The gridded population sampling methodology in Tunisia

Bayram Samet ŞAHİN[1], Hakan DEMİRBÜKEN[2]
M. Akif BAKIR [3]
[1]Hacettepe University, Institute of Population Studies, PhD Student, Social Research Methodology, Ankara, Turkey
[2]UDA Consulting, Project Director, Ankara, Turkey
[3]Gazi University, Department of Statistics, Ankara, Turkey

Abstract

Spatial sampling methods have been widely employed for household surveys especially in conflict regions. However, recently, these methods are also being used for household surveys where there is limited information available about the target population which would serve as sampling frame.

In the context of many African countries, administrative records are incomplete, and household or person register systems do not exist. At the same time, full household listing procedures are costly and can pose security risks. Therefore, survey statisticians explore alternative methods using a combination of satellite maps and area-based sampling in these situations. There are many alternative methods which integrate satellite images, listing and gridding of PSUs. In this study, a sampling methodology for a poverty survey conducted in Tunisia is presented. The methodology is based on gridded population sampling methodology where a uniform grid was overlaid on the city map and PSUs were selected from each stratum. Within each selected PSUs, randomly selected households and SSUs were interviewed.

Overall, the employed method had several strengths from practical aspect: it reduces the workload of field team, minimizes their discretion in sample selection, minimize enumerators' bias and allows for random selection with known probabilities. However, there are some weakness of the methodology such as the satellite images may be difficult to update, demarcation of boundaries is not easy to handle, and overlapping buildings into different grids.

Key Words: Spatial sampling, household survey, Sampling weights, GIS for sampling, Grid sampling

[1] This is another footnote.

Introduction

Household survey data are collected by governments, international organizations, and companies to support evidence-based policy development, the prioritization of national and international issues, allocation of funds and for tracking progress against policy and program goals. While households surveys are conducted in all countries to disseminate statistics in various fields, in countries where census data or other forms of official data are outdated, incomplete or inaccurate, households surveys are utmost important. The design of household surveys requires up-to-date population counts. However, in most developing countries, census data are not available, are outdated or are believed to be inaccurate. In those countries, researchers have begun to develop innovative sampling methods to address these challenges. One such method is gridded population sampling. In this method, gridded population datasets are considered as an alternative sampling frame.

This article considers the use of an alternative approach to collecting data from households – gridded population sampling. Similar methods are commonly used by developed countries and international organizations such as World Bank. In a study conducted by Thomson (2017), the 2010 Rwanda Demographic Health Survey was replicated by employing gridded population sampling. The WorldPop 2010 UN-adjusted 100m x 100m gridded population dataset was used as a sampling frame. It was found that gridded population sampling was a good alternative to a traditional household survey when census data was outdated or inaccurate.

This study stems from our experiences in administering a poverty survey of 960 households in Tunisia in 2019. The household survey was conducted by UDA Consulting (Turkey).

This article is organized into five sections. After a brief introduction in the first section above, the second section presents the background development of spatial sampling methods. The third section describes the methodology used for sample selection. The fourth section discusses the implementation strategy, which is followed by a discussion and conclusions in the final section.

Background of survey

As poverty in Tunisia has decreased by half since the turn of the century, inequality between regions has increased. The hinterlands of North West (NW) and Central West (CW) Tunisia have a poverty rate roughly three times greater than the rate found in coastal regions, and these areas are home to half of the country's poor. Despite claiming 50 percent of Tunisia's agricultural land and 82 percent of its forests, income levels in these two regions have remained low.

Unsustainable land management practices and discriminatory policies have inhibited profitable exploitation of resources. Limited incentives and a lack of opportunities for resource development, or local co-management, constrain the development of high-return activities and employment opportunities, and restrict forest village access to subsistence activities.

These lagging regions cover a mixed landscape including forests and grassland. Land degradation, largely due to over-grazing, deforestation, and other poor land management activities, has posed a challenge. Such degradation worsens the economic and environmental challenges faced by local populations, and further threatens their livelihoods. Land restoration and sustainable management are necessary to restore the productivity of the landscape. Better land use practices can make agriculture more productive and resilient and boost the incomes of agrarian and forest communities in Tunisia.

