Cloud Publications
International Journal of Advanced Remote Sensing and GIS
2016, Volume 5, Issue 10, pp. 1908-1914
ISSN 2320 - 0243, Crossref: 10.23953/cloud.ijarsg.78

Methodology Article | Open Access

Geospatial Techniques in Health Survey to Overcome the Lack of Sampling Frame

Elangovan A., Elavarsu G., Ezhil R., Prabu R., and Yuvaraj J.

National Institute of Epidemiology (ICMR), Ayapakkam, Chennai, Tamil Nadu, India

Publication Date: 14 October 2016

DOI: https://doi.org/10.23953/cloud.ijarsg.78

Copyright © 2016 Elangovan A., Elavarsu G., Ezhil R., Prabu R., and Yuvaraj J. This is an open access article distributed under the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract Sampling frame is a complete list of items or population from which a sample is drawn. More often researchers face difficulties in getting an up to date line list of sampling units, especially in the field where rapid changes are frequent. In household surveys it is difficult as newly constructed houses are seen as a continuous process in urban as well as rural areas. To overcome this issue, in this paper, we proposed a novel method by using Google-Earth with the GIS technologies. A study was done with the primary objective to identify the health impact of people living in and around quarries/crusher units at Thiruneermalai region nearer to Chennai. Since the boundary of the study area is found to be radial, it was difficult in getting the complete line list of the people living within the required radius. To overcome the issue the entire region was divided into small 30 x 30 meter grid cells. A total of 43652 sampling units (grid cells) formed the sampling frame. Among them, the required sample size was selected for the study. This method of GIS based grid sampling serve as an alternate tool for developing sampling frame with no or least cost.
Keywords *Geographical Information System; Grid Sampling; Sampling Frame*

1. Introduction

Sampling is an easy scientific method for investigating the entire population for assessing the outcome of interest [1-2]. Though various methods are available for deriving the representative samples from the population, simple random sampling method is considered to be more generalizable that represents the whole population since each individual in the sampling frame has equal chance of getting selected for evaluation [1-2]. Availability of sampling frame is a pre-requisite for deriving a sample using this method [3]. However in resource limited settings availability of reliable sampling frame is unlikely [4]. The available resources such as census enumeration did not have the location/ household identifier details moreover in the rapidly expanding population settings. The recent advances in technology related to Geographical Information System (GIS) and spatial analysis has paved way for overcoming limitations in the random sampling method. The GIS technology is being widely used in sampling natural resources and phenomena such as geological surveys, plant sampling, mineral deposition, bio-fuel exploration etc [5]. Spatial sampling does not rely on the sampling frame and building the sampling frame is not feasible in these cases. Sampling frame is

constructed in the spatial sampling by dividing the overlaid pattern into a defined number of units/grids/cells [5].

Further there is a paucity of literature, which demonstrates and depicts the usefulness and limitations of grid-sampling method in public health surveys. This method may reduce the cost of the survey to a considerable extent by minimizing the time, manpower and generating reliable estimates. With this background we explored the feasibility of using grid sampling method in estimating the health impact of hard stone quarry/crusher works in a sub-urban setting in Chennai in South India.

Public health is one of the major areas that have been made increasing use of Geographical Information System (GIS) techniques. In general, these efforts fall naturally within the domain of problems requiring use of spatial analysis as part of the solution [6-10]. GIS and other spatial analysis tools are therefore recognized as providing potentially transformational capabilities for public health efforts. A study used systematic GIS based grid sampling procedure for reconnaissance survey for assessing the percentage of positive for dengue and chikungunya caused mosquito breeding habitats [11], where the mean value of breeding habitats was calculated and assigned to each of the grid cells for mapping the mosquito breeding habitats. Efforts are continuously being done to improve the accuracy of grid based data and to find a simple and feasible method for treating population and housing grid data as the basis for sampling [12].

In order to fine tune the available methods, here we have proposed a novel sampling frame method, where we have included each and every point of the study area to form total sampling frame from where the required number of sampling units will be selected by simple random sampling method.

2. Study Area

The study area is identified as Thiruneermalai region nearer to Chennai which is located at 13°02′N 80°10′E/ 13.04°N 80.17°E on the southeast coast of *India* and in the northeast corner of *Tamil Nadu state (*Figure 1*)*. It is located on a flat *coastal plain* known as the *Eastern Coastal Plains*. Our objective is to measure the health impact of quarry works among the residents living within 5 KM radius form the quarries and to measure the related environmental parameters in the study area with specific attention to Tuberculosis.

