

RESEARCH ARTICLE

Open Access



Area based stratified random sampling using geospatial technology in a community-based survey

Carrie R. Howell^{1*} , Wei Su², Ariann F. Nassel², April A. Agne¹ and Andrea L. Cherrington¹

Abstract

Background: Most studies among Hispanics have focused on individual risk factors of obesity, with less attention on interpersonal, community and environmental determinants. Conducting community based surveys to study these determinants must ensure representativeness of disparate populations. We describe the use of a novel Geographic Information System (GIS)-based population based sampling to minimize selection bias in a rural community based study.

Methods: We conducted a community based survey to collect and examine social determinants of health and their association with obesity prevalence among a sample of Hispanics and non-Hispanic whites living in a rural community in the Southeastern United States. To ensure a balanced sample of both ethnic groups, we designed an area stratified random sampling procedure involving three stages: (1) division of the sampling area into non-overlapping strata based on Hispanic household proportion using GIS software; (2) random selection of the designated number of Census blocks from each stratum; and (3) random selection of the designated number of housing units (i.e., survey participants) from each Census block.

Results: The proposed sample included 109 Hispanic and 107 non-Hispanic participants to be recruited from 44 Census blocks. The final sample included 106 Hispanic and 111 non-Hispanic participants. The proportion of Hispanic surveys completed per strata matched our proposed distribution: 7% for strata 1, 30% for strata 2, 58% for strata 3 and 83% for strata 4.

Conclusion: Utilizing a standardized area based randomized sampling approach allowed us to successfully recruit an ethnically balanced sample while conducting door to door surveys in a rural, community based study. The integration of area based randomized sampling using tools such as GIS in future community-based research should be considered, particularly when trying to reach disparate populations.

Keywords: Area based, Geographic information systems, Stratified random sampling, Hispanic population, Rural population, Community based methods

* Correspondence: chowell@uabmc.edu

¹Department of Medicine, Division of Preventive Medicine, University of Alabama at Birmingham, Medical Towers 62, 1717 11th Avenue South, Birmingham, AL 35205, USA

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Obesity is a leading risk factor for the development of diabetes, cardiovascular illness, cancer and other chronic conditions that cause significant morbidity and mortality as well as increased health care costs [1]. Hispanics are the largest and fastest growing racial/ethnic minority group in the United States, comprising 17.3% of the population in 2014 [2], with disproportionately high obesity rates. Among adults living in the United States in 2015, the prevalence of obesity was 47% among Hispanics compared to 38% among non-Hispanic whites [3], highlighting the need to examine factors that contribute to this increased risk. To date, most studies among Hispanics have focused on individual risk factors of obesity, with less attention on interpersonal, community and environmental determinants. In order to conduct community level surveys to collect this type of data, it is crucial to ensure representativeness of both Hispanic and non-Hispanic populations in the study sample. Here we describe the use of a novel GIS-based population based sampling approach to minimize selection bias in a community based study.

Sampling for cross-sectional survey studies can be probability based or non-probability based. Probability based (e.g. random sampling) requires a defined population, where each possible unit has a known possibility of being selected [4]. Non-probability sampling methods (e.g. convenience sampling) have no known inclusion probabilities [5], producing bias and unbalanced sample representation [6–14]. Simple random sampling can also pose a problem for studies conducting research in minority populations. This method targets the whole population of interest and often results in minority under-representation. Stratified random sampling increases sample representativeness by dividing the study population into strata based on characteristics that are of interest to the researcher [15]. Random samples are then drawn from each strata to ensure adequate sampling of all groups. This approach reduces sampling bias; allows researchers to estimate within and between strata outcomes; and improves accuracy of results [15, 16].

Sampling design is important in large population studies with several national surveys utilizing stratified approaches to minimize bias. The US Census Bureau conducts the American Community Survey (ACS) to produce annually updated census data estimates based on geographic units (e.g. census tract and block group). The complex sampling design consists of first stratifying the US population by census block, then calculating population based sampling rates. Appropriate weights are applied in the analytical phase so that estimates represent the full population [17]. Similarly, the National Health and Nutrition Examination Survey (NHANES) employs a stratified, multistage cluster design that

oversamples specific subgroups to increase precision in health outcome estimates [18]. Smaller scale community based population studies should draw upon and incorporate aspects of these rigorous sampling designs to reduce sampling error and increase precision in estimates.

In recent years, technologies such as Geographic Information System (GIS) have been used to facilitate the sampling process in community-based research. Typically, GIS software have been used for data analysis and visualization [19]; however, health researchers have begun to realize its potential in facilitating the sampling and recruitment process, particularly in rural, developing countries [20–22]. To aid in sampling, GIS has been used to define populations in areas without formal census data [21, 22]; create clusters [22]; and stratify populations [20]. Area stratified random sampling methods use area units as the strata, such as census blocks, and produce samples comparable to random digit dialing recruitment approaches [20, 23, 24]. This method provides an innovative way to conduct community-based health survey research, particularly when the study area is small in population. Blending aspects of complex sampling design, such as those used in national surveys, with GIS methods has the potential to strengthen community based research. Here, we describe how geospatial data and Geographic Information Systems (GIS) were used to develop an area stratified random sampling protocol that ensured demographic balance in conducting a community-based, interviewer administered survey. The study's main aim was to examine social determinants of health and their association with obesity prevalence among a sample of Hispanics and non-Hispanic whites living in a rural community in the Southeastern United States.