The objectives of survey were (i) to measure the poverty of households robustly in the areas covered in the project scope using the SWIFT methodology developed by the World Bank Group to fill persistent data gaps about poverty measurement; (ii) to collect all determinants to later measure forest income with the SWIFT methodology; and (iii) to provide monitoring and evaluation tools to the World Bank Group.

The survey was designed to be quick to administer in households. One enumerator was tasked with surveying the household head or other knowledgeable household member over the age of 18 years. The questions in the survey were close ended in order to minimize the time to collect the information.

Literature review of spatial sampling

Using a population list (census) for sampling is the gold standard for obtaining representative, probabilistic samples from populations. However, designing and implementing social surveys in low- and middle-income countries is often challenging. In these settings, there is often no sampling frame or reliable frame that contains registers of residents for various reasons. In some cases, even if registers are available, they are not be up to date. In other cases, existing registers are known to be biased and, therefore, not useful for covering the target population.

The absence of sampling frames, or the existence of incomplete or biased sampling frames, is a major obstacle for probability-based sampling designs. To overcome this challenge, researchers have started to seek innovative sampling designs using GIS technologies. The fast development in GIS technology has provided opportunities for constructing an affective spatial sampling design. Researchers are developing innovative hybrid methods, such as combined satellite image interpretation and gridded population sampling, that allows for probability-based survey samples at a reasonable cost.

Spatial sampling methods have traditionally been deployed for natural phenomena such as assessing plant, soil type and mineral composition (Amblard-Gross et al. 2004; Dessard and Bar-Hen 2005; Di Zio et al. 2004; Fleishman et al. 2005). However, using spatial sampling for samples of human beings is a relatively new methodology (Arbia and Lafratta 2002).

The spatially distributed human population and its socio-economic characteristics show significant spatial patterns of segregation and a slower pace of change. This means that people with similar socio-economic characteristics tend to live closer to each other and the change in the spatial patterns of their socio-economic activities is relatively slow (Osborne and Rose 1999). With an increasing interest in the role of location in social science theories, having special contexts of social data and hence spatial sampling is becoming increasingly important. Recent literature suggests that the local/neighbourhood environment can greatly influence socioeconomic, demographic, economic, and health outcomes (Cerin et al. 2006; Gordon-Larsen et al. 2006; Liao et al. 2006; Schwartz 1996).

Spatial sampling methods differ from non-spatial sampling methods in terms of the units of the sampling frame. Since a sample is selected based on geographic locations in spatial sampling, the first step is to construct a frame of a finite population of identifiable geographic units such as grids, enumeration areas, or segments. Typically, this is performed by overlaying a geometric pattern onto the geographic area of interest. This procedure generates a sampling frame consisting of geographical units partitioning the geographic area into a finite number of identifiable units.

In the next step, one of the four classic methods of sampling (simple random, systematic random, stratified random, or cluster random sampling) or a combination of two or more methods can be employed to select sampling sites of n from the population with set of N units (Cressie 1990).

The methods of traditional spatial sampling adopted in natural sciences cannot be directly used for sampling from a human population for several reasons. Firstly, most of these methods heavily rely on overlaying a regular geometric structure onto the geographic area to identify a finite number of N units and assumes homogeneity in each unit. The use of regular grids, however, does not match with the highly irregular shapes of residential areas. Therefore, a regular grid design is of little use for constructing a sampling frame of households. Secondly, dwellings may span across two adjacent grids, which can violate the assumption of independency between grids. Thirdly, overlaying a regular grid assigns equal probability to each unit, but the number of dwellings can vary across grids. This necessitates the use of differential probability of inclusion in the sample design. Fourthly, the major focus of traditional spatial sampling methods has often been on collecting data for only a single attribute. Household surveys, however, require data collection for many attributes at both the household and individual levels.

Because of the drawbacks mentioned above, researchers have begun to adopt a more realistic approach to spatial sampling with varying probability of selection across geographic locations (Arbia and Lafratta 2002; Cressie 1990; Lafratta 2006; Stevens and Olsen 2004). The sampling methodology of research by Lee et al. (2006) somewhat addressed above-mentioned shortcomings, and their methodology is particularly relevant for this study.

Lee et al (2006) demonstrate the use of GIS to draw a sample of respondents in an urban environment with the main goal being to understand how the built environment facilitates/hinders walking and biking. Using the parcel data, they constructed a sampling frame of built-up environment, and respondents were drawn randomly with known selection probabilities from the developed sampling frame.

