

Implications of Choice of Second Stage Selection Method on Sampling Error and Non-Sampling Error

Evidence from an IDP Camp in South Sudan

Kristen Himelein

Utz Pape

Michael Wild



WORLD BANK GROUP

Poverty and Equity Global Practice

January 2024

Abstract

The most common sampling approach for cross-sectional household surveys in the developing world is a stratified two-stage design, where the first stage is usually a sample from a census-based area frame, and the second stage is a random sample of households from each of the areas selected in the first stage. To overcome the problem of outdated census frame information, it is common to conduct a household listing operation within these areas. However, these listing operations come with severe implications for survey costs, timeframe, as well as quality. To avoid such second-stage operations, some surveys choose alternate approaches for their second-stage operation. This paper compares five of these approaches, namely, satellite mapping, segmentation, grid square, the north method, and random walk, through simulations based on a census conducted in a refugee camp in South Sudan. The paper compares the

simulated approach with the estimates derived from the actual experiment and finds that all the resulting estimates are biased. Nevertheless, in addition to their practical challenges, the satellite mapping, segmentation, and grid square approaches exhibit the smallest bias. Although random walk shows the worst performance in the simulations, it regains ground in its implementation, especially vis-à-vis the north method, where implementation adds most significantly to its bias. In conclusion, most probability-based methods perform better than non-probability methods like random walk and are therefore preferable when no traditional household listing can take place. Although it is important to consider the theoretical properties of sampling approaches, implementation is at least as important. Training, implementation modalities, and monitoring of compliance are key factors in the overall performance.

This paper is a product of the Poverty and Equity Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at upape@worldbank.org and mwild@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

**Implications of Choice of Second Stage Selection Method on Sampling Error and Non-Sampling Error:
Evidence from an IDP Camp in South Sudan**

Kristen Himelein, Utz Pape and Michael Wild¹

JEL classification: C83, C13, C81

Keywords: Survey and Sampling Methods, Estimation, Micro-economic Data, Displacement

¹ Authors in alphabetical order. Kristen Himelein was a senior economist / statistician in the Poverty and Equity Global Practice at the World Bank. Utz Pape is a lead economist in the Poverty and Equity Global Practice of the World Bank. Michael Wild is a senior statistician in the World Bank's Data Group, Analytics and Tools section. All views are those of the authors and do not reflect the views of their employers including the World Bank or its member countries. The authors would like to thank Mario Gronert, Luca Parisotto and Marina Tolchinsky for providing research assistance and Toan Do for comments on the project proposal and earlier drafts of this paper.

1. Introduction

The most common sampling approach for cross-sectional household surveys in the developing world is a stratified two-stage design (Grosh and Munoz, 1996). Following stratification based on administrative boundaries, clusters are selected in the first stage with probability proportional to size from a national census-based frame. In the second stage, a canvassing operation is conducted in the selected clusters to compile an updated list from which households are randomly selected. While this methodology is straight forward to implement in the field and reliably produces unbiased estimates, there are several downsides.

The first downside is cost. The World Bank's Living Standards Measurement Study team, which provides technical assistance on large-scale household surveys around the world, estimates the field listing operation increases the overall budget for data collection by 25 percent. Due to confidentiality concerns, the data collected during a field listing operation, typically the name of the household head and address or location description of dwellings, does not have any analytical applications beyond as a component of the weight calculations.² At a time when typical survey costs are in the USD millions, reducing a significant cost component will increase the financial sustainability of data collection.

The second drawback to the traditional design relates to timeliness. At a minimum, listing operations are usually conducted several days, if not several weeks, before the main fieldwork. As populations shift, the quality of the list degrades as time passes. While this is generally not a major concern for static populations living in villages or cities, it is a major concern for those in IDP (Internally Displaced People) and refugee camps. The transient nature of such environments implies building an accurate sampling frame is a complicated process often fraught with inaccuracies. Structures, often tents, for example, can easily be enlarged or split, quickly changing the layout of the camp, potentially invalidating a pre-existing sampling frame.

There are also issues related to the subjectivity in a listing operation. Eckman (2013) found only an 80 percent overlap between the same blocks listed separately by different interviewers in the United States. Undercoverage during the listing operation impacts the representativeness of the final estimates if the undercoverage is non-random. For example, O'Muircheartaigh et al. (2007) showed undercoverage in the United States is higher in low-income and rural areas. If this finding extends to the developing world, poverty numbers may be underestimated. In addition, Barrett et al. (2002) find higher undercoverage of households occupied by non-Hispanic black respondents compared with non-Hispanic white or other race respondents. This potential bias introduced by racial differences between the interviewer and respondent is of particular importance in the developing world context when interviewers are often recruited in the capital city and sent to more remote regions for the survey.

This paper builds on the work done by Himelein et al. (2017) in describing five alternative sampling approaches considered for a household survey in Mogadishu (satellite mapping, segmentation, grid squares, "Qibla method," and random walk). In that paper, however, the authors used simulations which assumed perfect implementation. Therefore, while it was possible to compare the sampling error of the

² This may however change in the near future, as more and more machine learning applications discover the usability of this data for mapping purposes.

five methods, it was not possible to consider non-sampling error. This paper goes a step further by using simulations to describe the sampling error and a field experiment in an IDP camp in South Sudan to measure the total survey error of each design compared to a census, allowing for the disaggregation of the total error into sampling and non-sampling components. In addition, we attempt to separate the components of non-sampling error linked to the sample method from those common across all methods, such as interviewers selecting larger households and other issues in properly implementing the household survey protocols.

The next section briefly describes each method and highlights the literature as it relates to the relevant selection methods. Section 3 describes the data set and protocols for each method included in the experiment, followed by Section 4, which discusses implementation issues. Section 5 reports the results of the analysis, and section 6 concludes with further discussion of the overall performance and areas for future research.

2. Description of Methods

This paper compares five alternatives of second stage selection (satellite mapping, segmentation, grid squares, “Qibla” (or “walk north”) method, and random walk) to a human canvassing operation. We consider the human canvassing to be the gold standard of listing methods, though acknowledge Eckman (2013) has identified the limitations mentioned in the previous section.

2.1. Satellite Listing

While using satellite data to construct a sampling frame is common in land and agricultural surveys, household surveys are a more limited, though growing, application of the technology. In satellite listing, structures are identified from satellite imagery using either manual demarcation or an automated “computer vision” algorithm. Structures are then selected using simple random sampling, teams provided with GPS coordinates and maps to locate selected households. The main benefit of satellite listing is, if properly implemented, the results would match the precision of the gold standard of manual canvassing. The drawbacks include potential difficulties in identifying selected households due to the margin of error in the GPS machines, if structure identification is done using outdated maps, or if an automated model is trained on a context that does not readily translate into the current application. In addition, imaging cannot always consistently distinguish between residential structures and non-residential structures, leading to overcoverage issues. If a non-residential structure is identified, the selected point is declared out-of-scope and a replacement is used. If there are a large number of these points, however, assumptions would be required to adjust the denominator of the probability of selection calculations or the resulting weights could be biased. Large numbers of out-of-scope structures were not anticipated to be an issue in South Sudan though since the camps were predominantly residential. Another potential issue with automated algorithms is undercoverage resulting from a failure by the algorithm to identify a structure, such as if the roof is constructed of an organic material not sufficiently distinct from the ground cover. While this problem can be mitigated by improved imagery and well-designed ground truthing surveys, inaccuracies may still remain depending on the context in which the model was trained.

Examples from the literature include a study measuring disparities in health in Bobo-Dioulasso, Burkina Faso, in which Kassié et al. (2017) used satellite images and the cadastral map of the town for random sampling through a supervised classification method. Escamilla et al. (2014) used Google earth imagery and GIS software to manually digitize structures for a sampling frame for household survey in Lilongwe, Malawi. A random sample was then drawn from the list of households and interviewers used hand-held GPS devices to locate and interview households. A similar approach was used by Wampler, Rediske, & Molla (2013) for an ethnographic and water quality survey in Haiti. Specifically related to conflict, Lin & Kuwayama (2016) used high-resolution satellite imagery and manual identification to develop a sampling frame of man-made structures for their health survey in the Kerenik Camp in Darfur. Structures were then manually selected, and interviewers used hand-held GPS devices to navigate to the selected locations to conduct interviews.

The probability of selection for this method is simply the sampling fraction $\frac{j}{N_j}$ where j is the number of selected structures and N_j is the total number of structures. In the cases where it was necessary to select from multiple households within the dwelling, there would be an additional probability of selection for the household $\frac{i}{N_{ji}}$ where i is the number of households selected and N_{ji} is the total number of households in structure j . Therefore, the weight for this method can be represented as $w'_i = \frac{N_j N_{ji}}{ji}$. In the case of the experiment, the form simplifies to $w'_i = \frac{N_j N_{ji}}{j}$ as only one household was selected per structure.

2.2. Segmentation

Segmentation is a well-established practice of addressing primary sampling units (PSU) that are too large to list and can be done either prior to or after selection. Dividing large PSUs prior to selection is more efficient statistically because it keeps the selection to two-stages, but costlier, particularly if there are substantial numbers of large PSUs in the frame (Kish, 1965, p. 156). In addition, this approach only works if a reasonably updated frame exists. Field segmentation is the more common approach, which allows the larger PSUs to be selected and then does the segmentation as part of the fieldwork. This approach is less costly as it only requires segmenting the selected clusters and can be used if unexpectedly large clusters are found in the field but does reduce statistical precision due to the additional level of selection. Regardless of if segmentation is done pre- or post-selection, the segments should be approximately equal sized, and boundaries should follow identifiable landmarks on the ground to facilitate accurate implementation by field teams.

Assuming field segmentation, the weight for this method is based on the probability of selection at each of the stages. Probability p_1 is the probability of selection of a PSU from the total number of PSUs N_k , or $\frac{k}{N_k}$, where k is the number of PSUs selected. Probability p_2 is the probability of selection of a segment from the total number of segments within the selected PSU N_{kl} , or $\frac{l}{N_{kl}}$, where l is the number of segments selected. Probability p_3 is the probability of selection of a structure from the total number of structures within the selected segment N_{klj} , or $\frac{j}{N_{klj}}$, where j is the number of structures selected. As above, there

would be an additional layer of selection for households if there are multiple households within a structure, the probability for which can be represented as $\frac{i}{N_{klji}}$. The weight would therefore be $w'_i = \frac{(N_k)(N_{ki})(N_{klj})(N_{klji})}{klji}$.