Methods

Participants and setting

The population of interest resided in Albertville, Alabama where researchers had previously conducted a cervical screening study aimed at Hispanic women [25]. Located in Marshall County in the northeastern side of the state, Albertville has a population of 21,160 with 64.7% non-Hispanic white and 30.2% Hispanic as of the 2010 Census [26]. The city has two zip codes and is 26 square miles with a population density of 817 per square mile. The nearest metropolitan city with a population of

Table 1 Population and household characteristics in Albertville city

	Population	Households
Total	20,883	7401
Hispanic	5861 (28%)	1229 (17%)
Non-Hispanic	15,022 (72%)	6172 (83%)

over 150,000 is located 38 miles away. The median yearly income of Albertville is lower than Alabama as a whole (\$35,878 vs. 40,489). The Hispanic population is concentrated to approximately 17% of the households in the city (Table 1).

Data was collected from participants interviewed by trained research interviewers in door-to-door canvas between June and December 2013. To be included, participants had to be at least 19 years of age, not pregnant, speak English or Spanish fluently, and self-identify as non-Hispanic white or as Hispanic/Latino. Participants were compensated with a gift card for their time. All study procedures were reviewed and approved by the University of Alabama at Birmingham's Institutional Review Board.

Area stratified random sampling for recruitment

The goal to recruit an equal number of Hispanic and non-Hispanic participants would have been difficult to achieve by employing a completely random sampling procedure across the entire city. Therefore, a stratified random sampling procedure was created based on the Center for Disease Control and Prevention's (CDC) Community Assessment for Public Health Emergency Responses (CASPER) sampling methodology [27]. The CASPER approach was developed using cross-sectional epidemiological principles and is a form of a community needs assessment that provides a systematic approach to collecting household information on community public health status. The cluster sampling design involves two stages: selecting clusters based on household proportions and then interviewing a set random number of households in each cluster. The CDC recommends using GIS software in the selection of the sampling frame to allow users to select portions (clusters) of geographically defined areas, such as counties or cities. In addition, GIS software provides the ability to easily develop maps for community interviewers based on the selected clusters. For this reason, CASPER provides a toolbox for use in ArcGIS software to facilitate this methodology. Using this approach in our study involved three stages: (1) division of the sampling area into non-overlapping strata based on Hispanic household proportion; (2) random selection of the designated number of Census blocks from each stratum; and (3) random selection of the designated number of housing units (i.e., survey participants) from each Census block.

Stage 1: Divide the sampling area into non-overlapping strata based on Hispanic household proportion

To ensure that the interviewers would be able to reach sufficient Hispanic households, all Census blocks within Albertville were divided into four strata based on percentage of Hispanic households using GIS software.

Since Albertville city boundaries and Census block boundaries do not perfectly align with each other, a centroid criterion was used to determine whether or not a Census block belonged to Albertville city. As a result, 647 Census blocks were assigned to Albertville city. Of those, only 455 blocks contained households and the other 192 blocks were non-residential. Since the Hispanic population was concentrated in a relatively small geographic area, the 455 blocks were further divided into four unbalanced strata identified by Hispanic household proportion: < 10% Hispanic households, 10–30% Hispanic, 30–50% Hispanic, and \geq 50% Hispanic. Roughly 60% of the blocks were assigned to the \leq 10% of Hispanic households stratum, with 7% ($N=32$) of the blocks assigned to the >50% of Hispanic households stratum (see Table 2 and Fig. 1).

Stage 2: Randomly select the designated number of Census blocks from each stratum

Our goal was to recruit a total of 200 participants, with a distribution of 50% Hispanic and 50% non-Hispanic white (1:1 ratio). Maps denoted that the Hispanic population was largely concentrated in small area blocks (Fig. 1). Although smaller blocks suggest higher population density, they also contain fewer individuals and households compared with larger blocks. Since Hispanics comprised a smaller proportion of total households (17%), we needed to oversample blocks with higher concentrations of Hispanic households in order to reach an equal number of Hispanic and Non-Hispanic surveys. For these reasons we took the following approach to determine the number of Census blocks to select from each group, and the number of housing units to select from each Census block.

1. Considering the varying population size across blocks, it was determined to be more feasible to plan fewer surveys per block in more Hispanic population concentrated areas (i.e., strata 3 & 4 in Table 2), and more surveys per block in more non-Hispanic population concentrated areas (i.e., strata 1 & 2 in Table 2). As a result, we selected 10 blocks with 6 surveys per block from strata 1 and 2 and 12 blocks with 4 surveys per block from strata 3 and 4. These numbers were somewhat arbitrary, balancing the concern that selecting too many blocks which would increase cost, while taking care to not plan for an unrealistic quota of surveys per block when not feasible (e.g. the smallest block in the study area contained only 8 households).
2. For strata 1 and 2, distribution of Hispanic versus non-Hispanic surveys within each block roughly reflected the proportions of Hispanic and non-Hispanic households in the corresponding group.

