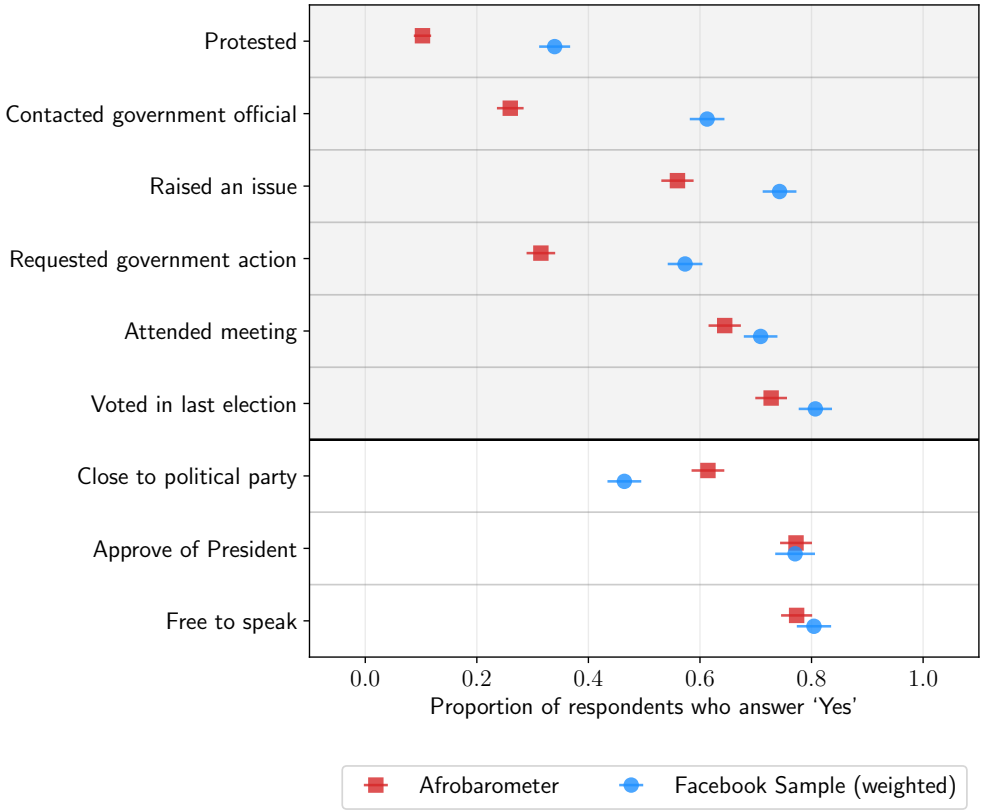

Figure 4: This figure shows responses from Afrobarometer and Facebook samples in Kenya, to questions related to political attitudes. It also compares self-reported behaviors including identifying with a political party, voting, and engaging in activities such as community meetings and protests. Estimates for both samples have been weighted using the individual weights provided in the Afrobarometer survey, and the raking procedure described above for the Facebook survey.

5.3 Can Facebook samples recover behavioral lab experiment results?

In Kenya, we also replicated a canonical behavioral experiment — Tversky and Kahneman's (1981) "disease problem" — which has been conducted in a wide range of contexts over time. This question asks respondents:

> *Imagine that your country is preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows.*

Then respondents are presented with two policy options: a single certain option, and a risky option that assigns probabilities to two different outcomes. In expectation, the payoffs of both

Table 3: Replication of Tversky and Kahneman (1981) Disease Problem

	Kenya Facebook Sample (unweighted)		**Tversky and Kahneman**		**Berinsky, Huber, and Lenz**	
Options	Save	Die	Save	Die	Save	Die
Certain	62%	36%	72%	22%	74%	38%
Risky	38%	64%	28%	78%	36%	62%

policies are equal, so differences in respondents' choices should be driven primarily by their appetites for risk. Respondents are randomly assigned to one of two conditions: either the policy options are framed in terms of a loss (number of deaths) or the policy options are framed as a gain (number of lives saved). The canonical finding that has been replicated in many samples across the globe suggests that people are loss averse — favoring the certain option when it is framed in terms of lives saved, and the risky option when it is framed in terms of lives lost (Tversky and Kahneman, 1981).