Although the conceptual idea of developed sampling frames in this study is somewhat similar to that adopted by Lee et al. (2006), the proposed sampling design is unique for several reasons. First, exploiting gridded population and GIS technologies, a sampling frame was constructed from scratch. Second, satellite images were visually inspected to detect buildings/dwellings for listing. Third, it was possible to implement the household listing to update and improve population data provided by the Global Human Settlement Layer (GHSL) of the European Commission's Science and Knowledge Service Joint Research Centre (JRC), in 2010 delivers data on people's presence in the world in the form of thematic maps; namely population density and settlement maps. Finally, the integration of GIS and the data collection system allowed navigation and identification of households in the sampled grids.

Methodology

Development of gridded population estimates

Three approaches for the preparation of gridded population estimates were found in the literature: "top-down", "bottom-up", and "CDR-enhanced". Standard "top-down" gridded population datasets are made by integrating census counts or official population estimates linked to respective geographic boundaries with satellite and GIS data layers. WorldPop, LandScan, GRUMP, GHS-POP and GPWv4 are among the providers. In order to predict the population of grids, census counts of relatively large areas are combined with the known geospatial covariates for some grids and implementing spatial weighting layer created based on covariates.

On the other hand, "Bottom-up" gridded population datasets require large research collaborations and are custom-made. This method is used to estimate population totals with very high-resolution satellite imagery when census data are unavailable, outdated, and unreliable. In this method, a statistical relationship between population density and spatial covariates is used to predict population density for small areas across a whole country based on the spatial covariates alone [Tatem, AJ (2017); Vaccination Tracking System (2017)]. Known populations of some small areas are modelled as a function of spatial covariates, and the remaining grid populations are estimated with this model.

CDR-enhanced gridded datasets are made by integrating aggregated call detail records (CDRs) from mobile phone operators with "top-down" or "bottom-up" population data. CDR-enhanced gridded population datasets are new and require large research collaborations including agreements with mobile phone operators. The result is a custom-made gridded population dataset for settings where populations are highly dynamic, for example during armed conflict, or during and immediately after a natural disaster [see Deville P, et al. (2014); Erbach-Schoenberg E, et al. (2016)]. Figure 1 displays the approach of the top-down and bottom-up methods visually.

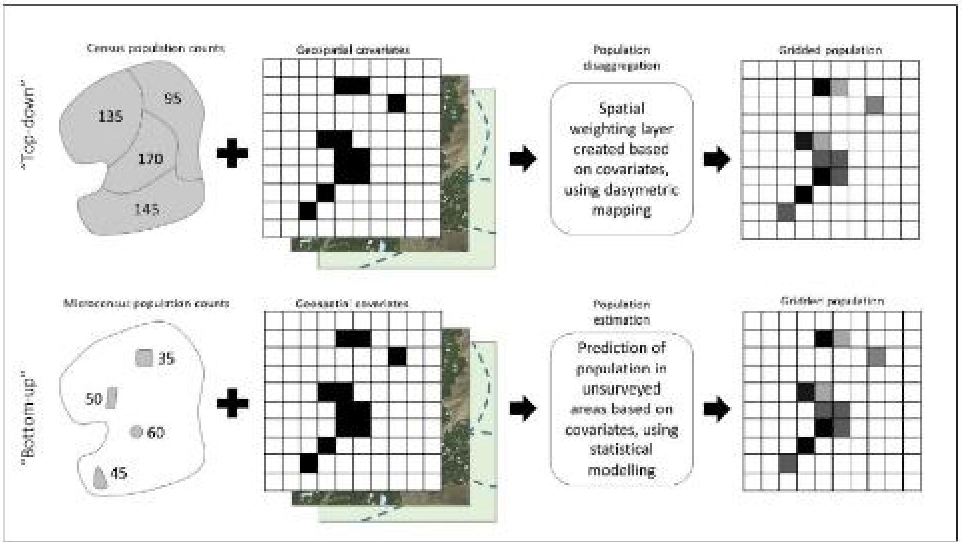

Figure 1 Methods for making gridded population data (Thomson DR, et al. (2017) reprint courtesy of Biomed Central)

For the Tunisia survey, the gridded population sampling method using a geographic information system (GIS) to partition areas of interest into logistically manageable grid cells for sampling was adopted. As enumeration areas (such as blocks of dwellings) were not available for Tunisia, GHSL gridded population data was used as a sampling frame. In the GHSL dataset, 250-meter x 250-meter grids are overlaid onto the study area. The population was predicted for each grid and the list of grids served as a sampling frame. The visualization of sampling frame is seen in the figure below.