Table 2 Proposed stratified sampling based on Hispanic household proportion

Strata	% Hispanic Households	Eligible Blocks	Selected Blocks	Surveys per Block	Surveys per stratum ^a	Surveys per Stratum by Hispanic Origin ^b	
						Hispanic	Non-Hispanic
Stratum 1	≤10%	268	10	6	60	5	55
Stratum 2	10–30%	101	10	6	60	20	40
Stratum 3	30–50%	54	12	4	48	36	12
Stratum 4	> 50%	32	12	4	48	48	0
Total		455^c	44		216	109	107

^aDerived by multiplying selected blocks and surveys per block

^bFor strata 1 and 2, distribution of Hispanic and non-Hispanic surveys within each block roughly equate to household proportions. Oversampling of Hispanics was planned a priori to reach recruitment goals thus proportions of Hispanic surveys in strata 3 and 4 were set higher than actual Hispanic household proportions

^cTotal of 647 blocks in Albertville, AL with 192 blocks with zero household, leaving n = 455 eligible blocks for sampling

Since oversampling of the Hispanic population was needed to achieve the recruitment goal, proportions of Hispanic surveys in strata 3 and 4 were set higher than the actual proportions of Hispanic households. Table 2 shows the proposed number of blocks to select from each group and numbers of Hispanic versus non-Hispanic surveys projected within each block. In total, we proposed 109 Hispanic surveys and 107 non-Hispanic surveys from 44 blocks.

Once the number of blocks from each group were determined, the CASPER toolkit developed by the CDC was utilized to generate random samples [27]. We used an add-on program developed for ArcGIS by the CDC to generate random samples using a polygon layer that represents the sampling area and non-overlapping clusters within the sampling area. In our study, the four strata were our sampling areas with Census blocks the non-overlapping clusters, accounting for the number of

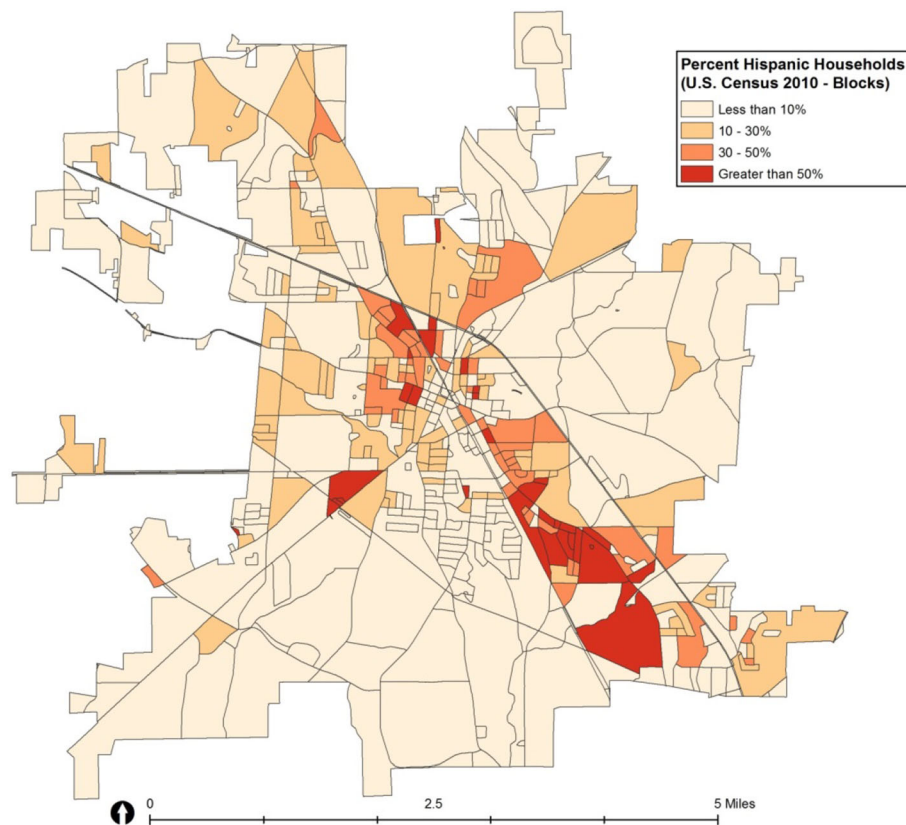


Fig. 1 Census blocks in Albertville, AL by Hispanic household proportion. Map of census block groups in Albertville, AL. Darker shading indicates higher Hispanic household proportions. Map developed using licensed ArcGIS software

housing units within each cluster. The random sampling procedure was repeated four times, once for each stratum. Figure 2 shows the 44 random blocks selected from the entire study area using this approach.