As shown in Table 3, results from our Facebook sample in Kenya are similar to the original sample and to other Western convenience samples (Berinsky et al., 2012). When the policies are framed in terms of the number of lives saved, a majority (62%) of respondents prefers the certain policy. When the policies are framed in terms of the number of people who will die from the disease, a similar majority (64%) of respondents prefers the risky option. As in existing studies, the difference in the proportion of respondents who chose the certain policy when it is framed in terms of lives saved vs. lives lost is significant with a p-value $<.01$.

6 Survey costs

Researchers typically weigh internal and external validity concerns against the relative costs of different survey tools. For Facebook surveys, these typically include the advertising costs for using the Facebook platform and optional incentive costs associated with paying respondents. We check whether surveys are cost-effective according to different definitions of cost-effectiveness.

Table 4 shows the reach and total spending of our survey advertising campaigns. The mean cost per completed survey was \$0.16 in Mexico[22] and \$1.07 in Kenya.[23] This is quite inexpensive in Mexico, but even in Kenya the cost is comparable with the cost of recruiting online convenience samples using platforms such as MTurk or Lucid.[24] In particular, these costs are substantially cheaper than the only feasible alternatives: online panels or in-person field surveys. In Supplementary Information Section 3, we explore variation in the cost of reaching different demographic groups. These costs vary because Facebook ads are deployed using a bidding system in which hard-to-reach populations are more expensive for advertisers to target.

[22]These statistics reflect the full campaign, including our low-education oversample. In Section 3 of the Supplementary Information, we include separate data for the initial sample (without education targeting) and for the oversample in which we targeted respondents with a high school degree or lower levels of education.

[23]This figure is derived from Facebook's count of completed surveys. This falls to \$0.56 if we consider all 2,323 surveys *initiated* on Qualtrics, and \$0.89 if we consider all 1,452 surveys completed on Qualtrics which were determined to be valid and were used in the analysis.

[24]We make this comparison only for context, since these platforms do not enable researchers to contact survey respondents in Kenya.

Table 4: Reach of survey

	Mexico	**Kenya**
Reach	439,056	318,960
Impressions	492,564	649,264
Clicks	24,314	8,206
Survey results	5,313	1,211
Total spent	$847.76	$1,293.91
Click through rate (%)	0.049	0.013
Completion rate (%)	0.011	0.002
Cost per click ($)	$0.04	$0.16
Cost per completed survey ($)	$0.16	$1.07

Notes: These survey results are those reported by Facebook (using the pixel) and therefore do not match our completions with respect to Qualtrics. The click through rate and the completion rate are defined with respect to the number of impressions.

7 Conclusion

In this paper, we assess the value of the Facebook advertising platform as a survey recruitment tool in the Global South. We provide a framework for quality evaluation that is based on internal validity, external validity, and cost. Using our experiences running quota-sampled surveys in Mexico and Kenya, we shed light on the tradeoffs and considerations involved in using Facebook for this purpose.

Given its broad user base, Facebook enables researchers to quickly and reliably recruit subjects from countries that are underrepresented on existing online subject recruitment platforms. We show that we can successfully and cheaply recruit hundreds of respondents in our countries of interest, with a broad range of demographic characteristics, geographic locations, and opinions. Facebook also presents a unique opportunity to rapidly gather real-time attitudes and sentiments in the face of current events. In our experience, hundreds of survey responses can be gathered within a few days, even when targeting specific demographic strata or geographic regions. Facebook is also cost-effective; Facebook offers a much cheaper way to reach respondents than online panels or face-to-face surveys.