2.3. Grid Squares

The grid squares approach breaks selected areas down into smaller units for listing, but instead of manually drawing boundaries or using an algorithm, a uniform grid is imposed on the area. This approach can be used either for PSU, in which case it would be similar to segmentation, or to the area as a whole, in which case each grid square would act like a PSU. The benefit is a decrease in pre-survey preparation time, but at the cost of greater difficulties in implementation if grid lines do not follow landmarks, as well as greater difficulties in calculating weights for households overlapping grid squares (Himelein et al., 2017). Elangovan et al. (2016) used a grid sampling methodology in the study of the health impacts of hard stone crushers in a residential neighborhood of Chennai. The authors found that 65 of the 300 selected grid squares were empty land, despite having excluded forests, bodies of water, etc. ex ante. In a mortality study in Iraq, Galway et al. (2012) used GIS and Google earth imagery for household sampling. The method used gridded population data for selection of clusters. The first cluster sampling stage of their study used the ‘Create Spatially Balanced Points’ (CSBP) function in the ArcGIS (v10) software. Boo et al. (2020) introduces a sampling design based on gridded population estimates as their sampling frame to implement a PPS design and derive sample size estimates for the number of grid cells.

Assuming the grid square method is applied to the area itself rather than a selected PSU, the weights for the grid method are similar to those for segmentation, where the cells are the PSUs, but without the additional step of selecting segments. The weights can therefore be represented as $w'_i = \frac{(N_k)(N_{kj})(N_{kji})}{kji}$.

2.4. North Method

The “Qibla method” described in Himelein et al. (2017), or what is called in this paper the “North method” method, is an attempt to assign probability weights to random point selection methods. Several random point selection methods can be found in the literature, particularly in relation to epidemiological studies. Grais et al (2007) used a methodology in which the closest household to a randomly selected point is selected for a study of vaccination rates in urban Niger, though did not attempt to calculate probabilistic sampling weights. Similar approaches were used by Kondo et al (2014) in a study of the city of Sanitiago Atitlán, Kumar (2007) in urban India, and Kolbe and Hutson (2006) in Port-au-Prince, Haiti. Shannon et al (2012) also used such a method to select points in a study of violence in Southern Lebanon in 2008 but used the radius of a circle to define an area to be field listed, and from which buildings and then households were selected for enumeration. The circle area and building density were used to calculate probability weights. The main difference between most random point selection methods and the North Method described here is that the North Method attempts to accurately estimate the probabilities of selection.

To accurately calculate weights for the North Method, the area of possible random selection points (RSPs) leading to the selection of a structure must be measured or calculated. A structure is chosen if the RSP falls within it or if walking north from the RSP the structure is encountered. The area of all points from which one and the same structure is selected, its selection area, is made up of the structure and the *shadow* it casts to its south without interference of any other structure. Shadow here refers to the union of all points south of the structure that by protocol should lead to its selection (Figure 6).

Since a structure with a larger selection area is more likely to have a random RSP landing within it than one with a smaller area, weights are required for unbiased estimates. As all starting points fall within the camp area and the camp area itself is made up of the selection area of the structures, the weight given to an observation is proportional to the ratio of the selection area to the total area of the camp. The weights for the North Method therefore require the calculation of the area of valid RSPs that lead the enumerator to select the structure to determine its selection probability.

Let the selection area be labeled as A_i , then the weight is $w'_i = \frac{N_{ji}}{(1 - (1 - A_j/K))^n}$, where K represents the sum over structures j of all areas A_j in the camp, and n is the number of RSPs. The inverse selection probability is multiplied by the number of households N_{ji} in structure j if there are multiple households within the same structure. See Särndal and Wretman (2003) and Himelein et al. (2017) for further discussion.

2.5. Random Walk

Random walk (or random route) surveys are extremely common in the developing world as a method to control costs when representative sampling frames are not readily available. These designs, however, are non-probabilistic and have been shown in the literature to generate biased estimates even under perfect implementation (Bauer, 2014, 2016; Himelein et al., 2017). The assumption of perfect implementation, however, is quite strong as interviewers have shown a preference for selecting respondents willing to participate in the survey (Alt, 1991), and a number of other studies found that data collected with random walk designs exhibit differences from known population statistics on gender, age, education, household size, and marital status (Bien et al. 1997, Hoffmeyer-Zlotnick 2003, Blohm 2006, Eckman & Koch 2016).

Probabilities of selection inherently cannot be calculated in a random walk sample design as no information is collected on how many structures are in the camp, or how likely it was that a given structure was the x^{th} structure along any path. Random walk must then assume all structures have the same selection probability, implying constant sampling weights. Therefore, the only component of the weights for the random walk is the sub-sampling of households within a selected structure: $w'_i = \frac{N_{ji}}{i}$.

2.6. Comparison of Methods

As mentioned above, stratified cluster samples with the canvassing of selected clusters is the most common sample design used to collect official socioeconomic statistics in the developing world, but in

other disciplines it is relatively rare. A review of published public health literature by Chen et al. (2018) found most surveys use probabilistic designs in the first stage, but random walk or similar methods in the second stage. Lupu and Michelitch (2018) suggest that the combination of random walk and quota sampling is the common approach for political science-themed surveys conducted in the developing world, with 77 percent of respondents to their expert survey using a variation on this design. Diaz de Rada and Martínez (2014) compare a combination of random walk and quota sampling (based on age and gender) to probability designs and find a more accurate estimation of age and educational attainment in the combined method than in the probability methods, but that the probability methods perform better for measuring unemployment. The authors cite the replacement protocols for the probability methods as a reason for the bias and attribute the use of quota sampling for the success in estimating age and education, compared to the gold standard of a high-quality probability sample design.

There are also a limited number of papers which directly compare two or three of the methods, but none that consider this wide range of alternatives. Chew et al. (2018) use a baseline convolutional neural network model on a gridded population sampling frame to select a sample of households in Nigeria and Guatemala. The authors found this technique to be on par with human canvassing in terms of accuracy, and to outperform other machine learning models based on crowdsourced or remote sensing data. Grais et al (2007) compared an unweighted random point selection methodology to a random walk in their study of vaccination rates in urban Niger. The authors do not find statistically significant differences between the methods, though the sample size was limited and both methods were non-probabilistic.

3. Design and Field Protocols

3.1. Experiment Design

This paper makes use of a dataset from the purposefully designed methodology experiment conducted in one section of the Protection of Civilians site 1 (PoC1, Figure 1), one of the largest IDP camps in Juba, South Sudan. To generate a gold standard as the basis of comparison, a household census was conducted between August and September 2017. During this exercise, 2,655 households were interviewed using a questionnaire designed to collect demographic information, dwelling characteristics, household consumption, and perception data. At the end of each census interview, households received a unique barcode that could be used to identify them later in the experiment.

To avoid changes in camp composition, immediately following the completion of the census fieldwork, the interviewers returned to the field to implement the experiment. Teams used each of the sample selection methods to identify which households would have been selected had that method been used for a survey. To avoid respondent fatigue, instead of re-asking the questionnaire, the interviewers simply scanned the unique bar code of the selected household. Once scanned, the barcodes created an observation in the method-specific dataset with the information captured in the census. Each sampling technique targeted about 322 interviews so that comparisons could be made between the methods using an identical sample size. There was, however, some non-response for each method if interviewers were not able to contact a household member who could provide access to the barcode, if the barcode had not been retained by the household, or if the barcode was not scanned correctly. Protocols for each individual method are listed below.

3.2. Satellite Mapping

The Satellite Mapping method used a geo-referenced listing of structures in the PoC camp based on imagery from March 13, 2017, approximately five months before the start of fieldwork. The interviewer team was given 322 randomly selected structures as well as a list of replacement structures from a list of all geo-referenced structures in the PoC camp. Interviewers navigated to selected points using the GPS coordinates of the structure. Non-residential structures were substituted with replacement points. If there was more than one household residing within the selected structure, one household was randomly selected.

3.3. Segmentation

The objective of segmentation is to decrease the listing burden, generally for speed, financial, or security concerns. For this experiment, the PoC camp was divided into 19 clusters each containing 12 blocks of approximately 9 to 12 structures (Figure 3). The size of the blocks varied because, to the extent possible, segment boundaries followed easily discernible landmarks. Since the segmentation was done using satellite maps, it was not possible to distinguish between administrative or residential structures. To select the households for the survey, 16 of the 19 clusters were selected, then 10 of the 12 blocks within the cluster. To select individual households, the enumerators conducted a listing of all structures within the selected blocks and randomly selected two structures to be interviewed. This selection method yields a sample size of 320 households. Similar to satellite mapping, if there were multiple households within the structure, one was randomly chosen for the interview.

3.4. Grid Method

The grid method is similar to segmentation, but instead of purposefully-drawn, approximately equal population segments, the PoC camp was overlaid with a grid of cells measuring 50 meters by 50 meters. For the fieldwork, 27 grid cells were selected. After a listing of structures in each selected cell, 12 structures were randomly selected in each cell. Within each structure, a random household was selected in the case of multiple households per structure. Structures that fell into more than one cell were assigned to a single cell hosting the majority of the area of the structure. This determination was made in the field.

The loss of control over the number of households within the primary sampling unit, which in this case is the grid cell, complicates the selection process. If all grid squares contained at least 12 households, then it would be necessary to select (with equal probability) about 27 squares to reach the target sample size of 322 households. The number of households in each grid square varied from 1 to 136 households, with 13 grid squares containing fewer than 12 structures (Figure 4). To reach the target sample size, grid squares were randomly ordered and the first 27 were selected. The expected sample size is then calculated by assuming that 12 structures are selected from each grid square in the case of grid squares containing more than 12 structures, and all households are selected if there are less than 12 structures in the grid square. If the expected sample size is 310 or less, an additional grid square is selected up to a maximum expected sample size of 328. The result is that while on average the total sample size was 322, the simulations gave a range between 317 and 328.

3.5. North Method

The North Method uses RSPs to determine the selected households. RSPs are chosen from the universe of all possible points with the boundaries of the PoC camp. To implement the North Method, 322 RSPs along with replacement RSPs were chosen. These points were random geo-coordinates within camp borders (Figure 5). If the RSP lay within a structure, the corresponding structure was selected. If not, starting at the selected RSP, enumerators walked directly north, using the compass application on their tablet, until a structure was encountered. If the structure was residential, the structure was chosen to be interviewed. In the case of multiple households present in the structure, one household was randomly chosen. If the structure was not residential or if the enumerator reached the boundary of the camp, a replacement RSP was used.