Stage 3: Randomly select the designated number of housing units from each Census block

Interviewers were provided with satellite maps (Fig. 3) for each block randomized with detailed instructions regarding how to randomly select the designated number of housing units within each block. The systematic random sampling method described in the CASPER toolkit [27] was adapted and modified to develop the study’s survey protocol:

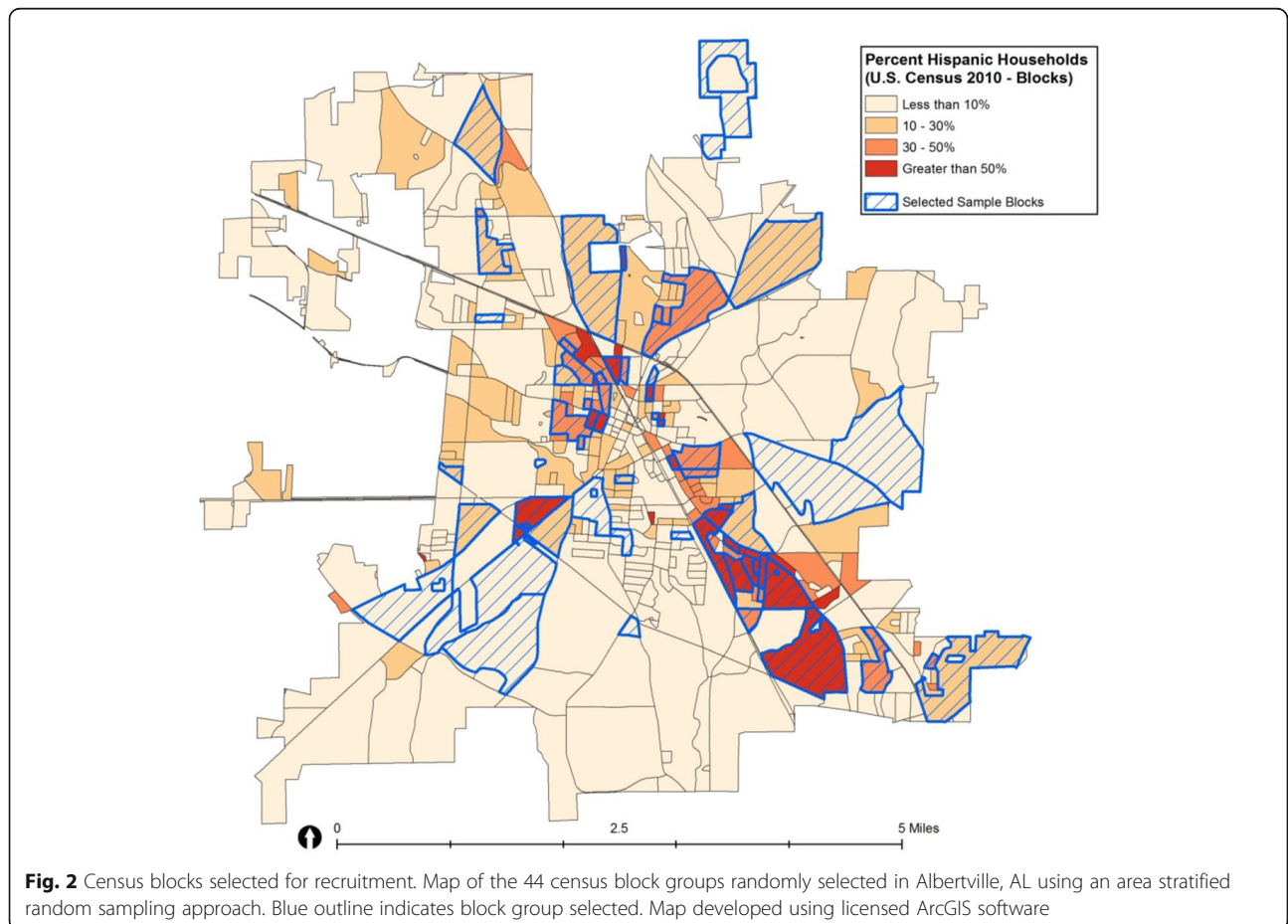
1. A starting point (address) for each sampling block was provided. This was the first house for the interviewers to survey.
2. After completing the first survey, interviewers would walk or drive in either direction to the next Nth house. This would be the next household for the interviewers to survey.
3. If no one answers the door, continue to the next Nth house.

4. Continue traveling through the sampling block, selecting every Nth house until they have completed the designated number of surveys for that sampling block.
5. If the interviewers circled back to the starting point and had not completed the designated number of surveys, they would then proceed through the block again and select every (N + 1)th house. For example, if Block A had an N of 8, in the next pass the interviewer would approach every 9th house.

The N used in the protocol was determined by dividing the total number of housing units by the designated number of surveys to complete in each block, and thus could vary from block to block. For example, if a block contained 50 housing units and the designated number of surveys was 6 for that block, the N would be 8. Values of N for each individual block were provided in the instructions to the interviewers. Additional instructions with regards to abandoned homes, businesses, duplexes and apartment complexes, multiple family homes, and trailer parks were also provided.