On the other hand, the Facebook advertising platform has a number of limitations. Internal validity may suffer because Facebook is unable to find respondents in particular strata. We could not fill a number of cells in Kenya, and education poses a particular problem because Facebook does not maintain educational status for many users in both countries. Internal validity may also suffer because Facebook does not correctly assign people to strata, and our success in this area is mixed. Facebook accurately assigned the vast majority of respondents to gender groups in both countries and to age cohorts in Mexico. Facebook mischaracterized about half of Kenyans' ages, and Facebook-assigned geographic locations matched self-reported locations in about two-thirds of cases in both countries. Internal validity may also suffer because responses are not i.i.d. samples from within their respective strata (for example, if people share the survey link with others in their network, or if the same person takes the survey multiple times).[25] Finally, as with all survey research, internal validity may suffer due to respondent inattention or haste. Encouragingly, we find that survey completion times are generally reasonable in both countries and that 98% of

[25] While you can "prevent ballot box stuffing" in Qualtrics which will block the link from being accessible again on the same device/browser, it is easily overcome using an incognito window, different browser, or different device, and therefore we are unable to prevent savvy duplicate survey takers.

respondents in Kenya passed an "easy" attention check, suggesting that truthfulness and focus are not critical problems in our sample.

From an external validity standpoint, we find that Facebook survey samples are biased with respect to both the population of Facebook users and the national population as a whole. Some biases are consistent across countries, such as the propensity to recruit more highly educated individuals, or the propensity for Facebook users to be more politically aware (either through expressing higher levels of concern about climate change in Mexico, or in reporting higher levels of political engagement in Kenya). Other biases vary between countries. The Mexican sample was skewed towards older age cohorts, whereas the Kenyan sample skewed towards younger age cohorts. Raked weights ameliorated but did not completely remove these biases. One encouraging finding is that we successfully replicated the results from Tversky and Kahneman's risk-framing experiment, suggesting that Facebook may have potential as a recruitment tool to broaden the audiences used for behavioral experiments.

With respect to costs, we find that these can be quite low, but they are highly dependent on the targeting strategy used. For some Kenyan cohorts, we paid up to $22.07 in advertising costs per completed survey. Facebook may be most cost-effective in urban areas or among otherwise internet-connected populations; these populations can be identified ex ante through the use of Facebook's estimates of Daily and Monthly Active Users (DAU/MAU). Naturally, the use of incentives for survey completion influences the cost-quality tradeoff. In Mexico, we successfully recruited respondents without incentives, likely because the survey was quite short and internet penetration is substantially higher than in Kenya. In Kenya, because much of the population uses rate-limited mobile internet, we provided a modest airtime credit as compensation. This incentive likely encouraged greater participation among resource-constrained respondents, but it also led to instances of gaming and viral sharing of the survey link. While link sharing complicates our efforts to satisfy the assumptions laid out in Section 2.3, future research could investigate strategies for leveraging the social nature of the platform, for example by engaging in snowball sampling that promotes forwarding of the survey link.

In addition to the tradeoffs discussed above, scholars should consider context before deciding to use Facebook to recruit subjects. Using Facebook as a method of respondent recruitment will be particularly successful in contexts where 1) phone and internet penetration is widespread, 2) literacy rates are relatively high, and 3) recent census data are available.[26] Researchers looking to adopt this method should also keep in mind that the nature of users' interactions with and expectations for Facebook might influence internal validity. While we have tested this method in two competitive democracies, in authoritarian states (where Facebook is permitted), citizens who answer surveys on the platform might have different assumptions about the government's surveillance of their responses. Researchers should consider these concerns in the survey design process, and future research could extend our analysis to other countries and contexts.

