Second Stage Sampling for Conflict Areas: Methods and Implications

Kristen Himelein, Stephanie Eckman, Siobhan Murray and Johannes Bauer[1]

Abstract: The collection of survey data from war zones or other unstable security situations is vulnerable to error because conflict often limits the options for implementation. Although there are elevated risks throughout the process, we focus here on challenges to frame construction and sample selection. We explore several alternative sampling approaches considered for the second stage selection of households for a survey in Mogadishu, Somalia. The methods are evaluated on precision, the complexity of calculations, the amount of time necessary for preparatory office work and the field implementation, and ease of implementation and verification.

Unpublished manuscript prepared for the Annual Bank Conference on Africa on June 8 – 9, 2015 in Berkeley, California. Do not cite without authors' permission.

Acknowledgments: The authors would like to thank Utz Pape, from the World Bank, and Matthieu Dillais from Altai Consulting for their comments on this draft, and Hannah Mautner and Ruben Bach of IAB for their research assistance.

---

[1] Kristen Himelein is a senior economist / statistician in the Poverty Global Practice at the World Bank. Stephanie Eckman is a senior researcher at the Institute for Employment Research (IAB) in Nuremberg, Germany. Siobhan Murray is a technical specialist in the Development Economics Research Group in the World Bank. Johannes Bauer is research fellow at the Institute for Sociology, Ludwigs-Maximilians University Munich and at the Institute for Employment Research (IAB). All views are those of the authors and do not reflect the views of their employers including the World Bank or its member countries.

1. Introduction

The collection of survey data from war zones or other unstable security situations provides important insights into the socio-economic implications of conflict. Data collected during these periods, however, is vulnerable to error, because conflict often limits the options for survey implementation. For example, the traditional two-stage sample design for face-to-face surveys in most developing countries first selects census enumeration areas (EAs) with probability proportional to size and then conducts a listing operation to create a frame of households from which a sample is selected. Such an approach, however, is not always feasible in conflict areas. At the first stage, updated counts are often not available, making probability proportional to size selection inefficient. At the second stage, survey staff canvas the entire selected area, requiring interviewers to spend additional time in the field approaching all households (see Harter et al, 2010 for description of household listing procedures). This may be too dangerous. To collect high-quality data in conflict areas, new methods of selection are needed, particularly at the second stage.

This paper explores several alternative sampling approaches considered for the baseline of the Mogadishu High Frequency Survey (MHFS). The baseline was a face-to-face household survey in Mogadishu, Somalia, conducted from October to December 2014 by the World Bank team and Altai Consulting. A full listing was deemed unsafe in Mogadishu. The additional time in the field and predictable movements increased interviewers' exposure to robbery, kidnapping, and assault, and also increased the likelihood that the local militias would object to their presence. The survey needed a sample design that would minimize the time spent in the field outside of the households, but also could be implemented without expensive equipment or extensive technical training. In addition, international supervisors from the consulting firm could not go to the field, necessitating a sample design that could be verified afterwards.

The consulting firm originally proposed a random walk procedure. While this methodology had the benefits of fast implementation and unpredictability of movement, the procedure gives biased results, even if implemented under perfect conditions (Bauer, 2014), and the circumstances in Mogadishu were far from ideal. Therefore the team considered four alternatives for household selection. The first option considered was to use a satellite map (of which many high quality options exist, due the arid conditions and political importance of the region) to map each structure. Of these, 10 would be selected. The second option considered was to cut EAs into 8-10 household non-uniform segments and ask enumerators to list and choose households from the segments.[2] The third option considered was to lay a uniform grid over the EA and ask enumerators to list and choose households from selected grid boxes. The final option considered was to start at a random point in the cluster and walk in a set direction, in this case towards Mecca, until the interviewer encountered a structure.

The paper will make use of data from the Mogadishu survey and geo-referenced maps and three example EAs to explore the following questions: (1) What are the implementation concerns for each method, including the options for verification and the impact of non-household structures? (2) What are the implications in terms of precision and bias for each of the methods described above? (3) What information is needed to calculate sampling weights for each method, and is this information available?

[2] Interviewers had a selection application on their smart phones that they used whenever subsampling was needed.

The next section briefly describes the literature as it relates to the questions above. Section 3 describes the data, section 4 gives detail on the methods considered, section 5 presents the results, and section 6 offers some discussion and conclusions.

2. Literature Review

The most common method for collecting household data in sub-Saharan Africa is to use a stratified two-stage sample, with census enumeration areas selected proportional to size in the first stage and a set number of households selected with simple random sampling in the second stage (Grosh and Munoz, 1996). Since often administrative records are incomplete and most structures do not have postal addresses, as is the case in Mogadishu, a household listing operation is usually necessary prior to the second stage selection. Due, however, to the security concerns cited above, listing was not feasible in Mogadishu.

A number of alternatives for second stage selection can be used when household lists are not available. The most common alternative is a random-walk in which the probabilities of selection are considered equivalent to simple random sampling. Random walk methodologies are commonly used in Europe (see Bauer, 2014, for recent examples), but are also implemented in the developing world. Specifically, the Afrobarometer survey, which has been conducted in multiple rounds in 35 African countries since 1999, and the Gallup World Poll, which conducted surveys in 29 sub-Saharan African countries in 2012, use random walk methodologies. Bauer (2014) tests the assumption that a random walk is equivalent to simple random sampling by simulating all possible random routes using standard within a German city and calculating the probability of selection for each household. The results show substantial deviation from simple random sampling expectations which lead to systematic bias. The simulations also assume perfect implementation of the routing instructions, which is unlikely given the limited ability to conduct in-field supervision and strong (though understandable) incentives for interviewers to select respondents who are willing to participate (Alt et al 1991).

The alternative methodologies discussed here use a combination of satellite maps and area-based sampling. As satellite technology has improved in quality and become more readily available, it has been increasingly used for research in the developing world. Barry and Rüther (2001) and Turkstra and Raithelhuber (2004) use satellite imagery to study informal urban settlements in South Africa and Kenya, respectively. Aminipouri et al (2009) use samples from high resolution satellite imagery to estimate slum populations in Dar-es-Salaam, Tanzania. Examples from the United States and Europe are less common, as usually there are traditional, reliable alternatives, but Dreiling et al (2009) tested the use of satellite images for household selection in rural counties of South Dakota.

While area-based selection methods are more common in agricultural and livestock surveys, Himelein et al (2014) used circles generated around randomly generated points to survey pastoralist populations in eastern Ethiopia, with the stratification developed from satellite imagery. A variation of this method was considered in Mogadishu, but the methodology surveys all eligible respondents living within the selected circle. The resulting uncertainty over the number of total number of selected households and the time spent in the field caused it to be discarded.