As it would be extremely difficult to determine the area of the shadow in the field, satellite imagery is used for these calculations. In the case of this experiment, the selection areas are calculated using Google Earth imagery taken on December 22, 2017, approximately one month after the census of households in the PoC camp. Given the dependence of the North Method on having current satellite imagery for accurate calculations, the availability of this imagery is a major consideration for this method. The weights would be over-estimated if new structures had been built in the shadow since the imagery was taken.

3.6. Random Walk

Random Walk obtains a sample by randomly selecting starting points for enumerators with generic but unambiguous instructions to select households at regular intervals on their path. For this experiment, enumerators conducted random walks using 21 RSPs (Figure 8). Starting as near as possible to the RSP, the supervisor chose any random point (like a street corner or a school). From this point, four enumerators walked each in one of the four cardinal directions. Walking in their designated direction away from the RSP, they counted structures on both the right and the left and each selected the fifth structure for interview. Enumerators were instructed to start with the buildings on the right if two buildings were opposite to each other. To select the next structure, enumerators continued along the cardinal path, and selected the next fifth structure. If the enumerator could not proceed on its cardinal path because she had reached the boundary of the PoC camp, enumerators were instructed to turn right at a 90-degree angle and continue counting until finding the fifth dwelling. Enumerators had to conduct six interviews along their paths.

4. Implementation Issues

4.1. Failure to Follow Survey Protocols

As noted above, even if field protocols are perfectly implemented, the estimates generated from Random Walk designs are likely to be biased. Enumerators furthermore often were unable or unwilling to follow the protocols. Streets and paths were not necessarily aligned with cardinal directions and obstacles further impeded the ability to follow a straight path. Additionally, since the selection method requires enumerator judgment, it is not replicable and therefore allows enumerators greater discretion to choose which households are “selected.” Figure 9 shows the paths taken by two teams of enumerators from

random starting points. The team starting from point 16 more or less followed the field protocols, traveling a straight line until reaching the edge of the camp, and then making a right turn. The team starting from point 11 had more difficulty following the protocols. The enumerator traveling west actually travelled in a south-westerly direction and the enumerator traveling east followed a jagged path. These deviations can further increase error in a method which is already known to deliver biased results.

4.2. Structure Identification Issues

A key challenge in using GPS-based sampling strategies is to efficiently match the information from the satellite maps to the information collected on the ground. In the Satellite Mapping method, the interviewers must be able to match the GPS coordinates generated on the satellite map to actual structures on the ground. In the case of the experiment fieldwork, interviewers were not able to match the GPS coordinates to a structure in 15 of 322 cases. In addition, in one case the structure was out of scope, identified as a shipping container being used as a school.

When using the North Method, there is the opposite issue of matching the GPS coordinate captured at the time of the interview to a structure on the satellite map. To calculate the weights for the North Method, the analyst must be able to identify the interviewed structure and calculate its 'shadow.' However, it was not possible to match the selected structure captured by a GPS reading at the time of the interview to a household in the satellite map in 10 of 322 cases in the experiment and 132 out of 2,655 households in the complete census of the camp. Due to GPS error, outdated maps, or interviewer error, the GPS positions of those interviews were not located within a structure in the satellite imagery. In these cases, the sampling weight of the closest (or the average for multiple closest) household(s) was used for the respective household.

4.3. Non-Response

The protocol for conducting interviews stipulated that households had to be visited three times if no knowledgeable adult was present. If, after the third visit, still no person was available to be interviewed, it was marked as a case of non-response. The cases of non-response were replaced for all methods except the census. For the methods which rely on simple random sampling (satellite mapping) or random point selection (North method and random walk), replacements consisted of additional random selections. For segmentation and the grid method, additional households were selected from the segment or grid square listing.

In the census, non-response was low, with interviewers unable to conduct interviews in only 36 out of 2,655 households, or 1.4 percent, due mainly to refusals or no adult being present in the household at the time of the repeated interview requests. For the sampling methods, the replacement rates were 1.1 percent for segmentation, 3.9 percent for the North method, 5.7 percent for satellite mapping, and 10.1 percent for grid square selection. The replacement rate for segmentation may, however, be artificially low as enumerators would have the incentive to not list households in the listing that they knew would not be home to respond. The North method may also be artificially low as random point selection similarly

gives the enumerator the possibility to evade the strict protocol. Enumerators could unofficially replace a non-responding household by going to an adjacent structure rather than obtaining a new random point.

Weights must be adjusted for all sampling methods to compensate for non-response. Hence the final weight is

$$w_{i,m} = w'_{i,m} * nr_{i,m}$$

Here, $w'_{i,m}$ is the selection weight for household i using method m , and $nr_{i,m}$ the non-response weight.

4.4. Multiple Households per Dwelling

In all methods except segmentation, structures are selected instead of households. In the case where a structure is occupied by only one household, there are no further stages of selection and the interviewer proceeds with the questionnaire. If there are multiple households, however, the interviewer must randomly select one for the interview. This additional selection increases the potential for non-sampling error as the interviewer must implement the randomization procedure correctly in a setting where it is difficult, if not impossible, to verify. If randomization is done correctly, there will be no additional bias, but the extra stage will decrease the efficiency of the estimate and increase its standard error.

The frequency of selecting structures with multiple households varies by method (Table 1). The impact is the lowest for grid squares (1.06), segmentation (1.08), and random walk (1.08). The percentage was much higher for the North Method (1.17) because larger structures have larger footprints and often have larger shadows, and, thus, are more likely to both be selected and to contain multiple households. The higher probabilities of selection, however, are accounted for in the weight calculations, and therefore the resulting statistics are unbiased, assuming the first stage of selection was implemented without bias. The highest percentage of structures containing multiple households, however, was found with the satellite mapping method (1.25). Since structures were randomly selected from a list of all structures, there is no theoretical reason why there should be more multiple household structures with satellite mapping, so this observation may be related to availability bias for larger households with available respondents.

5. Results

The objective of our analysis is to compare multiple sources of error and uncertainty in each of the five methods. In terms of the sources of error, we examine the bias inherent in the method design; non-sampling error common to all five methods; and non-sampling error specific to the method. To examine bias inherent to the method, we use simulations assuming perfect implementation based on the census data. Non-sampling error common to all methods is mainly availability bias (Cuddleback et al, 2004), which we explore by comparing estimated household size and other measures correlated with household size. We look also at variables uncorrelated with household size to explore method-specific error. In terms of uncertainty, we look at both the overall design effects as well as decompose those effects into the unequal weight effect (UWE) and the cluster effect (our survey has no stratification) to further understand how much of the observed uncertainty is related to the method and how much is specific to the somewhat unique circumstances of the South Sudan refugee camp context (Liu, Iannacchione, and Byron, 2002).

Finally, we also look at the mean square error (MSE) as this measure takes into account both bias and uncertainty.

5.1. Bias

5.1.1. Household Size and Other Correlated Demographic Variables

All simulation results generate estimated average household sizes which contain the true mean within the confidence interval (Table Excel). Compared with a census mean of 4.28, the simulation results for grid squares, North method, satellite mapping, and segmenting were all within 0.08 percent of the true mean, while the random walk results were almost thirty times higher at 2.2 percent - clearly biased compared to the probability methods. The experimental results all statistically significantly overestimated household size compared to the census mean. The survey methods yield means with biases of 12.4 percent for satellite mapping, 8.8 percent for segmenting, 13.6 percent for grid squares, 16.2 percent for the north method and 15.5 percent for random walk. This over-estimation is caused by a systematic tendency of enumerators to select larger households because they are more likely to find an adult respondent (Cuddleback et al, 2004). As larger structures often have more rooms, the results are further confirmed by a similar upward bias for the number of rooms in the experimental results. All of the experimental methods overestimate the number of rooms compared to the census mean by between 8.2 and 12.2 percent, with the largest overestimation generated by the methods for which the probability of inclusion is higher for physically larger structures: 10.3 percent for grid squares, 10.4 percent for the north method, and 12.2 percent for random walk.

Table 2 shows the distribution of household size by method and uses a likelihood test to check for differences from the census distribution. Satellite mapping is not statistically significantly different from the census distribution, grid squares and segmenting are weakly significantly different, and random walk and the north method are significantly different. Figure 11 shows the distribution of household by household size in the census and from the North method, with the latter showing the highest degree of bias compared to the census mean. The North method captures less than half the percentage of single member households as were found in the census (8.0 percent compared to 16.8 percent). Though similar patterns are found for all methods, the methods which do not select a specific structure, the North method and the random walk, show higher degrees of availability bias than those methods in which the selection can be verified.

Other demographic variables, including adult equivalent household size and the adult-to-member ratio of household members, are highly correlated with household size and therefore show similar patterns as household size. In the simulations, the results for the adult equivalent household size were within 0.05 percent of the census mean for all methods except random walk, which generated a bias of 1.4 percent. In the case of the adult ratio, the differences were less stark, with bias estimates ranging between 0.01 percent and 0.31 percent for the probability methods, compared to 0.41 percent for random walk. The experimental results also overestimate the mean for the adult equivalent household size measure, with the largest overestimation found for segmenting and random walk. For the ratio of adults to total household members, all methods underestimate the ratio compared to the census. This underestimation is again related to the tendency of the experimental methods to select larger households, which have larger numbers of children and consequentially lower adult ratios.

5.1.2. Variables Correlated with Household Size

Other variables considered in the analysis that are positively but not highly correlated with household size are if the household head had ever attended school (correlation = 0.191) and if the household head can read in any language (correlation = 0.183). While the simulation generates unbiased estimates for the probability methods (ranging from 0.03 percent to 0.44 percent), the random walk shows substantially higher bias with 1.63 percent for school attendance and 1.77 percent for being able to read. All experimental methods show lower percentages of school attendance (from 4.92 percent to 15.21 percent) and literacy (from 0.12 percent to 13.92 percent) than the census. This finding could reflect more difficulties in finding work for those heads with lower levels of education and, therefore, a higher likelihood of being found at home by the interviewer.