Results

The proposed sample included 109 Hispanic and 107 non-Hispanic participants to be recruited from 44



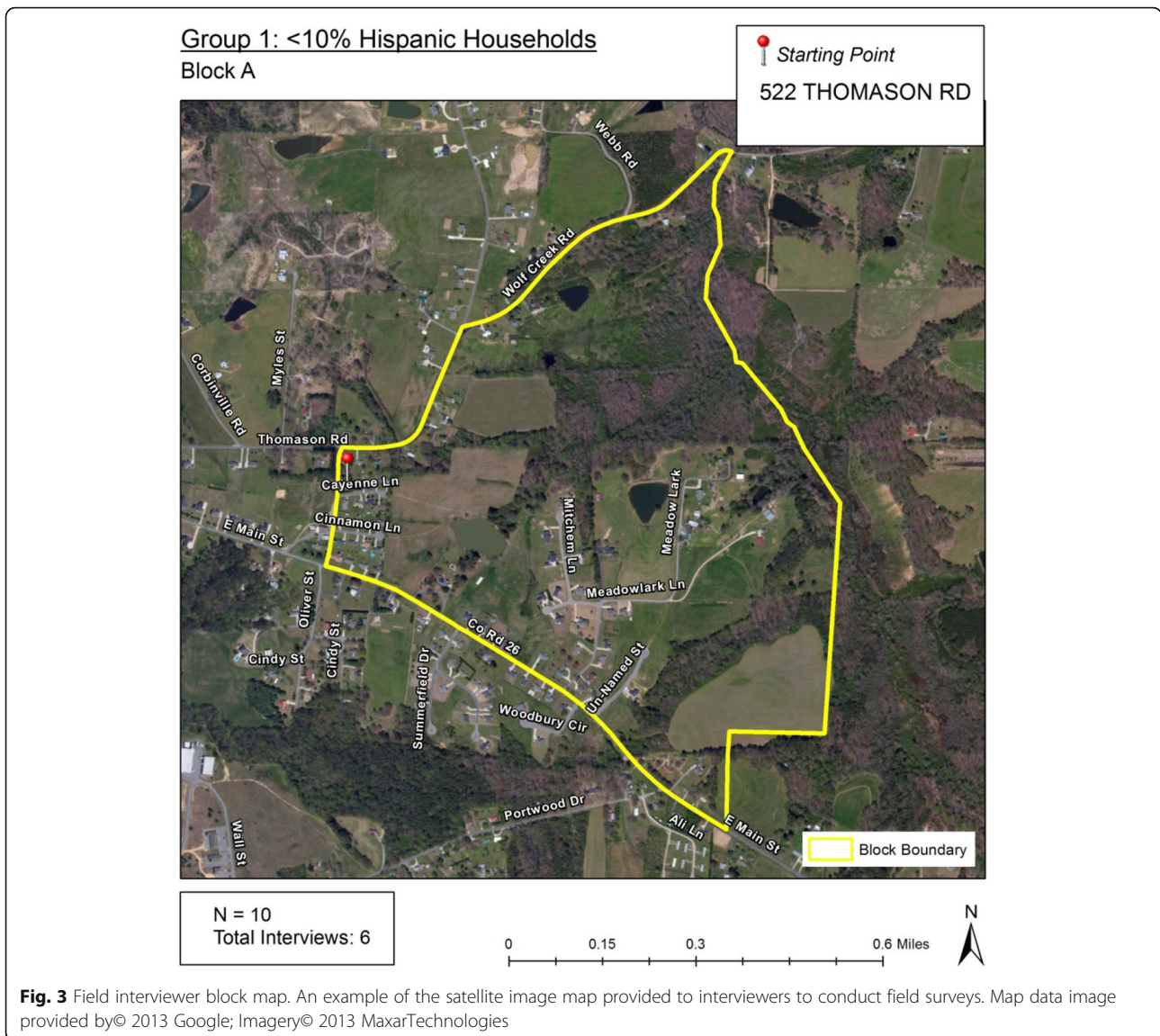


Table 3 Summary of actual sample

Strata	Proposed Sample				Actual Sample					p ^a
	Selected Blocks	Hispanic	Non-Hispanic	Group N	Selected Blocks	Hispanic	Non-Hispanic	Group N	% Hispanic Recruited	
Stratum 1 (≤10%) ^b	10	5	55	60	12	3	41	44	7%	1.0 ^d
Stratum 2 (10–30%)	10	20	40	60	12	15	35	50	30%	0.71 ^c
Stratum 3 (30–50%)	12	36	12	48	18	33	24	57	58%	0.07 ^c
Stratum 4 (> 50%)	12	48	0	48	22	55	11	66	83%	0.002 ^d
Total	44	109	107	216	64	106	111	217		

^aComparing the proposed distribution of surveys by ethnicity status to the proportions of surveys completed. *P*-values > 0.05 indicate that actual proportions did not differ from proposed population based proportions

^bProportion of Hispanic households in each individual block

^cChi-square test

^dFishers exact test

Census blocks. After exhausting all 44 blocks, interviewers were unable to meet recruitment goals for the proposed number of surveys in each block. Twenty additional blocks were selected using the same random sampling procedure described above, including two from strata 1 ($\leq 10\%$ of Hispanic households), two from strata 2 (10–30% of Hispanic households), six from strata 3 (30–50% of Hispanic households), and ten from strata 4 ($> 50\%$ of Hispanic households). More blocks with higher Hispanic population density were selected because field interviewers found that recruitment of Hispanic participants was particularly challenging. The final sample included 106 Hispanic and 111 non-Hispanic participants. The number of surveys completed from each block ranged from 0 to 11, with an average of 3.4 surveys per block (Table 3 and Fig. 4).