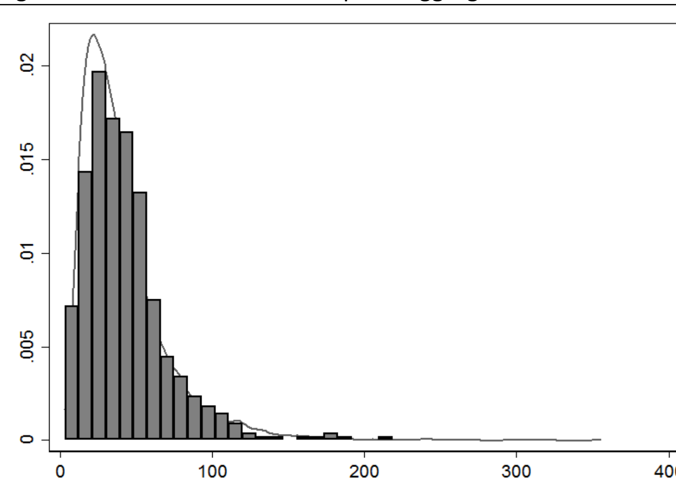
There are further examples of random point and satellite based selection in the public health literature. Grais et al (2007) also used a random point selection methodology in their study of vaccination rates in urban Niger, comparing the results to a random walk. They do not find statistically significant differences in the results from the three methods, though the sample size was limited, but conclude that interviewers found the random point selection methods most straightforward to implement than the random walk. Lowther et al (2009) uses satellite imagery to map more than 16,000 households in urban Zambia to select young children for a measles prevalence survey. They find the method straightforward to implement, but do not do a formal comparison with alternatives. Other public health studies such as the EPI studies use the "spin the pen" method to choose a starting household and then interview a tight cluster of households. This method is nonprobability (Bennett et al 1994) and was not considered for the Mogadishu study.

This paper brings together alternatives developed from this literature and applies them to a conflict environment. We take a rigorous approach using simulations and careful estimation of weights to compare the methods across a variety of potential field conditions. The results offer general guidelines for practitioners developing implementation plans for conflict settings.

3. Data and Methodology

To explore the challenges of the random walk and the four proposed alternatives, we simulated the use of each in three example PSUs from Mogadishu, Somalia. We purposefully chose three census enumeration areas as the PSUs for this exercise to illustrate the variation in physical layout present in Mogadishu. Maps of the three examples PSUs are shown in the appendix. The first is in Dharkinley district, a comparatively wealthy section of southwestern Mogadishu where the households are laid out relatively uniformly over gridded streets. The second is on the eastern edge of Heliwa district in the northeast of the city. This area is more irregular in layout with larger gaps between buildings. The third selected was in the more central Hodon district. It is densely populated with very irregularly laid out structures. (See maps in the appendix.)

Figure 1: distribution of consumption aggregates



Source: Authors' calculations based on Mogadishu High Frequency Survey data

To construct the dataset for the simulations, the values for consumption are taken from data collected by the MHFS. The survey covered both households in neighborhoods and those in internally displaced persons camps, but for the purposes of this simulation, we use only the neighborhood sample as the variation in the IDP sample is compressed due to reliance on food aid. Data was collected from the selected households on limited range of food and non-food items which we sum to calculate total consumption (see Mistiaen and Pape, forthcoming, for further details on these calculations.) There were 624 cases outside of the IDP camps with non-missing values on the two consumption measures. The distribution of consumption across these cases has a strong right skew to the distribution (Figure 1) with mean 43.0 and standard deviation 27.5. In assigning values for our simulations, we drew consumption totals from this distribution.

To simulate the variety of situations that may be found in the field, we use three different mechanisms for assigning consumption values to households in the three example PSUs. In the first, values are randomly assigned across the households in each PSU. In the second, the same values are reassigned to households to create a moderate degree of spatial clustering. In the third assignment mechanism, the spatial clustering of consumption values is more extreme. We study the ability of each of the proposed methods to estimate consumption under these three conditions. While these distributions may not mimic actual conditions, they are illustrative of the different situations encountered in the field.

For each of the combinations of methods and assignment mechanisms discussed above, 10,000 simulated samples were drawn and relevant probability weights calculated. Ten structures were selected within each PSU. In the cases of segmentation and grid point selection, where the sample was selected in two stages, two clusters were selected and then five structures within each clusters.

4. Alternative Sampling Methodologies

4.1 Satellite mapping

A full mapping of the PSU entails using satellite maps to identify the outline of each structure (see appendix). In this case, we used maps publically available on Google Earth and PSU outlines provided by the Somali Directorate of Statistics. From these maps, the structures inside each PSU can be assigned numbers and selected easily in the office with simple random sampling. Once selected, interviewers can be provided with the GPS coordinates of the household to locate it in the field. Mapping is the closest of the proposed study methods to the gold standard of a well-implemented full household listing. The main differences are that in a field listing, enumerators can exclude ineligible structures, such as uninhabited and commercial buildings, and include information not available from satellite maps, such as new construction and the identification of individual units within multi-household structures. Selection from a satellite mapping therefore requires an additional set of field protocols for addressing and documenting the above issues.

Satellite mapping can also be time consuming in terms of preparation. In the simplest approach to this method, cluster outlines can be overlaid with maps from Google Earth, maps printed, and buildings outlined and numbered by hand. A fully digitalized approach would take longer. Based on the experience mapping the three PSUs used in the paper, it takes about one minute per household to construct an outline. If the PSUs contain approximately 250 structures (the ones used here contain 68, 309, and 353 structures, respectively), mapping the 106 PSUs selected for the full Mogadishu High Frequency Survey would have required more than 50 work days.

Once the mapping has been completed, however, the calculation of the probability of selection, and by extension the survey weight, is straight-forward. The probability would be $\frac{n}{N}$, where n is the number of structures selected and *N* is the total number of structures mapped, plus any necessary adjustment for multi-household structures.

4.2 Segmenting

Segmenting is a standard field procedure of subdividing large PSUs into approximately equal sized smaller units for listing and selection purposes. The individual segments are then selected with simple random sampling, and listed by field enumerators, and households selected from these lists. Segmenting is less time consuming than a full mapping exercise in terms of office preparation, but still requires the manual demarcation of segment boundaries, which generally follow roads or other easily identifiable landmarks so that interviewers can identify the boundaries of the segments in the field. Also there would be additional time taken in the field to list the segment, which would have security implications.