5.1.3. Variables Uncorrelated with Household Size

We consider three variables, which are uncorrelated with household size. The variables are whether the respondent / household owns a mobile phone (correlation = 0.043), owns a mattress (correlation = 0.031), and wants to leave this location (correlation = 0.015). The simulation confirms that the probability methods yield largely unbiased results for wanting to move, ranging between 0.13 percent and 0.24 percent, but ownership of a mattress and mobile phone are slightly more biased ranging from 0.28 percent to 0.70 percent, again excluding random walk. For the random walk simulations, the bias was 0.71 percent for mobile phone ownership, 1.37 percent for desire to move location, and 3.02 percent for owning a mattress. The estimates from the experiments vary between methods and indicators. The satellite mapping generates the largest bias (ranging from 4.88 percent to 12.61 percent) followed by North method (from 1.93 percent to 7.53 percent) and segmenting (from 0.61 percent to 9.18 percent) with best results yield by grid squares (from 1.40 percent to 6.74 percent) and random walk (from 1.01 percent to 7.84 percent).

5.1.4. Poverty Variables

Simulations using the probability methods confirm largely unbiased results for consumption (total, per capita and per adult equivalent), with bias ranging from 0.01 percent to 0.2 percent for the satellite mapping, segmenting, and north methods, and being slightly higher for the grid squares methods, ranging from 0.38 percent to 0.82 percent. Bias was slightly higher for poverty measures (per capita and per adult equivalent), with the probability methods ranging from 0.07 percent to 0.54 percent. For the random walk method, the biases from the simulations were generally above 1 percent, and as high as 2.35 percent for per capita poverty. The results from the experiments show an upward bias for total consumption around 7 percent, except for satellite mapping with a bias of almost 25 percent, and a largely unbiased estimate from the grid squares methodology. Per capita and per adult equivalent consumption are largely biased downwards due to the upward bias in household size, except for satellite mapping in which the upward bias in household size is more than offset with a large upward bias in total consumption. Accordingly, per capita and per adult equivalent poverty measures are biased upwards (from 7.57 percent to 19.96 percent, respectively) with the exception for satellite mapping (downwards by 4.44 and 5.33 percent, respectively). The experimental random walk results are also upwardly biased at around 10 percent, consistent with what was found in other methods.

5.1.5. Summary of the Bias

To summarize the performance across the different indicators, we rank the bias for each indicator across methodologies but separately for simulations and experiments, and take the average of the rank across indicators for each methodology (Figure 11). Satellite Mapping, Segmenting and Grid Squares show similar performance for the simulation and the experiment, indicating that their relative performance vis-à-vis other methodologies remains similar. In contrast, the North Method's performance deteriorates substantially in the experiment. Given the relatively low complexity of the implementation of the North Method, it suggests that enumerators are not complying with the protocol. Interestingly, the Random Walk, which does not show a strong relative performance for simulations, regains ground in the experiment, performing substantially better than the North Method, even though it is a non-probabilistic method. Thus, implementation and monitoring of compliance with the protocol is critical to take into consideration when deciding on sampling approaches.

5.1.6. Meta-Analysis of Bias

To better understand the different factors that impact the accuracy of the five sampling methods being studied here, we undertook analysis of the simulated and observed results pooled across the five methods, controlling for difference characteristics of the particular questions and clustering the standard errors at the question level. The dependent variable for this analysis is the absolute value of the normalized mean of the bias, or the $(\text{observed value} - \text{census value}) / \text{census value}$. In addition to the type of sampling method, three question-level measures are included in the analysis: coefficient of variation, correlation with household size, and four versions of the Moran's I spatial dispersion. The coefficient of variation is a measure of the inherent variability of responses across the census values for a particular question and was included to control for higher variation variables being more prone to sampling error, which would show up as bias in the results. The correlation with household size was included as household size is a variable known to be impacted by availability bias, which is common across all five methods. Finally, two versions of the Moran's I spatial dispersion statistics (at 16m and 32m) were included to understand the impact of clustering within the PSU. If the spatial dispersion index were zero, there is no relationship between measured value of a certain household and those of their neighbors. In a case where an attribute is completely randomly distributed throughout the population, then the impact of the selection of sampling method is limited as you could speak with any 12 households and have a random set of responses.

Columns 1-4 in Table 2 in the appendix show the results for the pooled regressions for the simulated results. The results on method are consistent across all four specifications. Compared to the reference method of satellite mapping, the North Method is unbiased, while the Segmenting and Grid Square methods show minimal bias (0.1 percent and 0.2 percent, respectively). The Random Walk method shows 1.2 percent bias on average across the 14 questions. The additional controls for the coefficient of variation, correlation with household size, and spatial dispersion are also not significant.

Columns 5-8 in Table 2 show the results for the same specification using the observed results. The R2 for these models is substantially lower than for those with the simulated results. The base model including only the sampling methodologies has an R2 of 0.593 for the simulated results compared to only 0.097 for

the observed results and none of the variables for the sampling method are statistically significant in the observed models – indicating there is much more noise in the observed models compared to the simulated. The variables that contribute additional explanatory power also vary between the simulated and observed results. In both cases the addition of the coefficient of variation yields only a negligible increase. In the simulated models, there are similar small changes when the correlation with household size and the spatial dispersion measures are introduced. In the observed models, the R2 more than doubles to 0.215 when the correlation with household size variable is added, and the variable itself is strongly significant. This finding demonstrated the strong influence of availability bias across the five methods. The R2 also nearly doubles with the addition of the two spatial dispersion variables though the coefficients are not significant.

5.2. Normalized Root Mean Squared Error

The root mean squared error (RMSE), or the square root of the average of the deviation of the estimated mean from the true mean, accounts for both bias and uncertainty. Lower bias and lower variance yield lower values of RMSE and therefore lower RMSEs are preferable to higher RMSEs in evaluating methodologies. See section 7.3 in the appendix for more detail. In this application, because we have both continuous and dichotomous variables, we use the normalized RMSE (nRMSE) to facilitate comparability. Figure 12 and Figure 13 below compare the nRMSEs across the methods and questions. On average segmenting has the lowest nRMSE (1.67), followed by random walk (1.63), satellite mapping (1.70), grid squares (1.79) and the north method (2.14). As shown in Figure 12, however, the results for satellite mapping are skewed by one outlier value on total weekly household consumption. Excluding that value, the nRMSE for satellite mapping is 1.49. Overall, the method that perform the best is segmenting, which has the lowest or second lowest nRMSE for 12 of the 14 questions, and the highest or second highest for only two questions, followed by satellite mapping, which gives the best or second-best results in 9 questions and the worst or second worst results for 4 questions. The methods that perform the worst is the North Method, which does not give the best or second-best results for any of the questions and gives the worst of second worst results for 9 of 14 questions, followed by the random walk, which give the best or second best results for 4 questions and the worst or second worst results for 7 questions. The final method, grid squares, gives the best or second-best results for 3 questions and the worst or second worst results for 6 questions.

6. Discussion

We find that simulations arrive at the true household size distribution, while all simulations over-estimate household size. This over-estimation is caused by a systematic tendency of enumerators to select larger households because they are more likely to find an adult respondent. Specifically, the North method and the random walk show higher degrees of availability bias than those methods in which the selection can be verified, e.g., satellite mapping where a specific structure is chosen a priori. Also, for other indicators, including poverty estimates, we find that simulations obtain unbiased results while the actual experiments are biased, especially for variables correlated with household size. Summarizing the results across indicators shows that Satellite Mapping, Segmenting and Grid Squares have similar relative performances theoretically in the simulations and practically in the experiments. However, the North Method's performance deteriorates strongly in the experiment, allowing Random Walk to regain ground. Thus,

implementation matters. Pooling the analysis across indicators and using satellite mapping as reference, the North Method is unbiased, while the Segmenting and Grid Square methods show minimal bias (0.1 percent and 0.2 percent, respectively). The Random Walk method shows 1.2 percent bias on average across the 14 questions. In conclusion and in line with the literature, most probability-based methods perform better than non-probability methods like random walk. In addition, implementation of adherence with the survey protocol is extremely important and using appropriate methods and tools to cope with this challenge is absolutely mandatory for coming as close as possible to the theoretical results derived by the simulation for the probability-based methods. In practice – in a fragile setting like South Sudan – deviations from the survey protocol, measured as differences between the experiments and the simulations, have large influence on the actual bias of estimates.

7. Appendix

7.1. Simulation and Frame

To compare the efficiency of the different sampling frames and designs, we will apply an empirical sampling simulation. In this type of (Monte-Carlo style) simulation, either a true or synthetic population is used as the target population. By applying a specific sampling design, and repeated sampling (usually 1,000 repetitions) under this design, we can compare the resulting population estimates with the known true population values for each run of the simulation.

The resulting distribution of these estimates is called the sampling distribution, and the average squared deviation from the underlying population value is the Mean Squared Error (MSE) or when taking its square root, the Root MSE (RMSE). To facilitate the comparison, we use the relative version expressed in percentage deviation.

Empirical sampling simulations can be considered as the “[...] ultimate tool for investigators who want to know if one sampling strategy will work better than another for their population.” (Thompson, 2012). However, this requires the underlying simulation population to replicate as realistically as possible the target population.

7.2. Quality Metrics

A standard Measure in the assessment of a sampling designs is the Root Mean Squared Error (RMSE) and calculated as:

$$RMSE = \frac{\sum_{sim=1}^{1000} RMSE_{sim}}{1000} = \left[\frac{1}{1000} \times \frac{\sqrt{(\hat{Y} - Y)^2}}{Y} \right] \times 100$$

Expressed here as percentage deviation from the population mean Y and calculated for each parameter of interest.

Equation .. is only the empirical representation though and a result of rearranging the definition of the Mean squared Error,

$$\begin{aligned} MSE(\hat{Y}) &= E(\hat{Y} - Y)^2 = E[(\hat{Y} - \tilde{Y}) + (\tilde{Y} - Y)]^2 \\ &= E(\hat{Y} - \tilde{Y})^2 + 2E(\hat{Y} - \tilde{Y})(\tilde{Y} - Y) + (\tilde{Y} - Y)^2 \end{aligned}$$

And decomposing it into

$$MSE(\hat{Y}) = Var(\hat{Y}) + Bias(\hat{Y})$$

with \hat{Y} , \tilde{Y} and Y being the estimate from the sample, the mean of this estimate and the true value in the population respectively. Var is the corresponding variance, and $Bias$ the resulting bias component, which is defined as:

$$Bias(\tilde{Y}) = \tilde{Y} - Y$$

If the mean of the estimator and the population mean are the same, the bias is 0. And the MSE would be equal to the variance of the estimate, which is only a result of the sample size. However, in a real survey situation, the population mean is commonly unknown, the resulting MSE therefore captures both, the variance, and the bias. Since a sampling frame which is not covering the target population well, is likely to produce a different mean for the variable of interest than its true population mean, we may expect the bias to be different from 0.