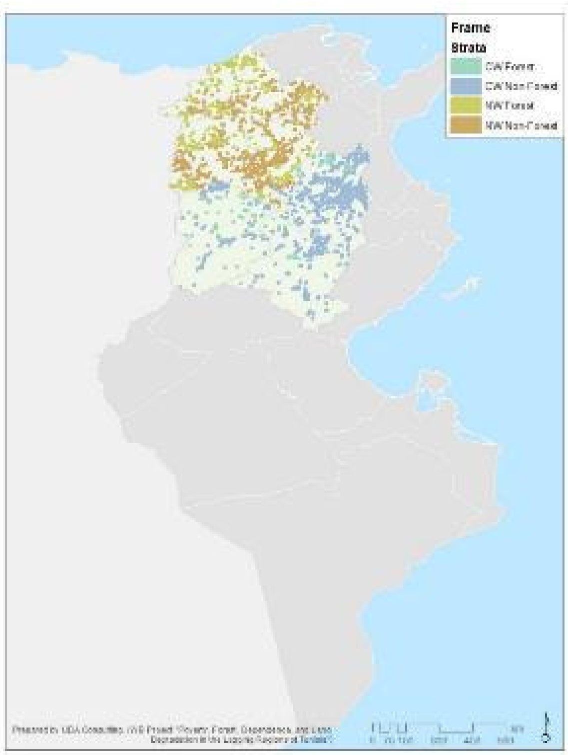

Figure 2 Sampling frame

Sampling design

The main characteristics of the sample design for the poverty survey in Tunisia can be explained as follows. The research team developed a two-stage, stratified, cluster random sampling method using geographical information system (GIS) software and global positioning system (GPS). The sampling frame was divided into four mutually exclusive strata to improve the representativeness and accuracy of the survey estimates. The stratification of the sampling frame was based on the distance of villages to forest areas. Villages were deemed to be “forest villages” if the distance from the central point of village to the nearest forest was less than 5 km. If the distance from the center of the village to the nearest forest area was more than 5 km, the villages were considered as “non-forest villages”.

Stratification was made according to two criteria. The first level of stratification, villages (households) in the rural areas of the western region, were divided into two strata: North-West (rural) and Center-West (rural). The second level of stratification was based on proximity to forests using a binary classification (forest village and non-forest village). Thus, the sampling frame is divided into 4 stratums.

Table 1 Stratification of sample

North-West (rural)
Forest
Non-Forest
Center-West (rural)
Forest
Non-Forest

Sample selection

The sample selection was performed in two stages. In the first stage, the grids were selected from stratums with probability proportional to grid’s estimated population size. Following the selection of grids, the dwellings in each grid were identified through visual inspection of satellite images, all of which were enumerated, and their coordinates were extracted and recorded in a database for listing.

In the second stage of selection, eight random households were automatically selected following the completion of a listing activity in each grid through simple random algorithms using UDA’s data collection system. Thus, a total of 120 grids out of 2,230 were randomly sampled from these four strata. The total number of sampled households and the allocation of the sample between strata was based on sample size calculations from the previously collected data from the LCMS survey of Tunisia. The map in Figure 3 presents the location of the selected grids for the household survey and listing activity.