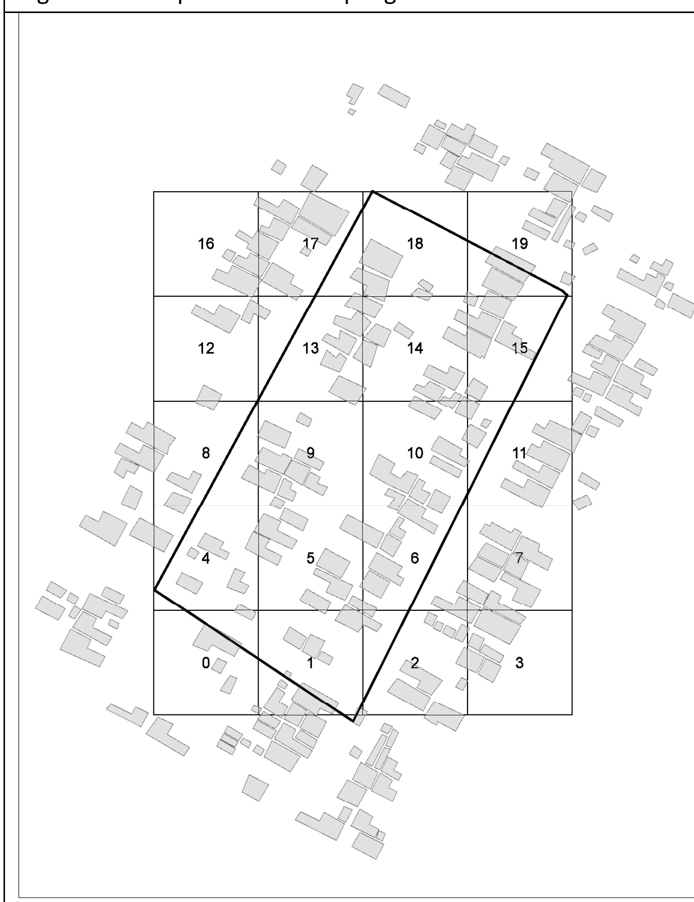
The calculation of the probability of selection is straightforward: the product of the probability of selection of the segment, and the probability of selection of the household within the segment. There are, however, implications for precision as the additional clustering introduced by the selection of segments could increase the design effect (though the magnitude would depend on the number of segments selected, the number of households selected per segment, and the degree of homogeneity within clusters for the study variables).

4.3 Grid

To implement the grid method, a uniform grid of squares (or another shape) is overlaid on the PSU map. Figure 1 shows an example using 50 x 50 meter squares for the Dharkinley PSU. The area of a grid point includes all of the area that lies both within the grid point and within the PSU boundaries. For example, in grid point 17 in figure 1, the majority of the structures would not be eligible as they lay outside of the PSU boundaries. Only the structures which lie in the bottom left corner are both within the grid and PSU boundaries.

One or more squares are selected with simple random sampling from the set of all squares that overlap the selected PSU. Depending on the survey protocols, a structure may be defined as eligible if all or part of it lies

Figure 2: Example of Grid Sampling Method



Source: Authors' diagram based on PSU boundaries and Google Earth images within the grid space. The more common protocol, including the structure if the majority lies within the grid point, has the benefit of simplifying the weight calculations, but the risk of subjective decisions made by interviewers in the field about where the majority of the building lies. Since the options for supervision and field re-verification were limited in this survey, it was decided to consider the structure as eligible if any portion of the structure lay within the grid boundaries.

To select a sample of households within the selected squares, a common approach would be for interviews to be conducted with all eligible respondents with the grid point. This could lead, however, to issues with verification as well as decreasing control over the final total sample size. Therefore, the protocol used in Mogadishu had interviewers list all households with the selected square and use the application to select households for the survey.

This variation of the grid method has the advantage that it requires less preparation time compared to mapping or segmenting. There are considerable drawbacks, however, in the ease of implementation and additional work to accurately calculate the selection probabilities. Since the squares do not follow landmarks on the ground, interviewers need to use GPS devices to find the squares' boundaries. This approach also still requires some listing work, which may have security implications depending on the size of the squares in the grid. The size can vary depending on the physical size of the PSU and the density of the population. Smaller squares require less listing work, but also mean that more buildings will lie on the boundaries between squares. Those selected structures which lie on boundary lines require additional time for field implementation, because the overlapping squares must also be listed, and also complicate the calculation of the probabilities of selection.

Let s_k be the number of squares selected in PSU_k and S_k be the number of squares that are partially or completely contained within PSU_k. For households that are entirely contained within square *j*, the probability of selection, given that PSU_k was selected, is:

$$\pi_{i|k} = \frac{n_j}{N_j} * \frac{s_k}{S_k} \qquad (1)$$

where n_j is the number of structures selected from square *j* and N_j is the total number of eligible structures in the square. $\frac{s_k}{S_k}$ is the probability of selection of the square when a simple random sample of size s_k is selected from the S_k squares in PSU_k.

If household *i* lies in both squares *j* and *j'*, the probability of selection is:

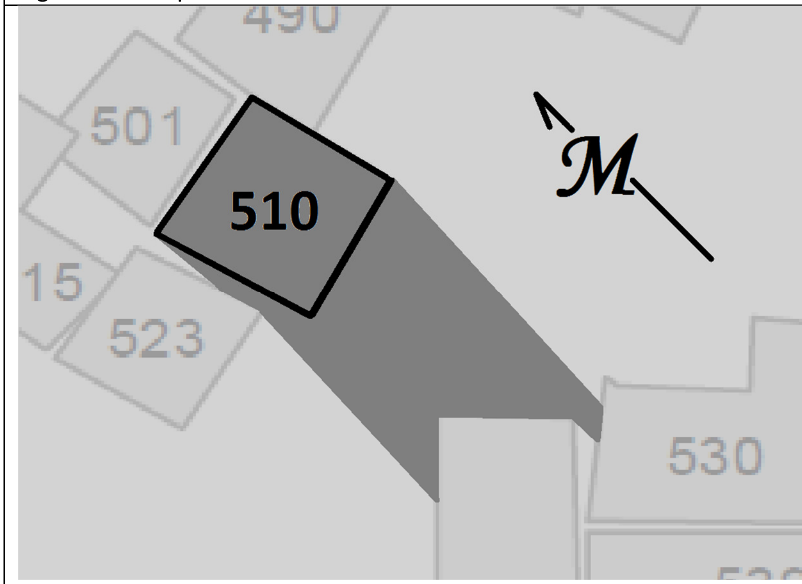
$$\pi_{i|k} = \left(\frac{n_j}{N_j} * \frac{s_k}{S_k}\right) + \left(\frac{n_{j\prime}}{N_{j\prime}} * \frac{s_k}{S_k}\right) - \left(\frac{n_j}{N_j} * \frac{n_{j\prime}}{N_{j\prime}} * \frac{s_k}{S_k} * \frac{s_k-1}{S_k-1}\right) \qquad (2)$$

In an extreme case of a structure lying on a four way intersection, there would be additional terms in equation (2). Interviewers would also have to spend significant time on additional listing, which greatly increases exposure in the field and provides disincentives to interviewers to report such households.

4.4 *Mecca Method*

This sampling approach involves selecting multiple random locations within each PSU and travelling from those points in a given direction until a structure is found. In Somalia, the consulting firm suggested using the direction of Mecca, since it is common for interviewers to have an application on their cell phones which shows this direction. If that structure is a household, the interview is done with the household. We are not aware of any surveys that have used this approach, but it clearly has intuitive appeal, as it is straightforward and inexpensive to implement in the field.

Figure 3: Example of Mecca Method



Source: Authors' diagram based on PSU boundaries and Google Earth images

Figure 3 gives a stylized example of this method. Household 510 will be selected whenever any of the points in the shaded region are selected. This region includes the area of the dwelling itself (its roof) and all points in its "shadow" – that is, all land inside the PSU that lies in the direction opposite Mecca, excluding points that lead to the selection of other buildings.