Additionally, algorithmic bias and ethics pose fundamental concerns for using the Facebook platform for research. First, there is very little understanding of how Facebook's ad targeting algorithm works and the extent to which this introduces selection on unobservables into survey samples. Future work could compare different ad targeting strategies to elucidate the the influence of Facebook's optimization algorithm. Second, there is a need for a clearer ethical framework around running remote surveys, including better guidance on the types of approvals needed from national ethics review boards. Institutional review boards will likely lag behind changing research environments. We recommend that researchers use their best judgement when designing remote

[26]The ability to reweight Facebook responses is predicated on access to reliable census data, which is not available in all countries.

surveys on social media, as they would if they were recruiting and surveying respondents face-to-face. Knowledge of the local context will continue to be invaluable in designing surveys in comparative contexts.[27]

All survey efforts involve tradeoffs between validity and costs. In this paper, we articulate a framework for evaluating these tradeoffs and assessing them in the context of Facebook surveys in Kenya and Mexico. With the global growth in mobile internet penetration and with the social distancing requirements necessitated by the current pandemic, it seems inevitable that the use of online survey samples will continue to increase over time and the questions we have raised here will become increasingly important to consider.

Acknowledgements

The authors thank Michaël Aklin, Baobao Zhang, Blair Read, Peter Howe, Jennifer Marlon, Anthony Leiserowitz, and André Grow for helpful comments. We thank Warsama Abdifitah, Kibuchi Eliud, Ahmed Hared, Lilian Ligeyo, Laban Okune, Gustavo Ovando-Montejo, Nelson Ngige, and Eunice Williams for translation assistance.

Funding

This work was supported by the Summer Institute for Computational Social Science (SICSS), the Russell Sage Foundation, the Alfred P. Sloan Foundation, and the Yale Program on Climate Change Communication.

[27] We do not have space to discuss the ethical questions related to researchers paying Facebook to recruit respondents. As with many aspects of research, this question is largely outside the realm of IRB guidelines and is instead left to the researchers to make their own decisions.

References

Afrobarometer (2016). Afrobarometer, 2016.

Araujo, M., Y. Mejova, I. Weber, and F. Benevenuto (2017). Using Facebook Ads Audiences for Global Lifestyle Disease Surveillance: Promises and Limitations. In *Proceedings of the 2017 ACM on Web Science Conference*, WebSci '17, Troy, New York, USA, pp. 253–257. ACM.

Berinsky, A. J., G. A. Huber, and G. S. Lenz (2012). Evaluating online labor markets for experimental research: Amazon. com's mechanical turk. *Political analysis 20*(3), 351–368.

Berinsky, A. J., M. F. Margolis, and M. W. Sances (2014). Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys. *American Journal of Political Science 58*(3), 739–753.

Berinsky, A. J., M. F. Margolis, M. W. Sances, and C. Warshaw (2019). Using screeners to measure respondent attention on self-administered surveys: Which items and how many? *Political Science Research and Methods*, 1–8.

Bicalho, C., M. Platas, and L. R. Rosenzweig (2020, May). "if we move, it moves with us:" physical distancing in africa during covid-19.

Boas, T. C., D. P. Christenson, and D. M. Glick (2018). Recruiting large online samples in the united states and india: Facebook, Mechanical Turk, and Qualtrics. *Political Science Research and Methods*, 1–19.

Broockman, D. E. and D. P. Green (2014). Do online advertisements increase political candidates' name recognition or favorability? evidence from randomized field experiments. *Political Behavior 36*(2), 263–289.

Coppock, A. and O. A. McClellan (2019). Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. *Research & Politics 6*(1), 2053168018822174.

Facebook Inc. (2019). Facebook reports third quarter 2019 results. (Accessed on April 15, 2020).

Grow, A., D. Perrotta, E. Del Fava, J. Cimentada, F. Rampazzo, S. Gil-Clavel, and E. Zagheni (2020, May). Addressing Public Health Emergencies via Facebook Surveys: Advantages, Challenges, and Practical Considerations. Preprint, SocArXiv.

GSMA (2020). The Mobile Economy 2020. Technical report.

Hirano, S., G. S. Lenz, M. Pinkovskiy, and J. M. Snyder Jr (2015). Voter learning in state primary elections. *American Journal of Political Science 59*(1), 91–108.