References

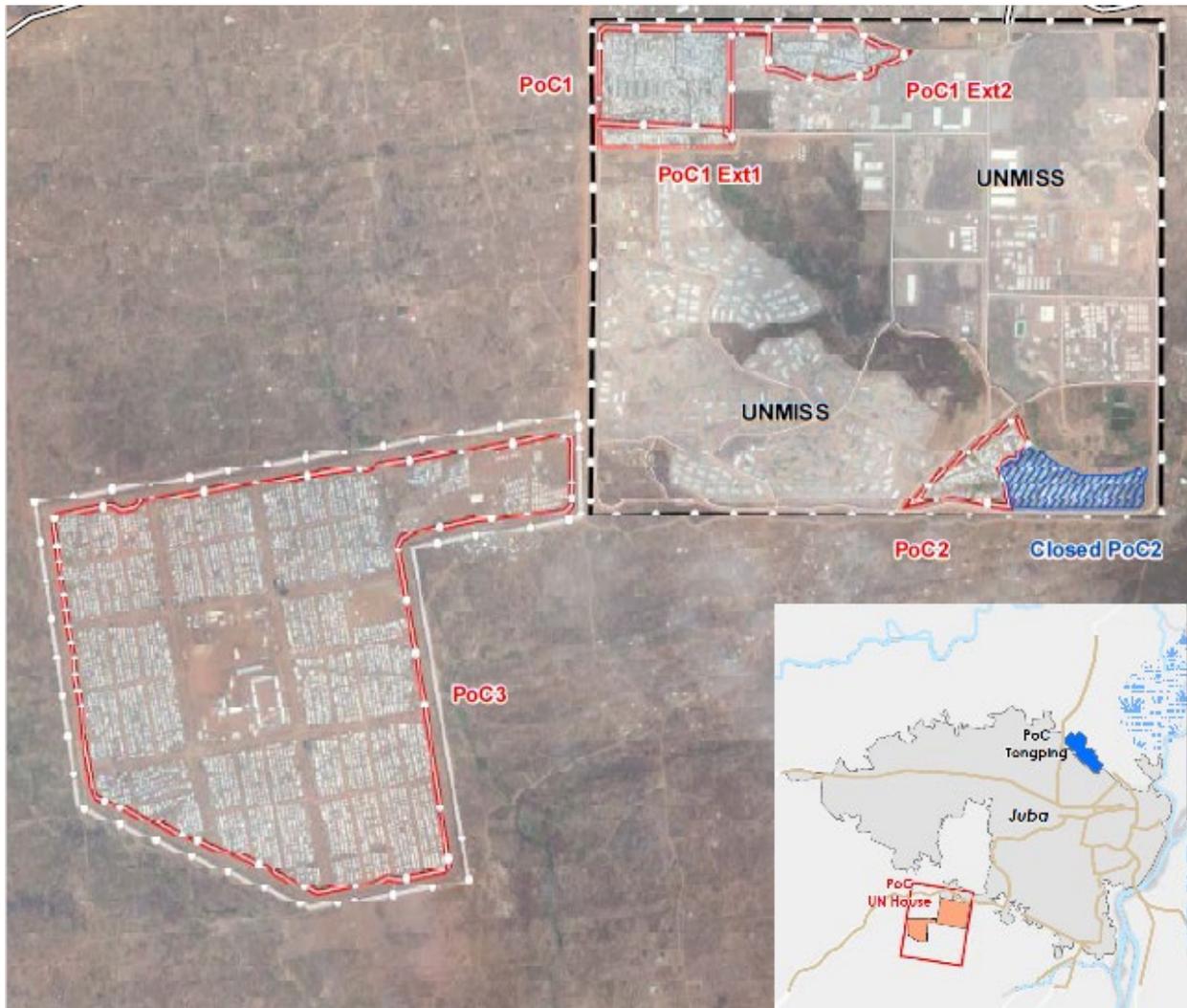
- Alt, C., Bien, W., and Krebs, D., 1991. Wie zuverlässig ist die Verwirklichung von Stichprobenverfahren? Random route versus Einwohnermeldeamtsstichprobe. *ZUMA-Nachrichten* 28, 65–72.
- Barrett, D.F., Beaghen, M., Smith, D. and Burcham, J., 2002. Census 2000 housing unit coverage study. In *Proceedings of the Section on Survey Research Methods, American Statistical Association* (pp. 146-151).
- Bauer, J.J., 2014. Selection errors of random route samples. *Sociological Methods & Research*, 43(3), pp.519-544.
- Bauer, J.J., 2016. Biases in random route surveys. *Journal of survey statistics and methodology*, 4(2), pp.263-287.
- Bien, W., Bender, D., & Krebs, D. 1997. DJI-Familiensurvey: Der Zwang, mit unterschiedlichen Stichproben zu leben. In *Stichproben in der Umfragepraxis* pp. 127–147. Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-322-86533-5_10
- Blohm, M. 2006. Data Quality in Nationwide Face-to-face Social Surveys. Retrieved from <http://www.ons.gov.uk/ons/about-ons/get-involved/events/events/q2006---european-conference-on-quality-in-survey-statistics-24-26-april-2006/agenda/session-17-wednesday.pdf>
- Boo, G., Darin, E., Thomson, D.R. and Tatem, A.J., 2020. A grid-based sample design framework for household surveys. *Gates Open Research*, 4.
- Chen, X., Hu, H., Xu, X., Gong, J., Yan, Y. and Li, F., 2018. Probability Sampling by Connecting Space with Households Using GIS/GPS Technologies. *Journal of Survey Statistics and Methodology*, 6(2), pp.149-168.
- Chew, R.F., Amer, S., Jones, K., Unangst, J., Cajka, J., Allpress, J. and Bruhn, M., 2018. Residential scene classification for gridded population sampling in developing countries using deep convolutional neural networks on satellite imagery. *International Journal of Health Geographics*, 17(1), p.12.
- Cuddeback, G., Wilson, E., Orme, J.G. and Combs-Orme, T., 2004. Detecting and statistically correcting sample selection bias. *Journal of Social Service Research*, 30(3), pp.19-33.
- Diaz de Rada, V. and Martín, V.M., 2014. Random route and quota sampling: Do they offer any advantage over probably sampling methods?.
- Eckman, S., 2013. Do different listers make the same housing unit frame? Variability in housing unit listing. *Journal of Official Statistics*, 29(2), pp. 249–259, doi: 10.2478/jos-2013-0021
- Eckman, S. and Koch, A. 2016. “Are High Response Rates Good for Data Quality? Evidence from the European Social Survey” Paper under review.
- Elangovan, A., Elavarsu, G., Ezhil, R., Prabu, R. and Yuvaraj, J., 2016. Geospatial Techniques in Health Survey to Overcome the Lack of Sampling Frame. *International Journal of Advanced Remote Sensing and GIS*, pp.pp-1908.

- Escamilla, V., Emch, M., Dandolo, L., Miller, W. C., Martinson, F., & Hoffman, I. (2014). Sampling at community level by using satellite imagery and geographical analysis, *Bulletin of the World Health Organization*; 92(9): 690–694. doi: 10.2471/BLT.14.140756
- Galway, L. P., Bell, N., Al Shatari, S. A. E., Hagopian, A., Burnham, G., Flaxman, A., ... Takaro, T. K. (2012). A two-stage cluster sampling method using gridded population data, a GIS, and Google EarthTMimagery in a population-based mortality survey in Iraq. *International Journal of Health Geographics*, 11(April). <https://doi.org/10.1186/1476-072X-11-12>
- Grais, R.F., Rose, A.M., Guthmann, J.-P., 2007. Don't spin the pen: two alternative methods for second-stage sampling in urban cluster surveys. *Emerging Themes in Epidemiology* 4, 8. doi:10.1186/1742-7622-4-8
- Grosh, M.E., Munoz, J., 1996. A manual for planning and implementing the living standards measurement study survey (No. LSM126). The World Bank.
- Himelein, K., Eckman, S., Murray, S. and Bauer, J., 2017. Alternatives to full listing for second stage sampling: Methods and implications. *Statistical Journal of the IAOS*, 33(3), pp.701-718.
- Hoffmeyer-Zlotnik, J. H. 2003. New sampling designs and the quality of data. *Developments in Applied Statistics*. Ljubljana: FDV Methodoloski Zvezki, pp. 205–217.
- Kassié, D., Roudot, A., Dessay, N., Piermay, J.-L., Salem, G., & Fournet, F. (2017). Development of a spatial sampling protocol using GIS to measure health disparities in Bobo-Dioulasso, Burkina Faso, a medium-sized African city. *International Journal of Health Geographics*, 16(1), 14. <https://doi.org/10.1186/s12942-017-0087-7>
- Kish, L. 1965. *Survey Sampling*. Wiley Classics Library Edition published 1995. John Wiley & Sons, New York.
- Kolbe, A.R., Hutson, R.A. 2006. Human Rights Abuse and Other Criminal Violations in Port-Au-Prince, Haiti: A Random Survey of Households. *The Lancet* 368 (9538): 864–73. doi:10.1016/S0140-6736(06)69211-8.
- Kondo, M.C., Bream, K.D.W., Barg, F.K. Branas, C.C. 2014. A Random Spatial Sampling Method in a Rural Developing Nation. *BMC Public Health* 14 (1): 338.
- Kumar, Naresh. 2007. "Spatial Sampling Design for a Demographic and Health Survey." *Population Research and Policy Review* 26 (5-6): 581–99.
- Lin, Y., & Kuwayama, D. P. (2016). Using satellite imagery and GPS technology to create random sampling frames in high risk environments. *International Journal of Surgery*, 32, 123–128. <https://doi.org/10.1016/j.ijssu.2016.06.044>
- Liu, J., Iannacchione, V., and Byron, M. 2002. Decomposing design effects for stratified sampling. In *Proceedings of the survey research methods section, American Statistical Association* (pp. 2124-6).