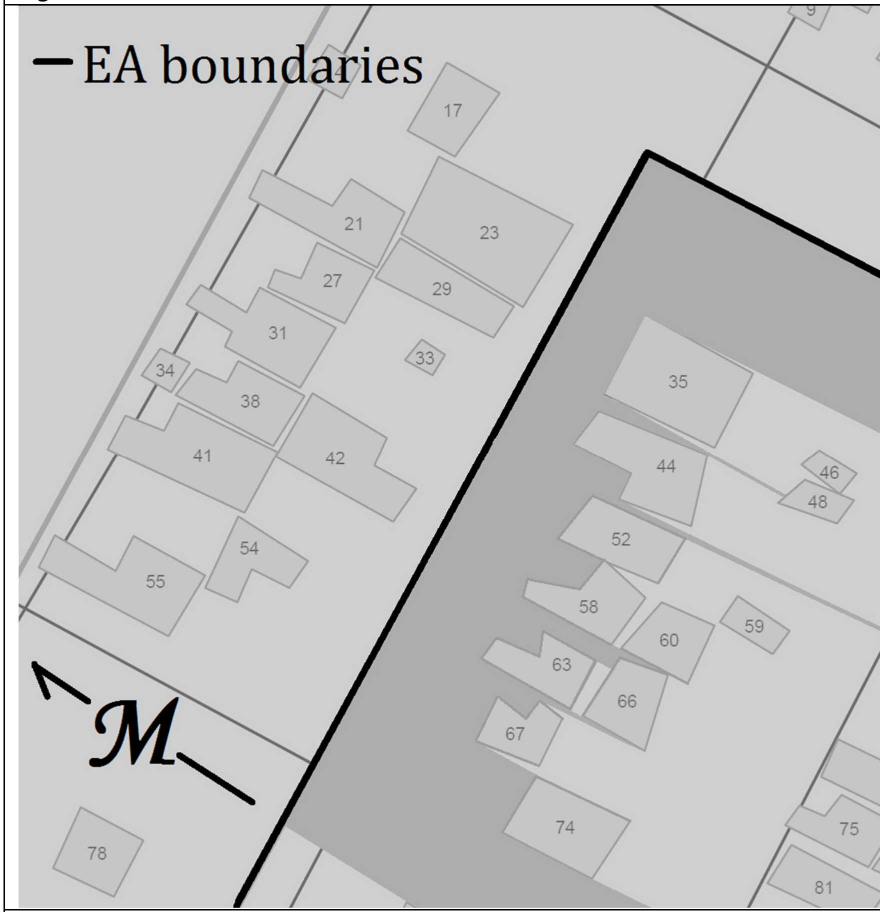
Despite its seeming ease-of-use, this approach contains many challenges. For one, it is not clear how non-residential structures should be handled. The interviewer could walk around business and vacant housing units, continuing in the direction of Mecca until she finds a residential unit. This approach would work in theory, but in addition to the difficulties in remote verification it would create, it would also complicate the calculation of probabilities of selection (discussed below). Therefore we do not suggest it. Instead, we suggest coding points that lead to non-household selections as out-of-scope, and selecting additional points to replace them.

Perhaps the biggest challenge with this method is the collection of the information needed to calculate probabilities of selection of the selected households. Figure 3 shows Household 510 and, in the shaded area, the set of all points that lead to the selection of this household. Each household, i, in the PSU has an associated selection region: call this region A_i. The probability of selection of household i (conditional on selection of PSU k), if c points in the PSU are selected, is one minus the probability that all c selected points are not in A_i:

$$\pi_{i|k} = 1 - \left(1 - \frac{Area\ of\ A_i}{total\ area\ of\ PSU_k}\right)^c \qquad (3)$$

(based on Särndal et al. 1992, p.50). This approach is essentially probability proportional to size selection with replacement, where the measure of size is the area of A_i. The weight is then the inverse.

Figure 4: Area which does not lead to selection of household



Source: Authors' diagram based on PSU boundaries and Google Earth images

Looking at Equation 3, the hardest quantity to calculate is the area of A_i. For the purposes of this paper, we use relatively recent Google Earth maps (two of which were from March 2014 and one from December 2013) to calculate the A_i region for each selected household. If high quality and recent satellite photos of the survey region are not available, calculation of the area of the selection regions will be much harder and may not be possible at all. Any structures added since the imagery was captured would not be included and therefore it would be difficult to calculate the area. Similarly, if new structures were built in the shadow, these areas would be incorrectly still included in A_i.

As it is likely that there will be incidences where field maps are either too dated to be useful or completely unavailable, we also consider two alternatives: using weights based on the estimated distance to the next structure in the opposite direction of Mecca, and ignoring the weights completely. Neither method can provide unbiased results, but under certain conditions they may be a good alternative for those that find themselves in second or third best scenarios. The simple distance would be unbiased if the length of the line was exactly proportional to the area of the shadow. While this specific condition is unlikely, the variation does have the benefit of not requiring digitized maps and being more flexible in accounting for new construction. No weighting would be approximately unbiased if dwellings were identical in size and equidistant.

Figure 4 illustrates another potential issue with this group of methods. There are points in this PSU that would not lead to the selection of any households. Consider the points in the direction of Mecca from Household 35, for example. If any of these points were selected, the interviewer would not find any household before she left the boundaries of the PSU. This issue raises questions for the field protocols. Should interviewers stop at the PSU boundary, or should they continue and select housing units outside of the selected PSU? If the former, how would the interviewer know where the PSU boundaries are? If the latter, the probabilities should be adjusted for the fact that the A_i region extends outside of the PSU, which is not straightforward. Additional structures outside of the boundaries of the PSU would need to be mapped, requiring additional preparation time. For the purposes of this paper, we mapped all households in a 50 meter buffer zone around the PSU boundaries. This increased the number of structures required from 309 to 408, 68 to 207, and 353 to 724, respectively, nearly doubling the required mapping time. A third option would be to allow interviewers to travel outside of the PSU in search of a selected household, but then remove these interviewed households outside the selected PSUs from the data set, because their probabilities of selection are too complex to calculate. This approach has preserves the probabilities of selection and is easy for the interviewer to implement, but deleting data is inefficient in terms of cost.

4.5 Random Walk

There are many different implementations of the random walk procedure. Each method invokes choosing a starting point within the selected area and then proceeding along a path, selecting every k^{th} household. The methods differ in how the path is defined. In this paper, we follow the method used by the Afrobarometer survey. The walking instructions are:

> "Starting as near as possible to the SSP [Sampling Start Point], the FS [Field Supervisor] should choose any random point (like a street corner, a school, or a water source) being careful to randomly rotate the choice of such landmarks. From this point, the four Fieldworkers follow this Walk Pattern: Fieldworker 1 walks towards the sun, Fieldworker 2 away from the sun, Fieldworker 3 at right angles to Fieldworker 1, Fieldworker 4 in the opposite direction from Fieldworker 3…. Walking in their designated direction away from the SSP, they will select the fifth household for their first interview, counting houses on both the right and the left (and starting with those on the right if they are opposite each other). Once they leave their first interview, they will continue on in the same direction, and select the tenth household (i.e., counting off an interval of ten more households), again counting houses on both the right and the left. If the settlement comes to an end and there are no more houses, the Fieldworker should turn at right angles to the right and keep walking, continuing to count until finding the tenth dwelling" (Afrobarometer, pg. 35).