Post-hoc chi-square and Fishers exact tests were used to test the proposed distribution of surveys by ethnicity status to the proportions of surveys completed. P -values > 0.05 indicate that actual proportions did not differ from proposed population based proportions. The proportion of Hispanic surveys completed per strata were similar to our proposed distribution for strata 1–3: 7% for strata 1 ($p = 1.0$), 30% for strata 2 ($p = 0.71$), and 58%

for strata 3 ($p = 0.07$). Although Strata 4 (83% Hispanic surveys, $p = 0.002$) had statistically different proportions, this was expected due to the need to oversample Hispanic surveys from this strata.

Discussion

Here we demonstrate the successful use of a novel area stratified random sampling technique utilizing GIS that ensured ethnic balance in the recruitment of our community canvased study sample. Field recruitment in community studies presents challenges in minimizing selection bias and ensuring demographic representation. Here, integrating GIS based technology with census data provided a standardized and objective approach to recruitment to address these issues. Specifically, we utilized GIS to create and visualize non-overlapping strata to determine individual stratum and to randomly select Census blocks within those strata. Our approach ensured the 1:1 ratio of Hispanics to non-Hispanics in our study, minimized selection bias, and provided an approach that was easy for the ‘boots on the ground’ interviewers to implement. Moreover, the distribution of completed Hispanic surveys by stratum closely matched our original proposed proportions (defined based on

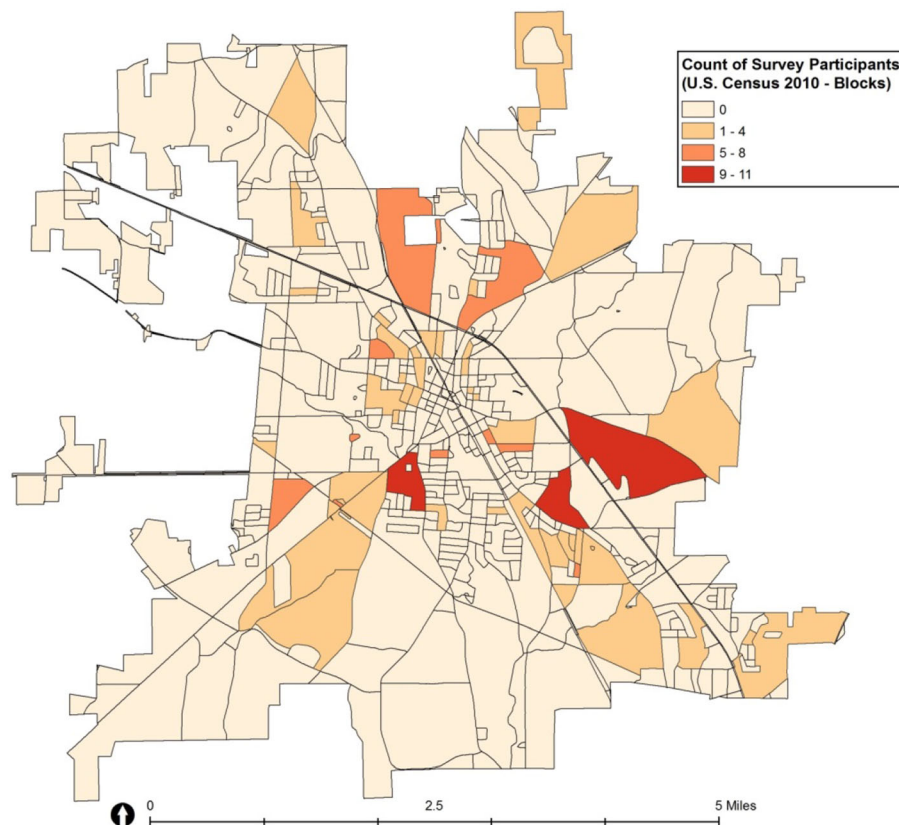


Fig. 4 Number of participants by Census block. Map of census block groups in Albertville, AL with the number of participants who completed a survey. Darker shading indicates more participants. Map developed using licensed ArcGIS software

percentage of Hispanic households in block), giving our sample geographic representation by Albertville block.

Utilizing GIS to facilitate community-based research, such as targeting areas for program planning or ensuring random sampling of survey respondents [28], has been implemented in recent population based studies. This method has been particularly useful in rural, developing countries [20–22, 29]. Defar et al. used GIS methods to conduct a cross-sectional survey in Ethiopia on maternal and child health care utilization in a similar two-stage process as the current study [29] while Wampler et al. used GIS to facilitate the random selection of households in specific areas in Haiti for water quality research [22]. Akin to the results here, a study that compared simple random sampling to stratified sampling by zip code and census tract found that area based stratified sampling ensured a higher representativeness of Hispanic residents in audits of tobacco retailers in an urban area [30]. In the public health realm, Lafontaine et al. developed a spatial random sampling method to conduct neighborhood built environment audits and concluded that this approach was more cost and time effective [31]. Likewise, using the approach herein resulted in recruiting our Hispanic sample in a more efficient manner.