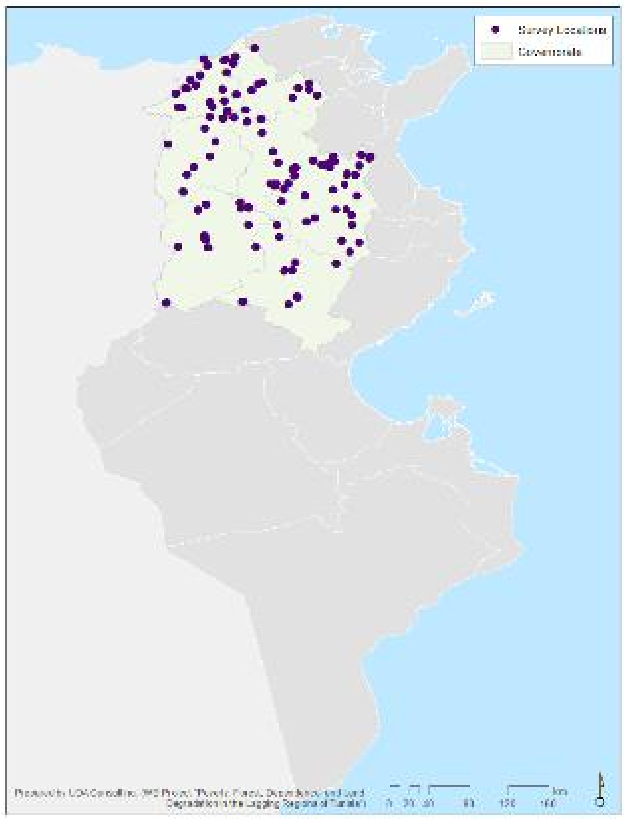

Figure 3 Selected Samples (Grids) for HH Survey and Listing

Listing

The listing activity aimed to improve the accuracy of population estimates, and to update the grid-level population counts. For this purpose, the sampled grids were first listed by listing team, which took 3 days of training on use of CAPI system and listing questionnaire. To facilitate access to dwellings and for easy reference of listing team, the coordinates of the sampled dwellings were loaded into the CAPI system to assist the listing team.

The purpose of using satellite mapping and listing was to eliminate the shortcomings of satellite mapping, as satellite mapping may not be up to date. The listing team were able to exclude ineligible buildings (such as non-residential and commercial buildings) and add new buildings which were not identifiable or apparent from satellite images. Figure 4 presents the satellite image for one of the sampled grids. The outcome of the listing served as the final sampling frame for the selection of households.

Figure 4 Satellite images of one grid for listing

Survey Implementation

Field work was conducted in three stages: the pre-test, the pilot, and the survey implementation. To develop the framework protocols for the survey, a pre-test and pilot were conducted. The first focused on system and field practices while the pilot finalized the protocols and tested the survey instrument.

The resulting methodology was designed to be relatively straightforward to implement in a low-capacity field environment. Each enumerator was given a tablet that had the selected sample points preloaded. In addition to the usual zoom and scrolling features, the device also displayed the position of the enumerator with respect to the sampled point and displayed a warning if the enumerator was beyond the acceptable distance from the sampled point. Figure 5 shows an example of a sample point in one grid.

Figure 5 Location of sample households in one grid

When the household was reached, the enumerator first sought consent from the household to participate in the survey. Two main challenges were encountered by enumerators. Some selected locations were not residential buildings, while in some cases, more than one household was found to be residing in the same dwelling. The first problem is very much related with the quality of listing activity done by the listing team, while the second problem was a likely result of the high poverty rate observed in the enumeration areas surveyed. Where more than one household was living in one dwelling, the enumerator was directed to conduct the survey to all households in the residence.

Weighting

To improve the accuracy of the estimates, it is necessary to perform a weighting procedure during the analysis. Weighting is assigned based on the recorded probabilities of selection at each stage of sampling unit selection. The applied weighting is formularized in Equation 1 below.

$$p_{hij} = P_1 P_2 = \frac{k_j n_{ij}}{N_j} \frac{m_{ij}}{n'_{ij}} \qquad (1)$$

where:

p_{hij} is the probability of selecting household h in grid i and stratum j;

P_1 is the probability of selecting a given PSU (grid) in stage 1 of the sampling strategy;

P_2 is the probability of selecting a given household in stage 2 of the sampling strategy;

k_j is the total number of PSUs selected in stratum j;

n_{ij} is the size (number of households) of grid i in stratum j;

N_j is the size (number of households) of stratum j;

m_{ij} Total number of households selected in grid i of stratum j;

n'_{ij} Total number of households listed in grid i of stratum j.