However, there are several documented problems with random walk methods. First, although the random walk methods do not necessarily produce equal probability samples, they do not collect any information with which to calculate probabilities of selection. For this reason, weights are not calculable for random walk samples; instead, the samples are analyzed as if they were equal probability. Bauer (2014) shows that this assumption is not correct (though the methodology differs from the one described above). Second, the method is difficult to verify. If a supervisor has GPS tracks from each interviewer, he or she can perhaps verify that the interviewers' direction of travel was correct, but cannot be sure that the rules about which households to select were implemented correctly. Even if implemented according to protocols, two interviewers starting at the same point and traveling on the same path may select different samples depending on the distance that they consider close enough to be included or in what sequence they count the dwellings. Finally, interviewers using random walk tend to select people who are at home, rather than those who live in the households specified by the rules (Alt et al 1991).

To simulate the random walk in the Mogadishu context, we replicate the Afrobarometer protocols to the extent possible. First a random point is selected. Since it is not possible to identify landmarks with the level of detail available on the maps, the randomly selected point is taken as the starting point. To simulate the direction of the sun, as random angle is chosen and the direction of the interviewer's path assigned at 90 degree intervals. For example, if 13 degrees from due north was selected, then the four paths would be at 13 degrees, 103 degrees, 193 degrees, and 283 degrees. From these lines, it was assumed that every dwelling within 5 meters on either side of the direction of walking was within the interviewer's line of sight. These dwellings were sequentially numbered and every fifth dwelling selected. If the interviewer reached the PSU boundary before selecting the requisite number of households, the path made a 90 degree turn and continued. If each of the four interviewers selected three households, the total cluster size would be 12. In order to ensure comparability with the other methods, each of whch aimed to select ten households, we dropped the last two selected households.[3] (See illustrations in appendix for further detail.)

5. Results

5.1 Simulations

For each of the methods discussed above and the three different methods of allocating consumption values to households (random, some spatial clustering, extreme clustering), we simulated 10,000 samples and calculated a mean for each one. We report the mean and standard deviation[4] of the distribution across all 10,000 samples and evaluate the different sampling approaches in terms of their bias and variance. If a sampling method is unbiased, the expected value of the sample means should be 40, the true mean consumption in each PSU. However, because our samples are quite small (only 10 cases at the most) and the underlying distribution is far from normal (see Figure 1), we should not expect that all methods will appear unbiased.

While generally it was possible to implement all of the methods in our simulations, there were notable challenges with two of the designs. In simulating the Mecca method, certain selected points did not lead to a selection within the EA. The impact was largely negligible in Heliwa or Hodon, where only 0.4 percent and 1.4 percent, respectively, of the total area led to no selection, but in Dharkinley, the smallest and most regular of the PSUs tested, 13 percent of the area led to no selection. In implementing the grid selection method, there was little control over the number of households in each grid point. In some cases, grid points were empty or did not have the minimum number of structures to achieve the expected sample size. In the most extreme case of the large and sparsely populated PSU of Heliwa, when 50 x 50 m grid points were used, 22 contained no structures. Of those remaining, a further 17 percent had less than the necessary five structures. Therefore the grid points were combined into 100 x 100 m squares. After combination, none of the larger grid points were empty, but 16 percent continued to contain less than the minimum. For the simulations, we dropped grid points without households, though this would likely not be possible in true field implementation, leading to cost inefficiencies.

[3] The analogous action in the field would be for the supervisor to rotate the additional interviews between interviewers to assure an even workload, though most likely in the design stage the cluster size would have been set to be evenly divisible among interviewers.
[4] The standard deviation of the distribution is the standard error of the estimate of the mean.

5.2 Bias and Variance

The mean, standard deviation, and coefficient of variation are shown in Table 1 for the seven methods and three PSUs under the three different consumption values. From this table we can evaluate how well each method worked in terms of bias and variance. From a true mean of 40, it was unsurprising that the full listing / satellite mapping method showed the most consistently unbiased results across the nine scenarios. Segmentation also showed consistently unbiased results. As expected, however, the segmentation method is more sensitive to clustering in the underlying distribution of the consumption values because the homogeneity within the segments.

The Mecca method also generally performed well when the results were weighted. The Mecca method with full weights calculated from the areas of the shadow showed an average difference of 1.3 from the true mean of 40 across the nine scenarios, in comparison to the 0.0 for full listing and 0.3 for segmentation. The method should be unbiased, as it is a probability method using appropriate weights, but the small sample size and skewed distribution of the underlying variable, as discussed in Section 4.6. The Mecca method with proxy weights, based on the distance from the selected point to the household, also performed well with an average difference of 1.5. As expected, the Mecca method without any weighting showed the most biased results of the methods considered, even when there was no clustering of wealth. It performed particularly poorly in Heliwa, where the sparse geography leads to long shadows and large differentials in the probabilities of selection.

The two remaining methods, the grid and random walk, showed more bias than the full listing, segmentation, and two of the three Mecca methods, but were better than the Mecca method with no weights. Both performed relatively well when the wealth values were randomly assigned, though the grid method was slightly better in these situations. The two methods also performed better in the more uniform Dharkinley and sparse Heliwa compared to the more chaotic Hodon PSU.

6.2 Implementation Issues

The final criteria on which the methods were evaluated was the ease of field implementation and remote supervision. The most straightforward to supervise remotely is the satellite mapping. Since the selection is done in the office, the interviewers can be sent to the field with their target locations loaded on a handheld GPS device, and supervisors can use tracks from the device to verify that interviewers visited the correct households. While there is still scope for cheating, such as going to the target location to record the GPS point and then interviewing another household, or claiming a refusal because the household appeared to be difficult or time-consuming, these behaviors are possible in all surveys, regardless of method, and can only be addressed through training. The Mecca method also provides the possibility for remote supervision by using GPS waypoints and tracks to let supervisors and central office staff verify that the interviewer travelled in the correct direction and interviewed households within the boundary of the select PSU (see Himelein et al, 2014, for an example of using interviewer tracks for supervision).

The grid and segmentation methods are more difficult to supervise because they are also more difficult for the interviewers to implement. When creating segments, best practice is to use clearly discernable landmarks used to draw boundaries, but these can change over time or not be correctly identified by the interviewers. If the interviewer incorrectly identifies the segment, it may be necessary to exclude the resulting data as it cannot be properly weighted.

The grid method provides additional challenges as the boundaries between the grid squares do not follow existing landmarks. The boundaries must therefore be programmed into the GPS and identified by the interviewers. As it is unlikely that they will be able to walk straight along the boundary, additional training may be required to correctly identify eligible structures. In addition, as was the case in Heliwa, it may be necessary to have large grid points in sparsely populated areas. This increases the time necessary to do the listing, exposing the interviewer to increased danger that may result in the inability to complete interviews.