It is important to note that we selected the number of blocks for randomization and recruitment based on feasibility but nonetheless in an arbitrary fashion. While this resulted in a balanced sample for our study, this will likely not translate into other scenarios. Since stratification by design results in subgroups that are over or under represented compared to the overall population [15], taking the actual population weights of each census tract into account when selecting blocks would have been more appropriate. Since the ultimate goal in sampling is to select a study sample that is representative of the population, applying population sampling weights and using model-based approaches such as raking prior to analysis are essential. Raking adjusts the sampling weights by forcing the survey totals to match proportions in the known population [32].

Our approach was not without challenges or limitations. When conducting the door to door surveys, interviewers were provided with detailed protocol and satellite maps. However, multiple issues arose. First, there was a significant number of houses that provided “no answer” and we had to implement the N + 1 sampling multiple times to reach recruitment targets. Time constraints also impacted interviewers. Some blocks sampled had a count number that was large ($N > 14$), which decreased sampling efficiency as driving from one house to the next could exceed 10 min. Another limitation of the study is that we used the population and household counts from the 2010 Decennial Census data, which may have underestimated the number of

Hispanics in Albertville at the time of data collection (2013). Further, the criterion used to divide the study area was Census block group and 2010 Census estimates were likely different than the true distribution of Hispanic households by block in 2013. Lastly, it is important to note that CASPER was designed for use in the United States and associated territories and uses data collected from the census bureau to create population based sampling areas and clusters. However, since CASPER was developed based on an epidemiological two-stage cluster sampling approach, it is possible to conduct this type of sampling in other countries where census type data are available using the CASPER protocol as a guide.

Conclusion

Overall, we developed a standardized area based randomized sampling protocol that allowed us to successfully recruit an ethnically balanced sample while conducting door to door community surveys. Minimizing selection bias in community-based surveys can be difficult; however, advancement in technological tools such as GIS provides novel approaches to address these biases. Based on our results here, we advocate the integration of area based randomized sampling in future community-based research, particularly when trying to reach disparate populations.

Abbreviations

GIS: Geographic information systems; CASPER: Community Assessment for Public Health Emergency Response; CDC: Centers for Disease Control and Prevention; ACS: American Community Survey; NHANES: National Health and Nutrition Examination Study

Acknowledgements

We especially thank Matthew Carle, Morgan Griesemer Lepard, Ynhi Thai, Meghan Meehan, Amancia Carrera, Sylvia Alavarez Mancinas, Susan Henry Barber, and Chris Caudill for their tireless efforts to canvas neighborhoods and interviews participants. We would also like to thank all our participants, the office of the Mayor of Albertville, the Albertville Police Department, support staff, and others who helped make this study possible.

Authors' contributions

WS, AC made substantial contributions to the design of the work. CH, WS, AC made substantial contributions to the analysis, interpretation of data and drafted the work. AN and AA made substantial contributions to the acquisition and interpretation of data. All authors read and approved the final manuscript and are accountable for the accuracy and integrity of the work presented.

Funding

This work was supported by grants from the University of Alabama with funding from the National Institute of Minority Health and Health Disparities (U54MD008176) and support from the National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, UAB Diabetes Research Center [1P60DK079626–01]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Minority Health and Health or National Institute of Diabetes and Digestive and Kidney Diseases or the National Institutes of Health or others supporting this work. All sources of funding had no role in study design; collection, analysis, and interpretation of data; writing the report; or the decision to submit the report for publication.

Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Ethics approval and consent to participate

This study was approved by the University of Alabama at Birmingham Institutional Review board and documented written informed consent was obtained from all participants prior to participation.

Consent for publication

Not applicable.

Competing interests

The authors declare they have no competing interests or financial relationships relevant to this article to disclose.

Author details

¹Department of Medicine, Division of Preventive Medicine, University of Alabama at Birmingham, Medical Towers 62, 1717 11th Avenue South, Birmingham, AL 35205, USA. ²School of Public Health, University of Alabama at Birmingham, 1665 University Blvd, Birmingham, AL 35233, USA.