The household selection weights (referred to as w_{sel}) was the inverse of the probability of selection defined in Equation 2. Survey weights may further be adjusted to allow for any unavoidable non-response or calibration to updated population estimates. Thus, the final weights can be defined as follows:

$$w_{final} = w_{sel} \times w_{ps} \times w_{nr} \qquad (2)$$

where:

w_{sel} = Inverse probability of selection in the sample design;

w_{ps} = Calibration coefficient for selection weights;

w_{nr} = Adjustment coefficient to account for rare non-response;

Discussion and Conclusion

The most appropriate sampling method in any survey depends on a trade-off between financial and human resources, desired precision, and tolerable bias. The gridded population sampling methodology used to collect data in Tunisia demonstrated that the implementation of such a design is feasible; however, concerns remain about the quality of gridded population datasets. The study showed that sufficient gridded population datasets are available for free by various providers and other forms of GIS information is accessible often through web domains. The availability of gridded population datasets and GIS information rendered stratification possible where household were selected within those stratums. By using satellite images, the quality of the gridded population dataset was improved prior to listing. Through the listing activity, the attributes of selected grids were supplied with up-to-date information that enabled the team to compute more accurate household selection probabilities and minimize coverage error.

The proposed method has some advantages and disadvantages. Below is a detailed discussion of the method with regards to the practice and literature.

Advantages of proposed method

The gridded population sampling methodology enables researchers to utilize the features of probabilistic random sampling methodology by creating a spatial sampling frame when an existing sampling frame does not exist or is not reliable. In social surveys, survey methodologists typically rely on existing sampling frames, such as household enumeration lists (from censuses and other sources). But, in many countries such sampling frames do not exist, are incomplete, or may lack attribute data needed to stratify and contextualize the sampling frame. In these cases, as demonstrated in this paper, a gridded population sampling methodology coupled with listing can be employed to overcome problems related to the sample frame. Additionally, geocoded survey data allows social researchers to integrate other forms of data to conduct spatial analysis (Downey 2006).

Often, conventional sampling methodologies are not practical in conflict/war zones and gridded population sampling methods are employed (LP Galway, et al, 2012). In the study conducted by LP Galway et.al. (2010) in Iraq to estimate mortality, a two-stage cluster sampling method was employed. In the first stage, by using gridded population dataset of Landscan, grids were randomly selected and in the second stage, starting households were selected by using satellite imagery in Google Earth. The feasibility and methodological quality of the methodology was elaborated. The comparison of the total clusters (grids) and households in each governorate of Iraq with the percentage of Iraqi population in each governorate suggested that the sample sufficiently captures the regional distribution of underlying Iraqi population. However, some bias might have occurred due to the use of the Landscan population dataset and Google Earth imagery which were not quantified.

Thomson et al (2012), and Sollom et al (2011) experimented gridded population sampling as an alternative to conventional methods where census data are not available. And funding organizations such as DHS have begun to recommend gridded population datasets as alternative sample frames (DHS Listing Manual, 2012). In the study conducted by Thomson et. Al. (2012), a representative household survey was conducted in Idjwi in 2010 (Thomson, D.R., Hadley, M.B., Greenough, P.G. et al., 2012). LandScan population estimates and satellite imagery were used in the design of sampling to have representative estimates.

Sollom et al (2011) conducted a household survey in Myanmar in 2011 to quantify human rights violations and access to health care. Because of the absence of a sampling frame, LandScan population dataset and the 2006 census data were used for sampling design. The results suggested the feasibility of gridded population sampling for Myanmar as it produced consistent estimates with other large-scale household surveys.

Gridded population sampling methodology supported by information technologies minimizes bias stemming from enumerators discretion. Enumerator bias in social surveys cannot easily be quantified but frequently occurs. Lemeshow and Robinson noted that enumerator discretion might create bias because enumerators may prefer to identify the location of households that are close in the absence of cluster listing. In this circumstance, the selection of households becomes non-random. Survey methodologists do not prefer to have a non-random sample selection because of enumerators' intervention. According to Brogan and her colleagues, segmenting the initial clusters into sub-segments would minimize enumerator bias and would produce the random selection of households. However, this approach still requires listing which is cumbersome for many researchers. Turner et al. proposed sketch mapping of sample clusters, creating segments of roughly equal size (equal across all selected clusters), randomly choosing one segment per cluster, and interviewing all eligible persons within selected segments. In this way, enumerator discretion is avoided, and field work becomes less sceptical. However, in national surveys, the cluster sizes are often too large to manage such a listing activity in a smooth manner, would make logistical management harder and more budget would be needed.