With both the grid and segmentation method, additional verification can be done for the listing portion by using satellite maps. If the interviewer enters fewer households into the selection tool than are expected based on the satellite map, it could be that they are purposely excluding or shirking. If the number is much higher without evidence of multi-household structures, it may be that the interviewers have gone to the wrong location, and this can be confirmed with the target household waypoint.

5.3 Replacements

One remaining potential issue not yet addressed relates to the use of non-response due to either refusals or are out of sample selections. Due to high transportation costs, most surveys in the developing world use replacements. This is done either through selecting additional households from PSU lists, as is recommended in the World Bank's Living Standards Measurement Study (Grosh and Munoz, 1996), or selecting a neighboring structure based on field protocols, such as selecting the dwelling immediately to the right (Lowther et al, 2009). While replacements for out of sample selections with new random points does not introduce bias, it is inefficient and increases costs. For non-response due to refusal, it is likely to be non-random, and therefore replacements will create at least some degree of bias in the data. The reason and method for the replacement may influence the degree. If refusals tend to come from the highest and lowest wealth households, as the opportunity cost of their time is high, and replacements come from the main part of the distribution, the use of replacements will attenuate the variation in the sample. This may cause the results to underestimate measures such as inequality that depend on accurately capturing the extremes of the distribution. When using a replacement method that uses near neighbors, if structures are abandoned or commercial buildings, those households living adjacent may be systematically different from the remainder of the PSU. In addition, those households near the boundary of the PSU would have a lower probability of selection since there are fewer households near them that would lead to them being selected as replacements.

Of the methods discussed above, segmenting and gridding require a short listing exercise at which time non-eligible structures can be excluded. Satellite mapping and the Mecca method rely on maps that cannot differentiate based on eligibility, and are therefore more vulnerable to issues with out of sample selections. In addition, regardless of method, the survey protocols should address procedures for the inevitable refusals, which may be more likely in conflict areas.

6. Discussion

Overall, each of the methods above could prove to be the best option for second stage sampling in a conflict zone depending on the context of the survey. Satellite mapping, segmentation, and the Mecca methods with full area weights are all probability methods for which it is possible (though perhaps not easy) to calculate weights, and thus all should produce unbiased estimates of the population mean.[5] In addition, the Mecca method with simple distance weights is a close approximation of an unbiased sample. The choice between the different methods is really one of cost and variance, as well as issues specific to the survey area. If there are no restrictions on time and back office resources, a full listing yields the most consistently unbiased and efficient design, provided recent maps are available and potential issues with out-of-sample buildings can be adequately addressed. The Mecca method provides promising results in simulations but has yet to be tried in the field. The simple distance variation of the Mecca method shows particular promise as it removes the requirement of updated satellite maps and greatly reduces the calculation burden for the weights. The non-probability methods, random walk and the unweighted Mecca method, do not produce unbiased results. Random walk, in particular, did not perform well in the simulations despite being common practice for many surveys. Given the expanding availability of satellite maps and decreasing costs of GPS technology, much of which is integrated into the phones and tablets used by interviewers, alternative methods based on probability sampling may reduce bias with little impact on cost or complexity of implementation.

In the case specifically discussed here, the Mogadishu High Frequency Survey, the team opted to use segmentation as a compromise between preparation time, ease of implementation, and the time and complexity necessary for the weight calculations. The survey was generally successfully fielded, but the team encountered a number of difficulties in the field. Teams occasionally faced high-level security threats and exploitative rent-seeking from local leadership. The complexity of the survey protocols, including the sampling design, slowed the implementation of the survey. Also a substantial number of observations had to be discarded because the interviewed points did not fall within the boundaries of the selected segments. Regardless of these challenges, however, it was possible implement a complex and yet rapid, high-quality survey in one of the most challenging urban contexts known to date.

[5] In this particular study, we saw small deviations from unbiasedness due to the small sample sizes used other issues discussed above.

References

Afrobarometer Network, 2014. Afrobarometer Round 6 Survey Manual.

Alt, C., Bien, W., Krebs, D., 1991. Wie zuverlässig ist die Verwirklichung von Stichprobenverfahren? Random route versus Einwohnermeldeamtsstichprobe. ZUMA-Nachrichten 28, 65–72.

Aminipouri, M., Sliuzas, R., Kuffer, M., 2009. Object-oriented analysis of very high resolution orthophotos for estimating the population of slum areas, case of Dar-Es-Salaam, Tanzania, in: Proc. ISPRS XXXVIII Conf. pp. 1–6.

Barry, M., Rüther, H., 2001. Data collection and management for informal settlement upgrades, in: Proc. International Conference on Spatial Information for Sustainable Development. Citeseer.

Bauer, J.J., 2014. Selection Errors of Random Route Samples. Sociological Methods & Research 0049124114521150. doi:10.1177/0049124114521150

Bennett, A., Radalowicz, A., Vella, V., Tomkins, A., 1994. A Computer Simulation of Household Sampling Schemes for Health Surveys in Developing Countries. International Journal of Epidemiology 23,6: 1282-1291.

Dreiling, K., Trushenski, S., Kayongo-Male, D., & Specker, B. 2009. Comparing household listing techniques in a rural Midwestern vanguard center of the national children's study. Public Health Nursing, 26(2), 192-201.

Eckman, S., Himelein, K., Dever, J., forthcoming. New Ideas in Sampling for Surveys in the Developing World, in: Johnson, T.P., Pennell, B.-E., Stoop, I., Dorer, B. (Eds.), Advances in Comparative Survey Methodology. 3MC.

Gallup, 2014. Farm workers pessimistic about their lives [WWW Document]. URL http://www.gallup.com/poll/169019/farm-workers-africa-pessimistic-lives.aspx (accessed 1.27.15).

Gallup, 2015. World Poll Methodology [WWW Document]. URL http://www.gallup.com/poll/105226/world-poll-methodology.aspx (accessed 1.27.15).

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.

Harter, R., Eckman, S., English, N., O'Muircheartaigh, C., 2010. Applied sampling for large-scale multi-stage area probability designs. Handbook of survey research 2, 169–199.

Himelein, K., Eckman, S., Murray, S., 2014. Sampling Nomads: A New Technique for Remote, Hard-to-Reach, and Mobile Populations. Journal of Official Statistics 30.

Lowther, S.A., Curriero, F.C., Shields, T., Ahmed, S., Monze, M., Moss, W.J., 2009. Feasibility of satellite image-based sampling for a health survey among urban townships of Lusaka, Zambia. Tropical Medicine & International Health 14, 70–78. doi:10.1111/j.1365-3156.2008.02185.x

Mneimneh, Z.N., Axinn, W.G., Ghimire, D., Cibelli, K.L., Alkaisy, M.S., 2014. Conducting surveys in areas of armed conflict, in: Hard-to-Survey Populations. Cambridge University Press.