- Lupu, N. and Michelitch, K., 2018. Advances in Survey Methods for the Developing World. *Annual Review of Political Science*, 21, pp.195-214.
- O’Muircheartaigh, C., English, E. and Eckman, S., 2007. Predicting the relative quality of alternative sampling frames. In *Proceedings of the Survey Research Methods Section, American Statistical Association*.
- Särndal, C., B., S., & and Wretman, J. 2003. *Model Assisted Survey Sampling*. New York: Springer Series in Statistics.
- Thompson, S.K., 2012. *Sampling* (Vol. 755). John Wiley & Sons.
- Shannon, H. S., Hutson, R., Kolbe, A., Stringer, B., & Haines, T. (2012). Choosing a survey sample when data on the population are limited: A method using Global Positioning Systems and aerial and satellite photographs. *Emerging Themes in Epidemiology*, 9, 1–7. <https://doi.org/10.1186/1742-7622-9-5>
- Wampler, P. J., Rediske, R. R., & Molla, A. R. (2013). Using ArcMap, Google Earth, and Global Positioning Systems to select and locate random households in rural Haiti. *International Journal of Health Geographics*, 12(1), 3. <https://doi.org/10.1186/1476-072X-12-3>

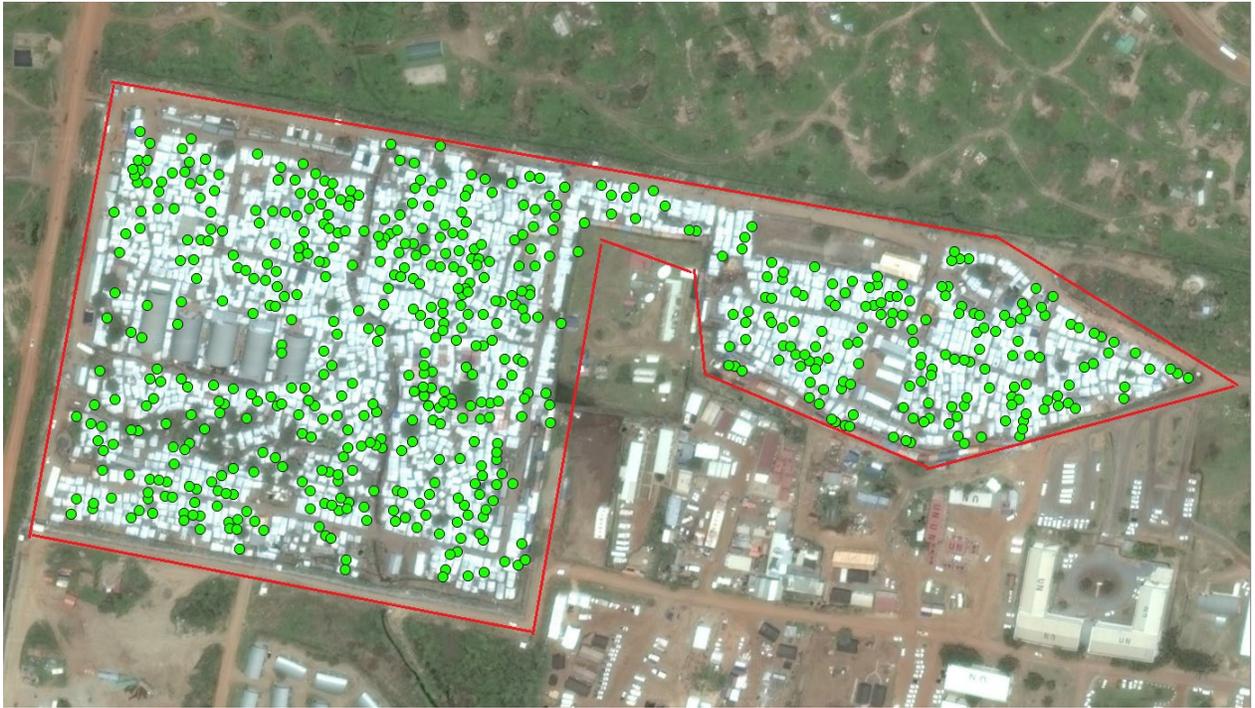
Figures

Figure 1: IDP Camp PoC1 in Juba, South Sudan



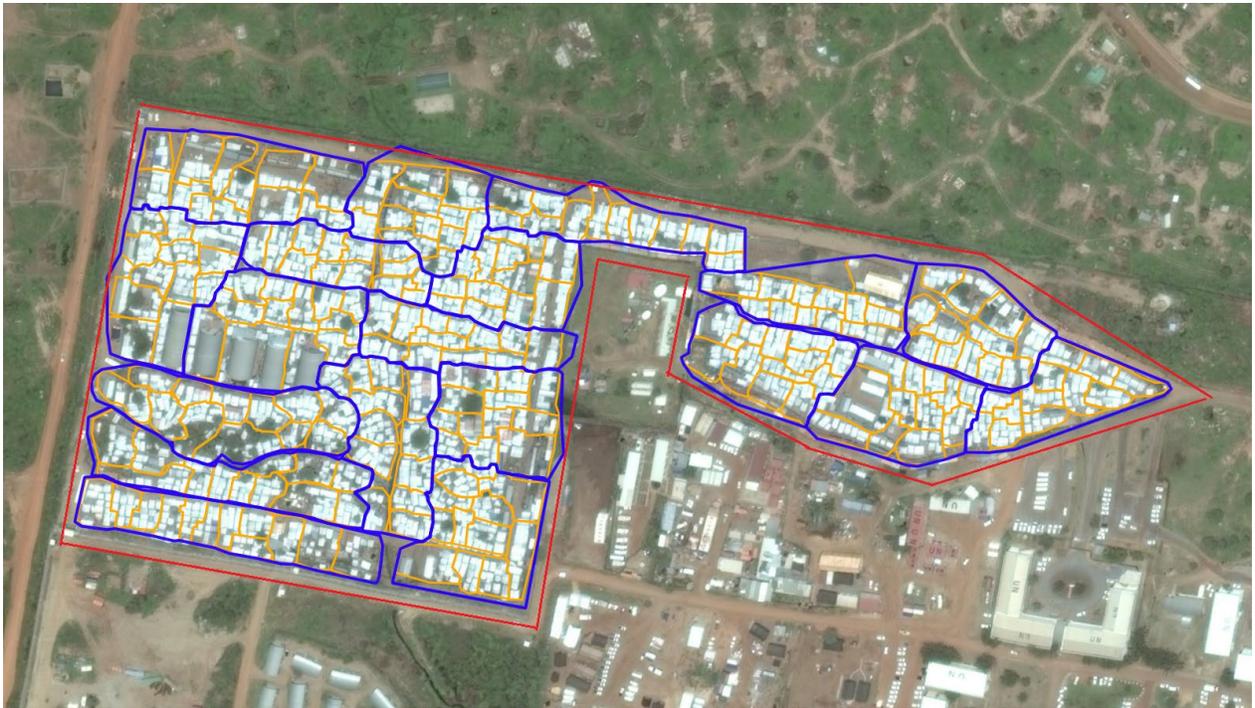
Source: UN OCHA, 2015.

Figure 2: Satellite mapping of residential structures



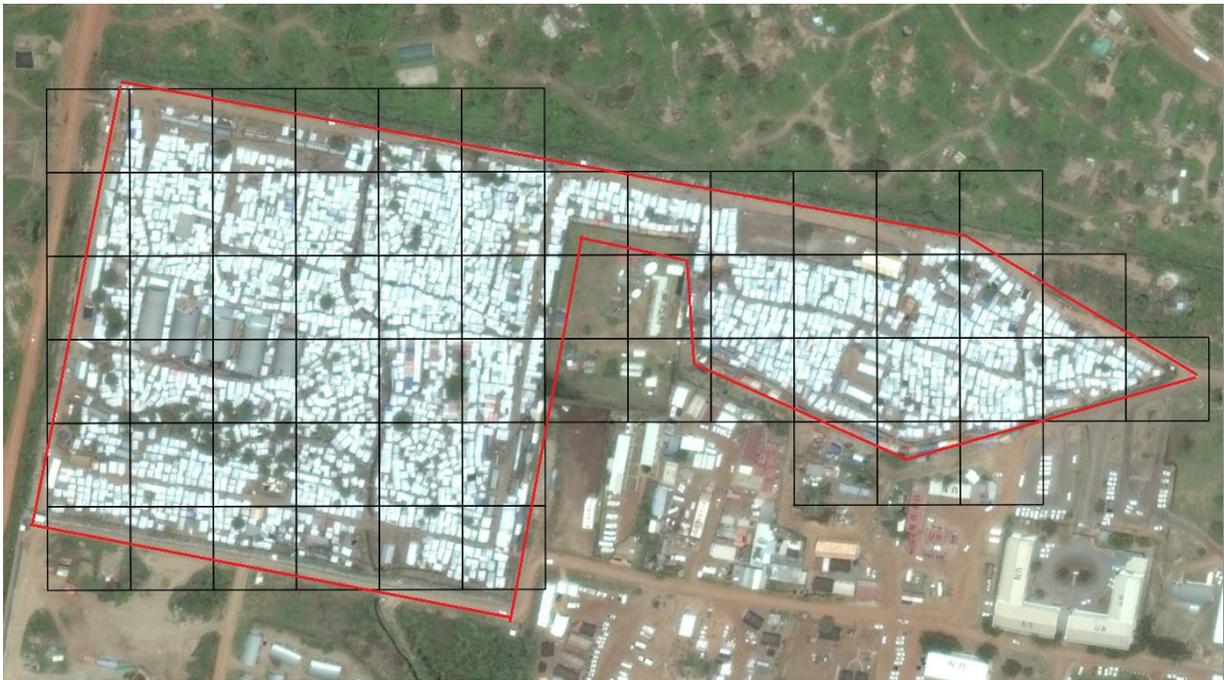
Source: Author's Calculation.

Figure 3: Enumeration Areas (blue) and Blocks (orange) for segmenting method.