In the Tunisia study, using satellite images, GPS of sample points and the availability of offline maps accessible through tablets/mobile phones, random points were generated, and enumerators were guided to sample households. Survey teams found this method easier to implement than conventional methods. The methodology was successfully implemented in a low-capacity environment without overly technical issues. This method has several strengths in terms of enumerator bias. First, it reduces the workload of enumerators, supervisors and coordinators. Second, it minimizes the enumerators' discretion in choosing sample households. Third, researchers are able to track the surveyed household locations to avoid errors and decisions made at the enumerators' discretion. For these reasons, this method is believed to minimize the bias stemming from enumerators.

In a study conducted by Harry et al (2012) in Lebanon, satellite imagery was used as a sampling frame and buildings from satellite images were randomly selected. The survey was conducted in Southern Lebanon in 2008 to estimate the extent of violence experienced by the population since the war with Israel in 2006 and gain information about attitudes towards civilian possession of arms. According to this study, the bias as a result of enumerators were minimized, and weight computation was possible which were not possible in case of absence of census records. However, the reliability of satellite images and GPS devices are the concerns of researchers.

Grais et al (2007) compared three methods to estimate vaccine coverage in Maradi, Niger in 2006. The spin-the-pen, GPS and grid methods were employed, and results were compared. The results of the survey suggested that GPS and grid methods performed quickly in an urban context and yielded similar results to the spin-the-pen random walk method while less bias was introduced.

Disadvantages of proposed method

Due to the complexity of the population distribution across geographical locations, statistical, surface and cartographic methods are employed to estimate the spatial distribution however, the estimates of gridded population and ground realities still diverge. The gridded sampling has some drawbacks.

In the Tunisia study, the GHSL gridded population dataset was used for sampling purposes. The population dataset was stratified and grids were selected by PPS method. In this dataset, the average estimated number of households per grid was 49 (maximum is 1050 and minimum is 3). The summary statistics of sample grids is provided in the table below.

Table 2 The summary statistics of sampled grids by strata

Strata	Number of Estimated HH (GHSL)	Number of listed households
Stratum-1	3176	691
Stratum-2	5405	668
Stratum-3	7911	1197
Stratum-4	2823	780
TOTAL	19315	3336

As displayed in the table above, there is a significant difference in households between the GHSL estimation and the listing results. At the grid level comparison, in 93 out of 115 sample grids, the GHSL overestimated the actual number of households and in 22 of the sample grids, GHSL underestimated the number of households. The mean of difference between GHSL estimates and listing was 139 households and no pattern was observed across the stratum.

The accuracy of gridded population varies and is often unquantified (Thomson, 2017). According to Thomson (2017), accuracy of estimations depends on several components: the accuracy of the input census data, the geographical scale of the input census data, the type of model covariate data and the model algorithm itself. Thomson (2017) also suggests using a 1km x 1km gridded population dataset instead of a 100m x 100m gridded population dataset. Many studies focused on assessing the accuracy of the gridded population estimates against census data, but the evidence is still not sufficient to evaluate accuracy of top-down gridded population dataset.

The labour-intensive process of detecting buildings/structures through satellite images and up-to-datedness of satellite images raise concerns about the gridded population sampling methodology. The Tunisia methodology requires visual inspection of satellite images in sample grids for listing operations. This process is labour intensive and done manually, and it poses a risk of human error to some extent. This is one of the shortcomings of the methodology. Furthermore, the resolution and up-to-datedness of satellite images have pivot role in providing high quality survey estimates. The dwellings should be visible without extensive tree-cover or cloud-cover. In our study, Google Earth images were inspected, and it was noticed that the images were up to date. However, in some locations where there is population movement due to seasonal migration, conflict, and epidemics should not rely on satellite imagery.

The probabilistic sampling methodology requires sampling units to be mutually exclusive across clusters and/or stratums. However, there were some cases were some buildings fell into two separate grids, which violates independency of probabilistic sampling methodology. In some cases, some households can be spread across two adjacent grids, which can violate the assumption of independence in the distribution within each grid. Additionally, overlaying a regular grid assigns equal probability of selection to each unit, but the number of households in each grid can vary. Therefore, the use of weights is necessary to compensate selection probabilities.

Stratification of the sampling frame can be challenging if sufficient GIS information is not available. Stratification is the process by which the survey population is divided into subgroups or strata that are as homogeneous as possible using certain criteria. The purpose of stratification is to enhance the sample representativeness with a given total sample size, thereby reducing sampling errors. In gridded population sampling designs, the contextual stratification will not be possible if sufficient GIS information is not available.