Särdnal, C.E., Swensson, B., Wretman, J.H., 1992. Model assisted survey sampling. Springer.

Turkstra, J., Raithelhuber, M., 2004. Urban Slum Monitoring.

Table 1: Main Results

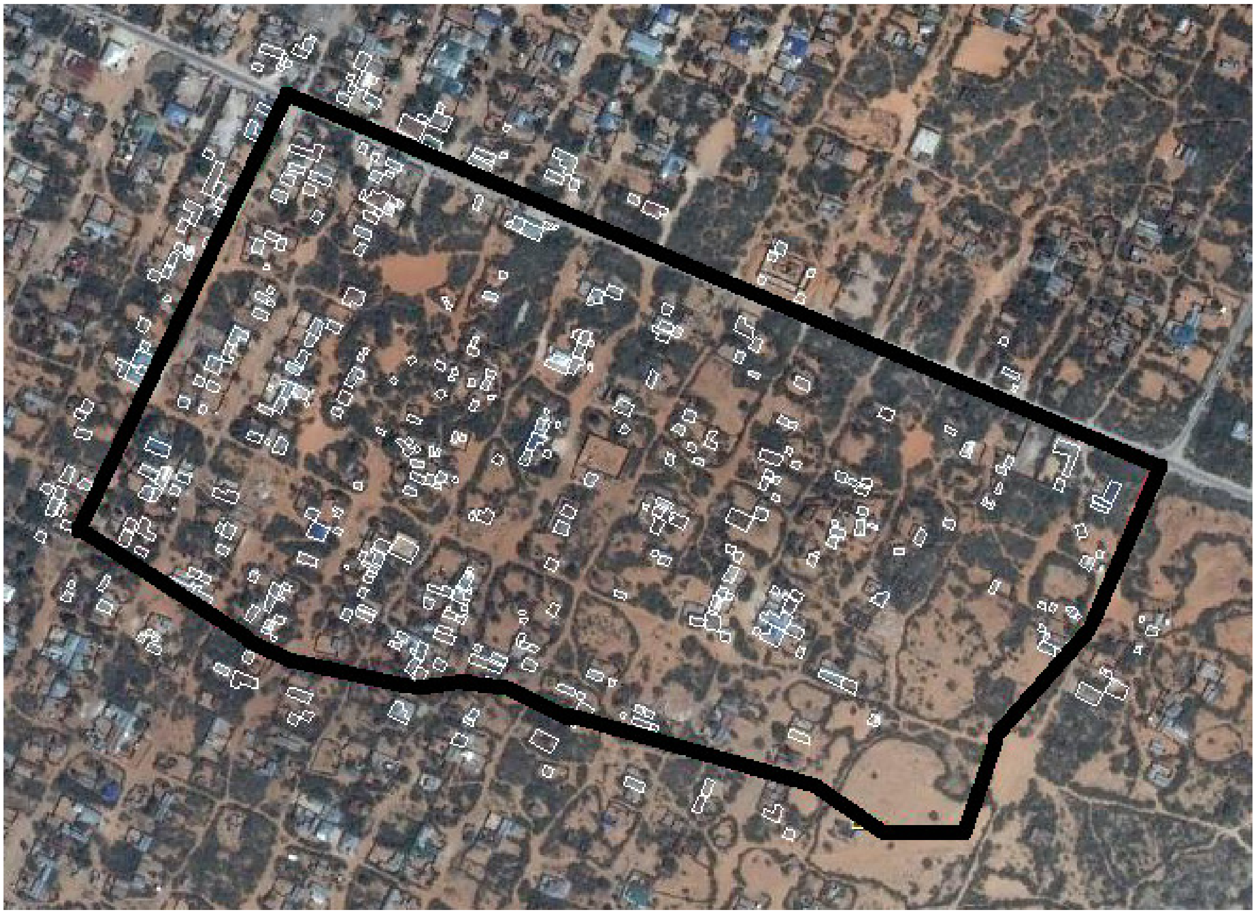
Method/Clustering	Dharkinley (EA4)			Heliwa (EA5)			Hodon (EA6)		
	Mean	*std error*	*cv*	*mean*	*std error*	*cv*	*mean*	*std error*	*cv*
True Mean	40.0			40.2			40.0		
Full Listing / Satellite Mapping									
Randomly assigned	40.0	11.7	0.29	40.0	9.6	0.24	40.0	9.5	0.24
Some spatial clustering	40.0	11.6	0.29	40.0	9.6	0.24	40.0	9.6	0.24
Extreme clustering	40.1	11.7	0.29	40.0	9.6	0.24	40.0	9.5	0.24
Mecca method (accurate weights)									
Randomly assigned	40.9	10.4	0.26	39.7	13.7	0.34	39.3	14.1	0.36
Some spatial clustering	39.4	15.7	0.40	37.6	15.1	0.40	41.5	13.3	0.32
Extreme clustering	37.8	16.7	0.44	37.8	14.5	0.38	41.4	13.8	0.33
Mecca method (proxy weights)									
Randomly assigned	41.2	12.5	0.30	39.5	14.7	0.37	39.3	14.9	0.38
Some spatial clustering	39.4	17.8	0.45	37.3	15.7	0.42	41.7	14.5	0.35
Extreme clustering	37.7	17.2	0.46	37.6	15.2	0.40	41.7	15.0	0.36
Mecca method (no weights)									
Randomly assigned	49.4	14.1	0.29	37.7	8.8	0.23	38.0	8.1	0.21
Some spatial clustering	37.1	9.5	0.26	33.4	7.0	0.21	45.5	9.2	0.20
Extreme clustering	35.8	8.1	0.23	34.1	7.9	0.23	44.5	8.4	0.19
Segmentation									
Randomly assigned	40.2	11.5	0.29	40.0	10.2	0.25	40.1	10.8	0.27
Some spatial clustering	40.9	17.3	0.42	40.0	17.3	0.43	39.9	17.0	0.43
Extreme clustering	40.9	17.7	0.43	40.0	19.2	0.48	40.0	19.3	0.48
Grid									
Randomly assigned	40.2	13.2	0.33	40.7	11.5	0.28	39.7	11.0	0.28
Some spatial clustering	38.1	20.5	0.54	40.9	18.6	0.46	45.3	20.1	0.44
Extreme clustering	38.8	24.9	0.64	42.6	24.0	0.56	45.1	22.6	0.50
Random walk									
Randomly assigned	38.4	11.5	0.30	39.3	9.1	0.23	39.5	9.1	0.23
Some spatial clustering	39.2	11.8	0.30	40.8	12.0	0.29	45.8	17.7	0.39
Extreme clustering	38.8	10.5	0.27	39.2	10.3	0.26	46.2	20.9	0.45