Huff, C. and D. Tingley (2015). "who are these people?" evaluating the demographic characteristics and political preferences of mturk survey respondents. *Research & Politics 2*(3), 2053168015604648.

Independent Electoral and Boundaries Commission (2017, August). Speech by Independent Electoral and Boundaries Commission Chairperson, Wafula Chebukati at the Announcement of Presidential Results at Bomas of Kenya on August 11th 2017.

INEGI (2015a). Population. Accessed: 20 August, 2019.

INEGI (2015b). Porcentaje de la población de 15 años y más con algún grado escolar por entidad federativa según nivel de escolaridad y sexo, años censales seleccionados 2000 a 2015. Accessed: 20 August, 2019.

Jäger, K. (2017). The potential of online sampling for studying political activists around the world and across time. *Political Analysis 25*(3), 329–343.

Kapp, J. M., C. Peters, and D. P. Oliver (2013). Research recruitment using facebook advertising: big potential, big challenges. *Journal of Cancer Education 28*(1), 134–137.

KNBS (2010). 2019 population and housing census. Accessed: 9 June, 2020.

Kosinski, M., S. C. Matz, S. D. Gosling, V. Popov, and D. Stillwell (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist 70*(6), 543.

LAPOP (2018–2019). Mexico. Accessed: 16 October, 2019.

Latin American Public Opinion Project (2017). Americas barometer, 2016/17.

Litman, L., J. Robinson, and T. Abberbock (2017). Turkprime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior research methods 49*(2), 433–442.

Lumley, T. (2020). survey: analysis of complex survey samples. R package version 4.0.

Mercer, A., A. Lau, and C. Kennedy (2018). For weighting online opt-in samples, what matters most?

Pötzschke, S. and M. Braun (2017, October). Migrant Sampling Using Facebook Advertisements: A Case Study of Polish Migrants in Four European Countries. *Social Science Computer Review 35*(5), 633–653. 00011.

Ramo, D. E. and J. J. Prochaska (2012). Broad reach and targeted recruitment using facebook for an online survey of young adult substance use. *Journal of medical Internet research 14*(1), e28.

Ramo, D. E., T. M. Rodriguez, K. Chavez, M. J. Sommer, and J. J. Prochaska (2014). Facebook recruitment of young adult smokers for a cessation trial: methods, metrics, and lessons learned. *Internet Interventions 1*(2), 58–64.

Rosenzweig, L. R. and Y.-Y. Zhou (2019). Team and nation: Sports, nationalism, and attitudes toward refugees in kenya and tanzania. Accessed: 9 May, 2020.

Rotondi, V., R. Kashyap, L. M. Pesando, S. Spinelli, and F. C. Billari (2020). Leveraging mobile phones to attain sustainable development. *Proceedings of the National Academy of Sciences*.

Ryan, T. J. (2012). What makes us click? demonstrating incentives for angry discourse with digital-age field experiments. *The Journal of Politics 74*(4), 1138–1152.

Samuels, D. and C. Zucco Jr (2014). The power of partisanship in brazil: Evidence from survey experiments. *American Journal of Political Science 58*(1), 212–225.

Santoso, L. P., R. Stein, and R. Stevenson (2016). Survey experiments with google consumer surveys: Promise and pitfalls for academic research in social science. *Political Analysis 24*(3), 356–373.

Ševčíková, H. (2020). Explorer of World Population Prospects.

Tversky, A. and D. Kahneman (1981). The framing of decisions and the psychology of choice. *Science 211*(4481), 453–458.

United Nations (2019). *World Population Prospects Volume 1, Volume 1,*.

U.S. Securities and Exchange Commission (2019). Facebook Inc 2019 Annual Report 10-K.

World Bank (2020). World bank open data.

Zhang, B., M. Mildenberger, P. D. Howe, J. Marlon, S. A. Rosenthal, and A. Leiserowitz (2018). Quota sampling using facebook advertisements. *Political Science Research and Methods*, 1–7.