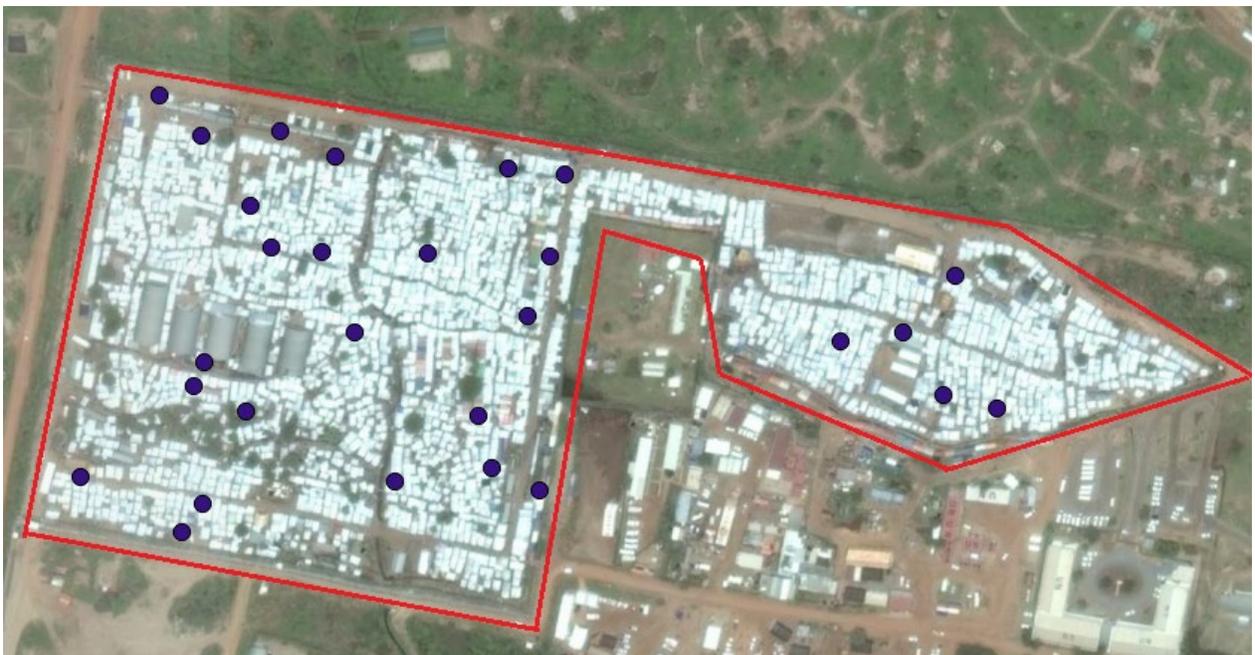
Source: Author's Calculation.

Figure 4: Grid overlay over the camp.



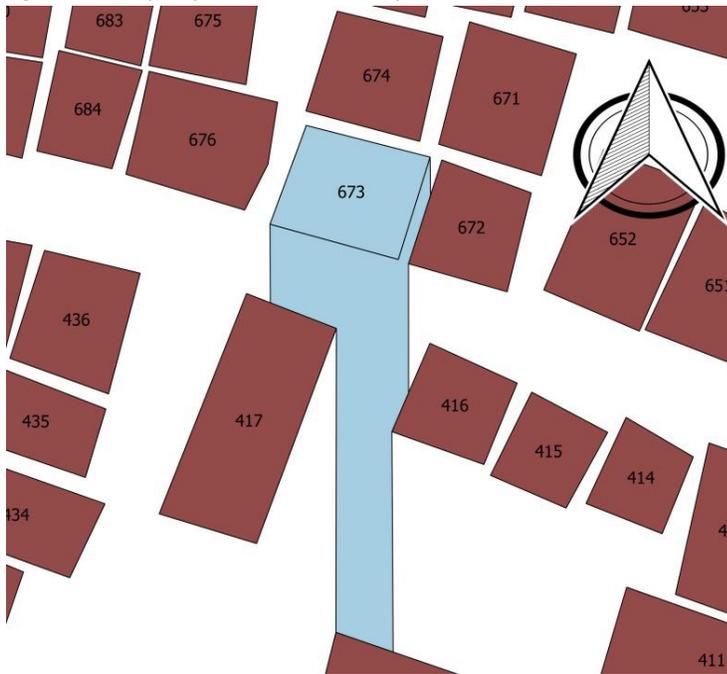
Source: Author's Calculation.

Figure 5: Random coordinates for the North Method



Source: Author's Calculation.

Figure 6: Example of the selection area of a structure.



Source: Author's calculation.

Figure 7: Areas leading to selection of given household in North method



Source: Author's Calculation.

Figure 10: Household size

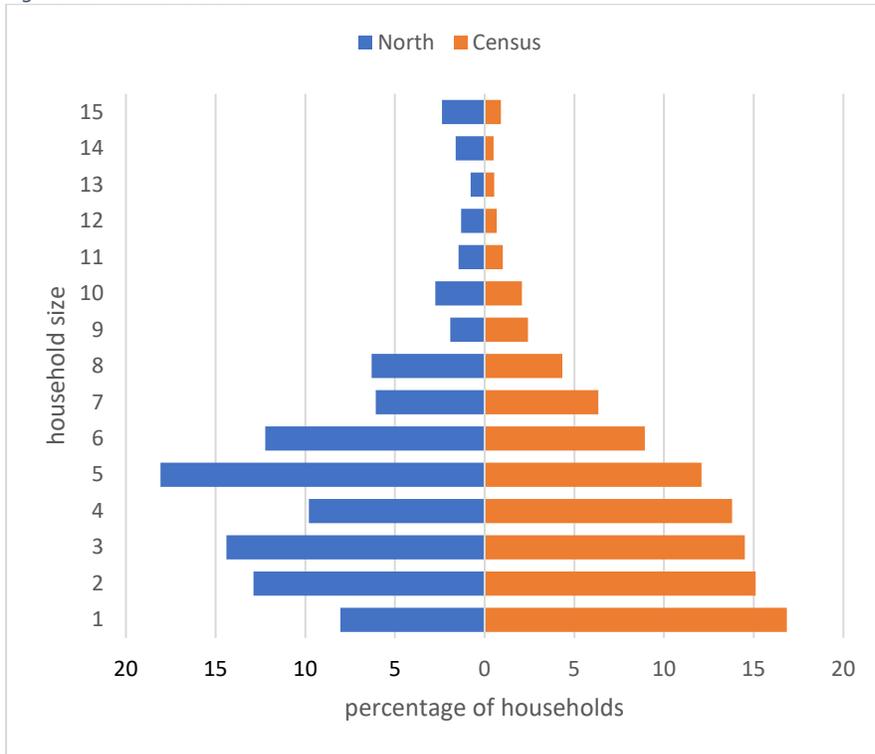


Figure 11: Rank of absolute bias for simulations and experiments.