Received: 22 May 2020 Accepted: 29 October 2020

Published online: 10 November 2020

References

- World Health Organization. Obesity and overweight fact sheet no. 311. 2013. Available from: <http://www.who.int/mediacentre/factsheets/fs311/en/index.html>.
- Steppler R, Brown A. 2014, Hispanics in the United States statistical portrait. 2016. Available from: <https://www.pewresearch.org/hispanic/2016/04/19/2014-statistical-information-on-hispanics-in-united-states/>.
- Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity among adults and youth: United States, 2015–2016. *NCHS Data Brief*. 2017;(288):1–8.
- Tashakkori A, Teddlie C. *Handbook on mixed methods in the behavioral and social sciences*. Thousand Oaks: Sage; 2003.
- Doherty M. Probability versus non-probability sampling in sample surveys. *N Z Stat Rev*. 1994;4:21–8.
- Fanzana B, Srunv E. A venue-based method for sampling hard-to-reach populations. *Public Health Rep*. 2001:216–22.
- Klein JD, Thomas RK, Sutter EJ. Self-reported smoking in online surveys: prevalence estimate validity and item format effects. *Med Care*. 2007;45(7):691–5.
- Roster CA, Rogers RD, Albaum G, Klein D. A comparison of response characteristics from web and telephone surveys. *Int J Mark Res*. 2004;46(3):359–73.
- Schillewaert N, Meulemeester P. Comparing response distributions of offline and online. *Int J Mark Res*. 2005;47(2):163–78.
- Schonlau M, Zapert K, Simon LP, Sanstad KH, Marcus SM, Adams J, et al. A comparison between responses from a propensity-weighted web survey and an identical RDD survey. *Soc Sci Comput Rev*. 2004;22(1):128–38.
- Spijkerman R, Knibbe R, Knoops K, Van De Mheen D, Van Den Eijnden R. The utility of online panel surveys versus computer assisted interviews in obtaining substance use prevalence estimates in the Netherlands. *Addiction*. 2009;104(10):1641–5.
- Bethell C, Fiorillo J, Lansky D, Hendryx M, Knickman J. Online consumer surveys as a methodology for assessing the quality of the United States health care system. *J Med Internet Res*. 2004;6(1):e2.
- Chang L, Krosnick JA. National surveys via RDD telephone interviewing versus the internet: comparing sample representativeness and response quality. *Public Opin Q*. 2009;73(4):641–78.
- Malhotra N, Krosnick JA. The effect of survey mode and sampling on inferences about political attitudes and behavior: comparing the 2000 and 2004 ANES to internet surveys with nonprobability samples. *Polit Anal*. 2007;15(3):286–323.
- Teddlie C, Yu F. Mixed methods sampling: a typology with examples. *J Mixed Methods Res*. 2007;1(1):77–100.
- Elfil M, Negida A. Sampling methods in clinical research; an educational review. *Emerg (Tehran)*. 2017;5(1):e52.
- American Community Survey Design and Methodology. Chapter 4: Sample design and selection. 2014. https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_ch04_014.pdf?#. Accessed 20 July 2020.
- National Health and Nutrition Examination Survey, 2015–2018: Sample Design and Estimation Procedures. https://www.cdc.gov/nchs/data/series/sr_02/sr02-184-508.pdf. Accessed 20 July 2020.
- Cromley EK, McLafferty SL. *GIS and public health*: Guilford Press; 2011.
- Kondo MC, Bream KD, Barg FK, Branas CC. A random spatial sampling method in a rural developing nation. *BMC Public Health*. 2014;14:338.
- Lin Y, Kuwayama DP. Using satellite imagery and GPS technology to create random sampling frames in high risk environments. *Int J Surg*. 2016;32:123–8.
- Wampler PJ, Rediske RR, Molla AR. Using ArcMap, Google Earth, and Global Positioning Systems to select and locate random households in rural Haiti. *Int J Health Geogr*. 2013;12:3.
- Aquilino WS, Wright DL. Substance use estimates from RDD and area probability samples: impact of differential screening methods and unit nonresponse. *Public Opin Q*. 1996;60(4):563–73.
- Lete C, Holly EA, Roseman DS, Thomas DB. Comparison of control subjects recruited by random digit dialing and area survey. *Am J Epidemiol*. 1994;140(7):643–8.
- Scarinci IC, Garces-Palacio IC, Morales-Aleman MM, McGuire A. Sowing the seeds of health: training of community health advisors to promote breast and cervical cancer screening among Latina immigrants in Alabama. *J Health Care Poor Underserved*. 2016;27(4):1779–93.
- US Census Bureau. QuickFacts: Albertville, AL. <https://www.census.gov/quickfacts/fact/table/albertvillecityalabama/PST045218>.
- Centers for Disease Control and Prevention. Community Assessment for Public Health Emergency Response (CASPER), sampling methodology. <https://www.cdc.gov/nceh/casper/sampling-methodology.htm>.
- Quon Huber MS, Van Egeren LA, Pierce SJ, Foster-Fishman PG. GIS applications for community-based research and action: mapping change in a community-building initiative. *J Prev Interv Community*. 2009;37(1):5–20.
- Defar A, Okwaraji YB, Tigabu Z, Persson LA, Alemu K. Geographic differences in maternal and child health care utilization in four Ethiopian regions; a cross-sectional study. *Int J Equity Health*. 2019;18(1):173.
- Lee JGL, Shook-Sa BE, Bowling JM, Ribisl KM. Comparison of sampling strategies for tobacco retailer inspections to maximize coverage in vulnerable areas and minimize cost. *Nicotine Tob Res*. 2018;20(11):1353–8.
- Lafontaine SJ, Sawada M, Kristjansson E. A direct observation method for auditing large urban centers using stratified sampling, mobile GIS technology and virtual environments. *Int J Health Geogr*. 2017;16(1):6.
- Battaglia MP, Izrael D, Hoaglin DC, Frankel MR. Practical considerations in raking survey data. *Surv Pract*. 2009;2(5):1–10.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