Figure 1: *Study area with radial boundary*

We tried to get a complete list of houses of the study area but end up with failures as our boundary of study area is circular and not part of any existing administrative boundaries. To overcome this problem, we found an alternative strategy to develop a sampling frame by using Google-Earth [13], GPS and GIS technologies.

3. Materials and Methods

GIS/GPS based grid sampling method is used for recruitment of community residing nearby quarries. We identified the quarry site and created a buffer zone of 5 KM around the quarry over Google-Earth (Figure 1). Google-Earth has been used all over the world along with GIS tools for better research in public health [14-16].

Once the study area is identified, we blocked all water bodies, forest area and other areas where people are not residing. Hence it is ensured that only those areas where people residing are considered for selection. The entire 5 KM buffer-zone was overlaid with 30x30 meter small grids developed using Keyhole Mark-up Language (KML) programme [17]. There were 43652 such grid cells created (Figure 2) and each cell has been assigned with a unique number. The longitude, latitude, distance from centre point (quarry) and their unique identification of each grid cells have been stored in a file, from where 300 grid cells have been selected by simple random sampling method. The sample size of 300 was arrived by considering the approximate prevalence of morbidity conditions required by the study. Samples of two selected grids are shown in Figure 3.

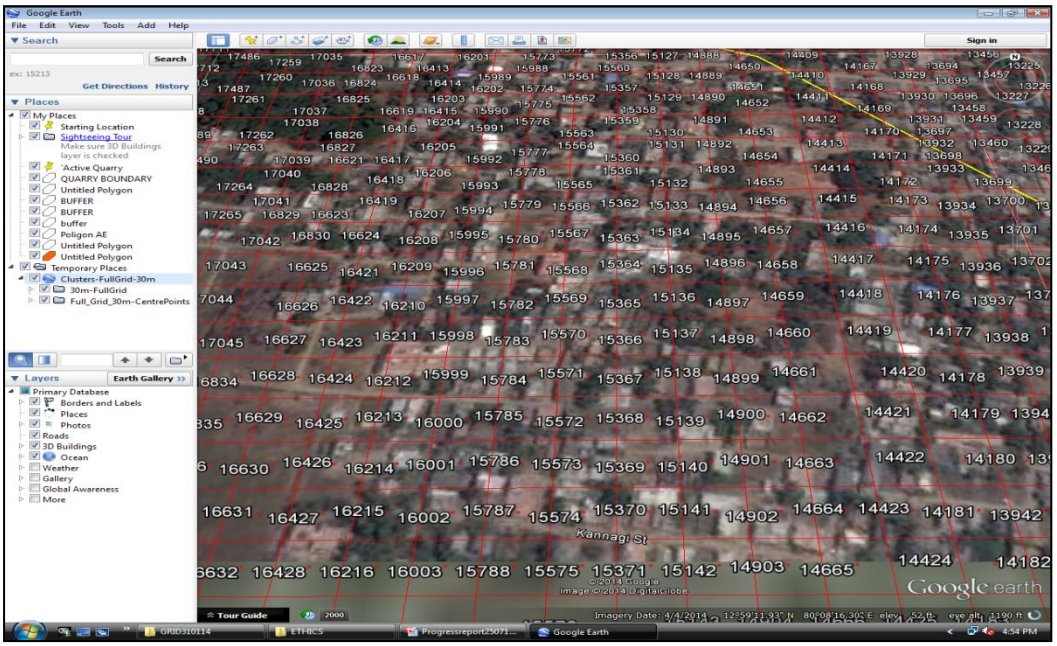

Figure 2: *Study area divided by 30x30 meter grids*

Our paramedical field teams who are well trained in using Global Positioning System (GPS) visited the selected grid with help of Tablet-PC, where all the 300 selected grids were overlapped on Google-Earth map [18]. Tablet-PC was used to assist the team in identifying the way and direction of the selected houses available in any particular grid of a street.

The longitude and latitude of grids were re-confirmed with the GPS readings before collecting the data from the particular households. Standard questionnaire was developed and the relevant information related to this study was collected. A maximum of 4 households were enumerated from each grid. If more numbers of household available in a single grid, we randomly selected 4 households as required

by the study. The study was initiated during February 2014 and a total number of 1905 subjects including 869 males and 1036 females were covered.