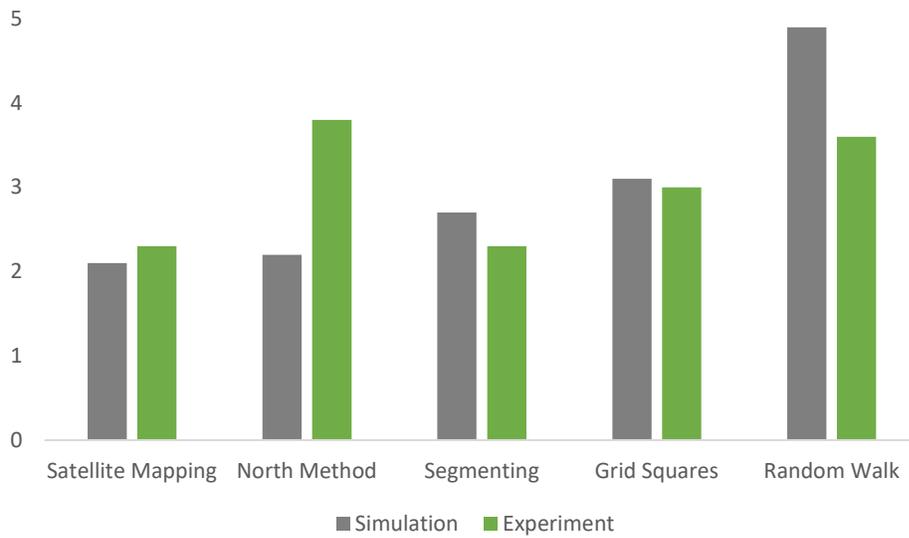


Figure 12. Normalized Mean Square Error by Method

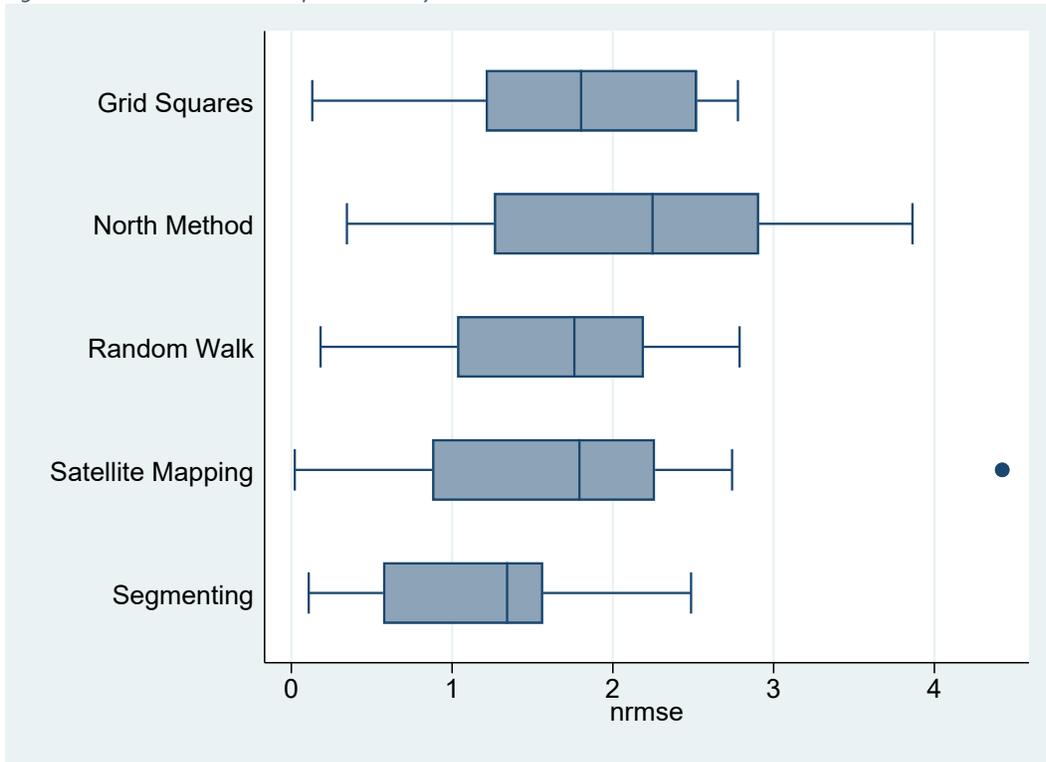


Figure 13. Normalized Mean Square Error by Method and Question

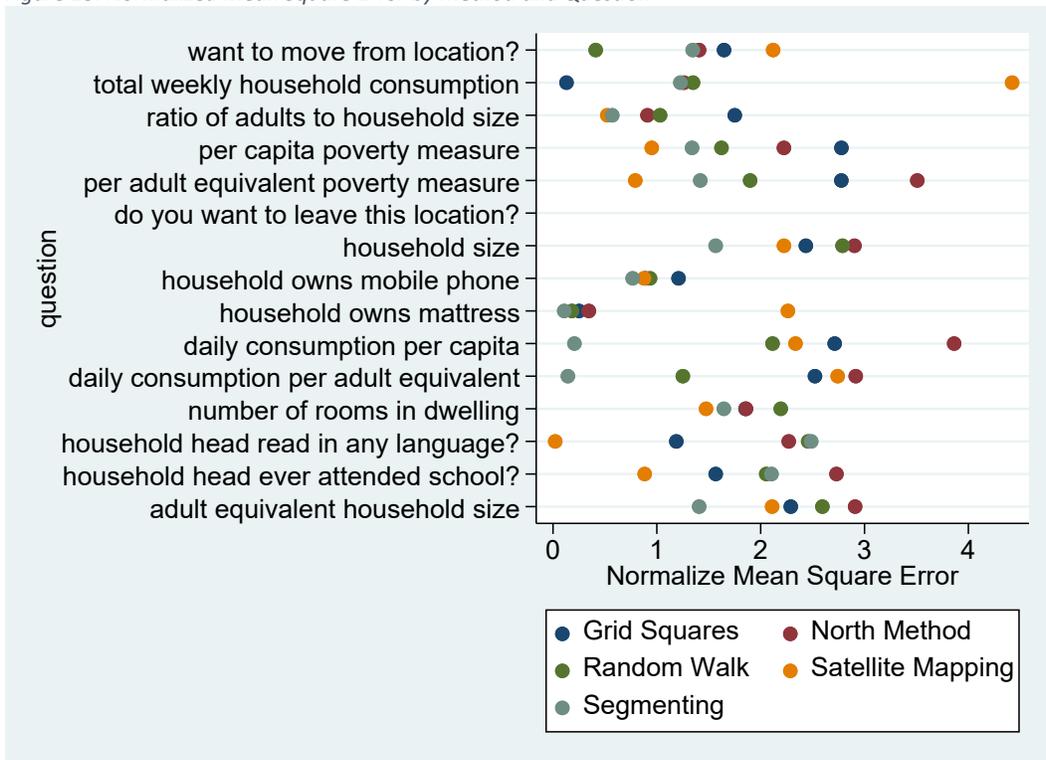


Table 1: Replacement Rate and Mean Number of Households per surveyed structure.

method	Replacement Rate (%)	Average households per structure in sample
Census	1.4	--
Grid Squares	10.1	1.06
North Method	3.9	1.17
Random Walk	--	1.08
Satellite Mapping	5.7	1.25
Segmenting	1.1	1.08

Source: Author's Calculation.

	Census	Grid Squares	North Method	Random Walk	Satellite Mapping	Segmenting
1	16.8	17.1	10.3	11.2	13.2	12.0
2	15.1	8.4	13.4	10.6	12.5	10.2
3	14.5	10.2	15.4	16.4	15.2	11.9
4	13.8	12.3	8.0	13.5	11.5	15.5
5	12.1	19.2	16.5	9.8	12.2	13.9
6	8.9	11.4	11.5	8.6	9.7	10.8
7	6.3	7.5	4.9	10.1	9.5	8.7
8	4.3	4.2	6.2	6.9	5.2	5.7
9	2.4	2.2	1.7	4.0	2.0	3.5
10+	5.7	7.4	12.1	8.9	9.0	7.9
likelihood-ratio chi2(9)		16.7	34.5	33.7	9.4	16.4
pr		0.053	0.000	0.000	0.397	0.060

Source: Author's Calculation.

method	Census		Grid Squares		North Method		Satellite Mapping		Segmenting		Random Walk	
			simulati on	experim ent	simulati on	experim ent	simulati on	experim ent	simulati on	experim ent	simulati on	experim ent
ratio of adults to household size	0.694	bias (%)	0.23%	-9.76%	-0.12%	-5.07%	0.01%	-2.91%	-0.31%	-3.19%	-0.41%	-5.75%
		deff	1.53	1.74	1.35	1.29	1.00	1.30	1.24	1.28	1.16	1.07
total weekly household consumption	2,870	bias (%)	0.82%	-0.74%	0.01%	7.17%	0.20%	24.80%	0.13%	6.93%	1.02%	7.57%
		deff	1.49	0.98	1.26	1.19	1.00	1.29	1.07	1.25	0.97	0.99
daily consumption per adult equivalent	164.5	bias (%)	0.42%	-14.35%	-0.24%	-16.59%	0.09%	15.37%	0.02%	0.81%	0.12%	-7.02%
		deff	1.51	1.11	1.30	1.192	1.00	1.349	1.07	1.200	1.12	1.078
daily consumption per capita	127.8	bias (%)	0.38%	-15.43%	-0.37%	-21.98%	0.03%	13.10%	-0.05%	-1.17%	-1.09%	109.45%
		deff	1.51	1.08	1.30	1.225	1.00	1.335	1.10	1.231	1.10	1.094
adult equivalent household size	2.86	bias (%)	0.00%	11.76%	-0.02%	12.76%	-0.04%	7.88%	-0.01%	14.46%	1.41%	16.21%
		deff	1.51	1.25	1.32	1.492	1.00	1.129	1.23	1.019	1.18	1.318
household size	4.28	bias (%)	-0.08%	13.57%	-0.05%	16.18%	-0.06%	12.39%	0.04%	8.77%	2.21%	15.54%
		deff	1.53	1.53	1.34	1.29	1.00	1.26	1.23	1.17	1.18	1.02
household head ever attended school? (%)	0.571	bias (%)	-0.44%	-8.73%	-0.15%	-15.21%	-0.03%	-4.92%	-0.08%	-11.78%	-1.63%	-11.44%
		deff	1.52	1.49	1.36	1.295	1.00	1.310	1.25	1.238	1.19	1.066
household head read in any language? (%)	0.587	bias (%)	-0.29%	-6.62%	-0.02%	-12.65%	-0.15%	-0.12%	-0.17%	-13.92%	-1.77%	-13.70%
		deff	1.54	1.51	1.36	1.29	1.00	1.29	1.25	1.28	1.20	1.067
household owns mobile phone (%)	0.528	bias (%)	0.50%	6.74%	0.57%	-4.96%	0.66%	4.88%	0.69%	4.29%	0.71%	5.21%
		deff	1.54	1.38	1.34	1.28	1.00	1.30	1.22	1.23	1.31	1.076
per adult equivalent poverty measure	0.430	bias (%)	0.41%	15.80%	0.14%	19.96%	0.28%	-4.44%	0.54%	8.02%	1.69%	10.66%
		deff	1.51	1.38	1.34	1.30	1.00	1.28	1.18	1.25	1.07	1.071

per capita poverty measure	0.604	bias (%)	0.07%	15.80%	0.10%	12.65%	0.14%	-5.33%	0.47%	7.57%	2.35%	9.11%
		deff	1.54	1.26	1.33	1.32	1.00	1.33	1.18	1.27	1.22	1.077
number of rooms in dwelling	1.38	bias (%)	0.12%	10.33%	-0.07%	10.36%	-0.03%	8.21%	-0.11%	9.21%	0.94%	12.22%
		deff	1.60	1.73	1.31	1.14	1.00	1.46	1.19	1.42	1.43	1.02
household owns mattress (%)	0.44	bias (%)	0.28%	-1.40%	0.69%	-1.93%	0.65%	12.61%	0.70%	-0.61%	3.02%	1.01%
		deff	1.56	1.20	1.34	1.28	1.00	1.31	1.15	1.25	1.02	1.071
want to move from location? (%)	0.56	bias (%)	0.24%	2.29%	0.14%	-7.53%	0.13%	11.81%	0.14%	9.18%	-1.37%	7.84%
		deff	1.53	1.08	1.35	1.22	1.00	1.22	1.06	1.31	1.15	1.28

Table 2. Multivariate Regressions on pooled simulated and observed results

	simulations					observed results		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	coef/se	coef/se	coef/se	coef/se	coef/se	coef/se	coef/se	coef/se
Reference: Satellite Mapping								
Grid Squares	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)	0.002*** (0.001)	-0.003 (0.032)	-0.003 (0.032)	-0.003 (0.032)	-0.003 (0.033)
North Method	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.015 (0.031)	0.015 (0.031)	0.015 (0.031)	0.015 (0.031)
Random Walk	0.012*** (0.002)	0.012*** (0.002)	0.012*** (0.002)	0.012*** (0.002)	-0.014 (0.028)	-0.014 (0.028)	-0.014 (0.028)	-0.014 (0.029)
Segmenting	0.001* (0.000)	0.001* (0.000)	0.001* (0.000)	0.001* (0.000)	-0.043 (0.028)	-0.043 (0.029)	-0.043 (0.029)	-0.043 (0.029)
Coefficient of Variation		0.004 (0.003)				-0.006 (0.041)		
Correlation with HH Size			-0.003 (0.002)				0.071*** (0.020)	
Moran's I @ 16m				0.105 (0.116)				-2.051 (1.378)
Moran's I @ 32m				0.022 (0.121)				-3.094 (2.644)
constant	0.002*** (0.001)	-0.001 (0.002)	0.003*** (0.001)	0.001 (0.002)	0.109*** (0.023)	0.114** (0.048)	0.085*** (0.023)	0.118** (0.048)
n	70	70	70	70	70	70	70	70
R2	0.593	0.615	0.613	0.607	0.097	0.098	0.215	0.190

note: *** p<0.01, ** p<0.05, * p<0.1