Appendix

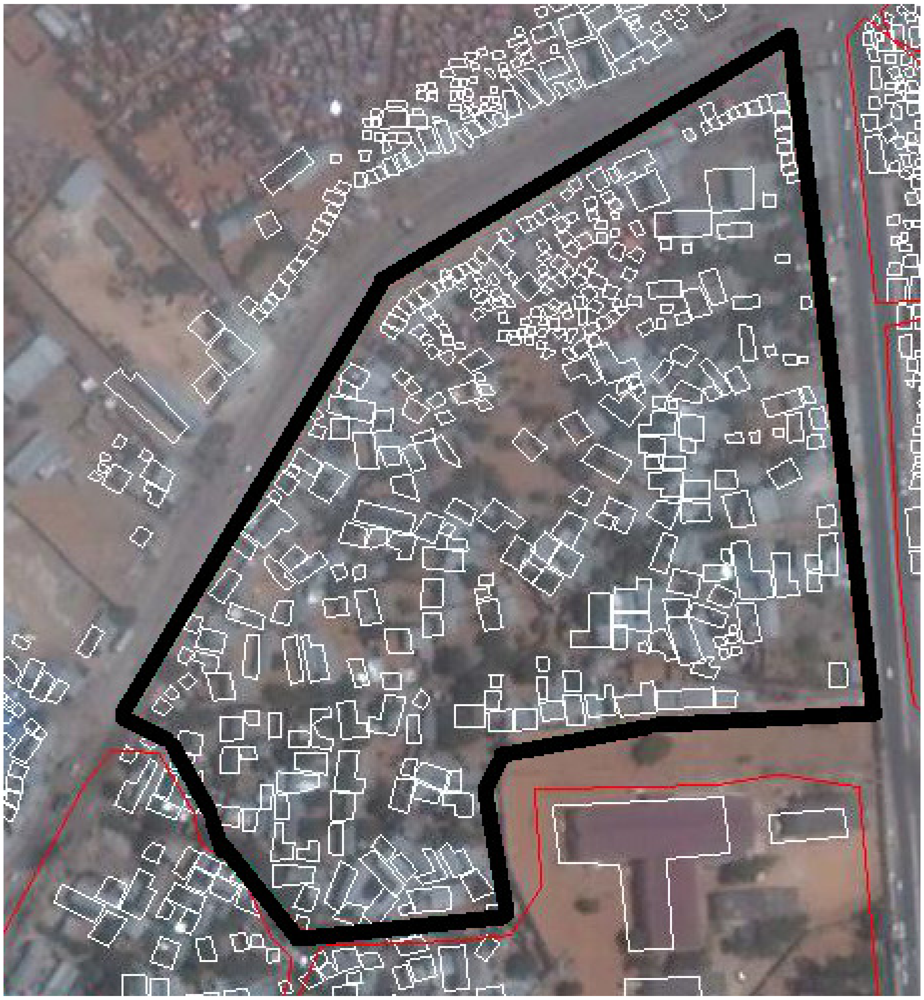
Table A1 : Description of sample PSUs						
Location	Total PSU Area (m^2)	Total PSU + Buffer Area (m^2)	Area in which no households would be selected with Mecca method (% of total)	Number of Structures	Number of Structures (including buffer)	Imagery date
Hodon	42,615	95,707	1.4%	309	408	March 14, 2014
Dharkinley	24,390	65,447	13.0%	68	207	December 25, 2013
Heliwa	345,157	477,252	0.4%	353	724	March 14, 2014

Dharkinley

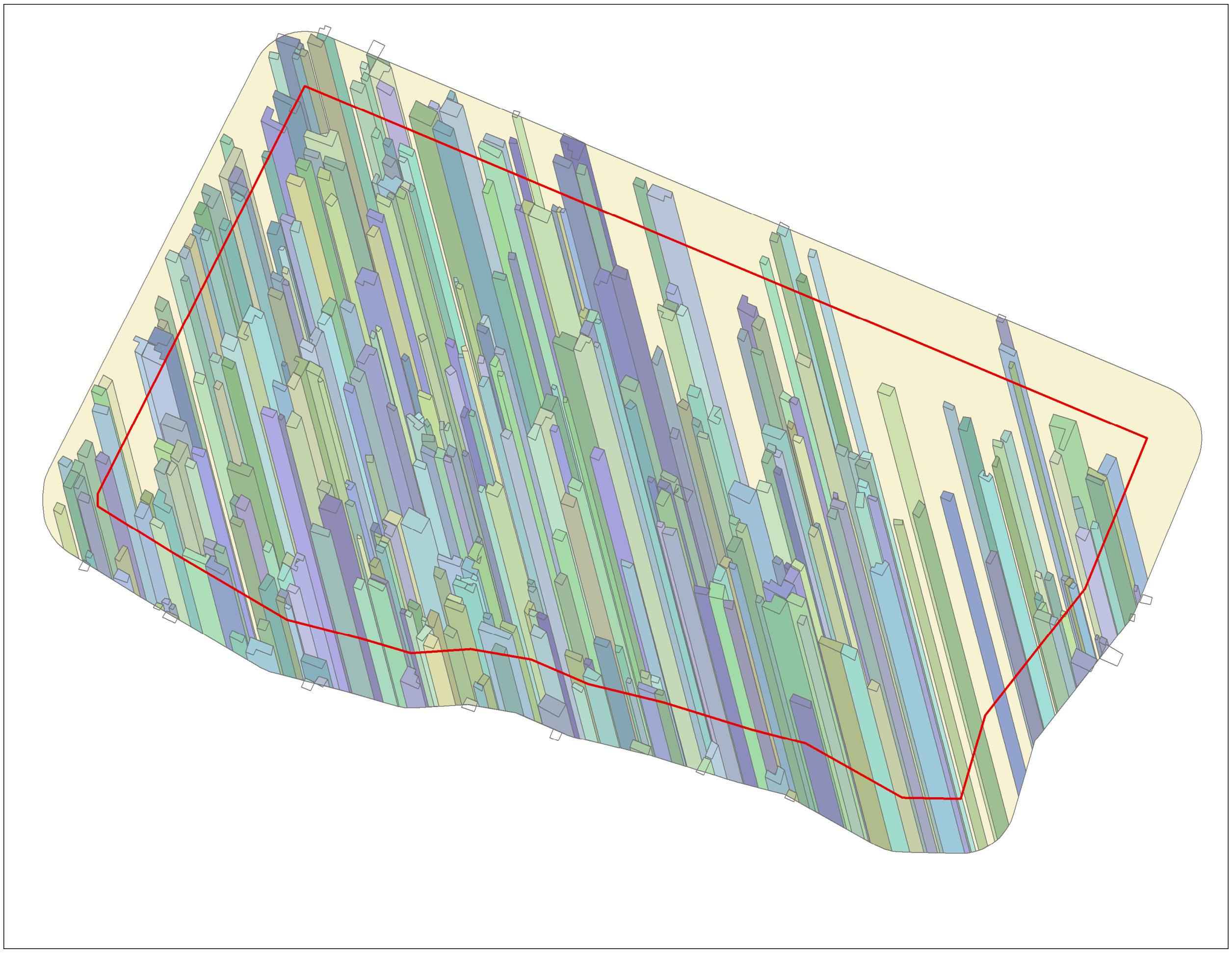
Heliwa



Hodon



Example of “shadows” for Mecca method



Example of path of random walk