Figure 3: *Samples of two selected grids*

4. Results

Nearly 65 (22%) grids fall on empty land and the remaining 235 (78%) fall on structural buildings (Table 1). Hence one can expect that nearly 25% of grids may fall in vacant places in spite of the exclusion of water bodies, forest etc. Among the 235 grids whom fall on structural buildings, 208 (89%) grids were having residential structures and the remaining 27 (11%) grids fall on companies, schools, office etc., (Table 2). These factors have to be kept in mind while calculating the sample size and required number of extra grids has to be added with the sample size.

Table 1: *Distribution of sampled grids (N=300)*

Grid	Description of grid	No.	%
Fall on structural building	grids with habitations	208	
	companies, school, office etc.,	27	
	total	235	78
Fall on empty land	vacant land	53	
	water bodies	8	
	roads	4	
	total	65	22

Table 2: *Distribution of grids having structural buildings (N=235)*

Grid	Description of grid	No.	%
Habitations	grids with eligible HH	204	
	exclusion criteria	2	
	refusal	2	
	total	208	89
In-eligible structurals	companies, school, office etc.,	27	11

The relevant information was entered into Mapinfo [19], GIS software to perform further advanced spatial analysis. Figure 4 shows the distribution of grids from where the samples were collected. The cases found within the grids fall between 3 and 5 KM is considered are controls and between 1 and 3 KM is considered as cases.

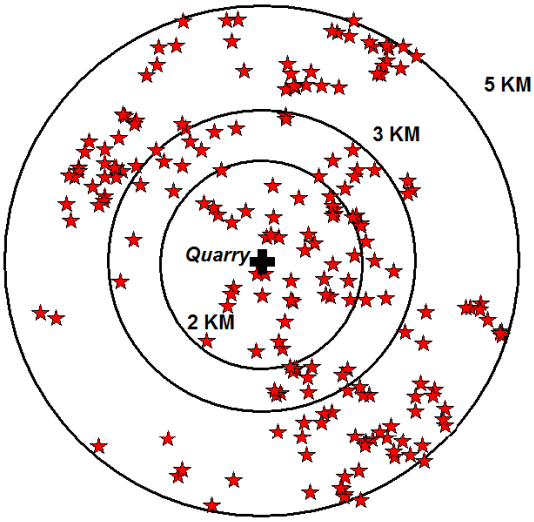

Figure 4: *Distribution of households within 2, 3 and 5 KM radius (N=204)*

5. Discussions

Use of spatial technology in identifying sampling units has been used in public health since many years. Baba Soumare et al. used computer software which randomly select a site to identify sheep/goats from that particular site [20], which may lead to have more number of empty points requiring more man power to search for a valid site/point and does not provide a complete sampling frame as large portion of areas are not been selected. Similar approach of creating random points was used by Kumar N. [5] to identify a household. These procedures may effective for collecting samples from ground water, soil, ambient air etc. It does not provide complete sampling frame as many of the points were not been selected. But the method proposed in this paper will form a complete sampling frame without leaving a single point/location as the entire study area is covered by grids. The size of the grid may vary depends upon the size of the sampling unit of any particular study.

Google-Earth is particularly very useful tool for public health, which can be used as free of cost. The new settlements can be viewed on the map as the Google-Earth is periodically updating their imagery data. Google Earth also provides higher resolution satellite images which can be used as a pre-navigation tool. Viewing grids or sample points in Google Earth before visiting the actual field helped field investigators to identify a right house without much difficulty and hence it reduces the cost, time and manpower.

6. Limitation

In the recent time GIS technology has been used in public health particularly in mapping mosquito breeding habitats, quantifying the ambient air pollution and in other socio-demographic surveys. Though there are many advantages of using grid-sampling method in public health surveys certain limitations exist such as a) empty cells which will take more manpower to find the valid cells; b) some households spread across two adjacent grids, which violates the principle of independence in the distribution within each grid; c) varying number of households across grids, which requires differential probability of inclusion by adding appropriate weight.

The important limitation is that the micro level Google imageries which include roads and structural building are available only for important regions/cities and not for all the regions. Removing non-residential areas like water bodies, companies etc., from being selected are considered as a big challenge.

7. Conclusions

House hold survey is often difficult as newly constructed houses are continues process in both urban and rural areas. Many times researchers find difficulties in getting up-to-date line list, especially in the field setting where rapid changes are frequent. Since Google-Earth capture periodical imageries of the globe along with their spatial locations, this demonstration of using grid sampling method with the combination of Google-Earth with other GIS technologies serve as an alternate tool for developing sampling frame with no or least cost.