This study demonstrated the use of gridded population datasets and satellite images for constructing a spatial sampling design for household surveys. Gridded population sampling methodology is at an early stage and numerous approaches have been tested so far. The findings of these approaches are promising and may have many methodological strengths over conventional sampling methods. However, these methods have limitations and drawbacks, and researchers should focus on how to improve this approach. Land cover maps of countries should be developed to identify the areas where there is no residence. After that, the population models should be executed to estimate population for each grid. By this way, the accuracy can be improved. Second, the error components of gridded population models should be elaborated by multidisciplinary research teams. It is expected that this case study encourages further research and resources to examine implementations of these methods in various geographical locations.

References

- Center for International Earth Science Information Network Columbia University (2015) Gridded Population of the World, Version 4 (GPWv4): Data Quality Indicators, Beta Release; NASA Socioeconomic Data and Applications Center (SEDAC): Palisades, NY, USA.
- Eckman, S., Himelein, K., Dever, J., forthcoming. New Ideas in Sampling for Surveys in the Developing World, in: Johnson, T.P., Pennell, B.,ÄêE., Stoop, I., Dorer, B. (Eds.), Advances in Comparative Survey Methodology. 3MC.
- Grais, R.F., Rose, A.M., Guthmann, J.,ÄêP. (2007). Don't spin the pen: two alternative methods for secondstage sampling in urban cluster surveys. Emerging Themes in Epidemiology 4, 8. doi:10.1186/1742,Äê7622,Äê4,Äê8
- Himelein, K., Eckman, S., Murray, S. (2014). Sampling Nomads: A New Technique for Remote, Hard,Äêto,Äê Reach, and Mobile Populations. Journal of Official Statistics 30.
- Kish, L. (1965). Survey Sampling. New York: Wiley
- Kristen Himelein, Stephanie Eckman, and Siobhan Murray (2014) Sampling Nomads: A New Technique for Remote, Hard-to-Reach, and Mobile Populations. Journal of Official Statistics 30(2):191-213. https://doi.org/10.2478/jos-2014-0013
- Kumar, N. (2007). Spatial sampling design for a demographic and health survey. Population Research and Policy Review, 26(5-6), 581-599. https://doi.org/10.1007/s11113-007-9044-7
- Kumar N., Liang D. and Linderman M (2013) An Optimal Spatial Sampling Design for Social Surveys
- Laaksonen. S. (2007): Weighting for Two-phase Surveyed Data. Survey Methodology 33. 2. 121-130.
- Lafratta, G. (2006). Efficiency evaluation of MEV spatial sampling strategies: A scenario analysis. Computational Statistics & Data Analysis, 50(3), 878-890.
- Lemeshow S, Robinson D. (1985) Surveys to measure programme coverage and impact: a review of the methodology used by the expanded programme on immunization. World Health Stat Q. 38(1):65-75.
- LP Galway, Nathaniel Bell, et.al (2012) A two-stage cluster sampling method using gridded population data, a GIS, and Google EarthTM imagery in a population-based mortality survey in Iraq, International Journal of Health Geographics 11(12)
- Pirowski, T.; Bartos, K. (2018) Detailed mapping of the distribution of a city population based on information from the national database on buildings. Geodetski Vestnik 62, 458-471.
- Robert F. Chew, Safaa Amer, Kasey Jones, Jennifer Unangst, James Cajka, Justine Allpress & Mark Bruhn (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(12).
- Thomson, Dana & Stevens, Forrest & Ruktanonchai, Nick & Tatem, Andrew & Castro, Marcia. (2017). GridSample: An R package to generate household survey primary sampling units (PSUs) from gridded population data. International Journal of Health Geographics. DOI: 16. 10.1186/s12942-017-0098-4.
- Thomson, Dana R. (2017) Gridsample.org: Tutorial to Implement a Gridded Population Survey. Version 1. Flowminder Foundation: Stockholm.
- Thomson, D.R., Hadley, M.B., Greenough, P.G. et al. (2012) Modelling strategic interventions in a population with a total fertility rate of 8.3: a cross-sectional study of Idjwi Island, DRC. BMC Public Health 12, 959 doi:10.1186/1471-2458-12-959