References

[1] Royal Geographical Society with IBG, London. Sampling Techniques. https://www.rgs.org/OurWork/Schools/Fieldwork+and+local+learning/Fieldwork+techniques/Sampling+techniques.htm

[2] Stephen B., Hulley, 2001: *Designing Clinical Research- An Epidemiologic Approach*. Second Edition. Philadelphia, PA: Lippincott, Williams & Wilkins. 336.

[3] Michelle C. Kondo, Kent D.W. Bream, Frances K. Barg, and Charles C. Branas. A Random Spatial Sampling Method in a Rural Developing Nation. *BMC Public Health*. 2014. 14; 338.

[4] Mark V. Wilson and Oregon State University. Simple Random Sampling in the Field. http://oregonstate.edu/instruct/bot440/wilsomar/Content/SRS.htm

[5] Kumar, N: Spatial Sampling Design for a Demographic and Health Survey. *Popul. Res. Policy Rev*. 2007. 26; 581-599.

[6] Mona Choi, R.N. Spatial Analysis in Epidemiology. *Health Informatics Research*. 2013. 19 (2) 148-149.

[7] Rushton, G. Public Health, GIS, and Spatial Analytic Tools. *Annual Review of Public Health*. 2003. 24 (1) 43-56.

[8] Stefania Bertazzon. GIS and Public Health. *ISPRS International Journal of Geo-Information*. 2014. 3 (3) 868-870.

[9] Riner, M.E., Cunningham, C., and Johnson, A. Public Health Education and Practice using Geographic Information System Technology. Public Health Nursing. 2004. 21 (1) 57-65.

[10] Robinson, T.P. Spatial Statistics and Geographical Information Systems in Epidemiology and Public Health. *Advances in Parasitology*. 2000. 47; 81-128.

[11] Palaniyandi, M., Anand, P.H., and Maniyosai, R. GIS Based Community Survey and Systematic Grid Sampling for Dengue Epidemic Surveillance, Control and Management: A Case Study of Pondicherry Municipality. *International Journal of Mosquito Research*. 2014. 1 (4) 30-38.

[12] Ding Wenxiu, Xiaole, Li, Zhiqiang, Li, Donaixia, Zhang Yimei, and Temu Qile. Population and Housing Grid Spatialization in Yunnan Province based on Grid Sampling and Application of Rapid Earthquake Loss Assessment: The Jinggu Ms6.6 Earthquake. *Geodesy and Geodynamics*. 2014. 5 (4) 25-33.

[13] Google Company. Google Earth, https://www.google.com/earth/KML.

[14] Theodore B. Lefer, Matthew R. Anderson, Alice Fornari, Anastasia Lambert, Jason Fletcher and Maria Baquero. Using Google Earth as an Innovative Tool for Community Mapping. *Public Health Reports*. 2008. 123 (4) 474-480.

[15] Raoul Kamadjeu. Tracking the Polio Virus down the Congo River: A Case Study on the use of Google Earth™ in Public Health Planning and Mapping. *International Journal of Health Geographics*. 2009. 8; 4.

[16] Taylor, B.T., Fernando, P., Bauman, A.E., Williamson, A., Craig, J.C., and Redman, S. Measuring the Quality of Public Open Space using Google Earth. *Am J Prev Med*. 2010. 40 (2) 105-112.

[17] Wikimedia Foundation, Inc. Keyhole Markup Language
https://en.wikipedia.org/wiki/Keyhole_Markup_Language

[18] Ping Yu, Maximilian de Courten, Elaine Pan, Gauden Galea, and Jan Pryor. The Development and Evaluation of a PDA-based Method for Public Health Surveillance Data Collection in Developing Countries. *International Journal of Medical Informatics*. 2009. 78 (8) 532-542.

[19] Pitney Bowes Software. MapInfo Pro. The World's Premier Desktop GIS and Mapping Application. www.mapinfo.com/product/mapinfo-professional/

[20] Baba Soumare, Stefano Tempia, Vittorio Cagnolati, Abdullatif Mohamoud Guido Van Huylenbroeck, Dirk Berkvens. Screening for Rift Valley Fever Infection in Northern Somalia: A GIS based Survey Method to Overcome the Lack of Sampling Frame. *Veterinary Microbiology*. 2007. 121 (3-4) 249-256.