Methods for Retrospective Geocoding in Population Studies: The Jackson Heart Study

Jennifer C. Robinson, Sharon B. Wyatt, DeMarc Hickson,
Danielle Gwinn, Fazlay Faruque, Mario Sims,
Daniel Sarpong, and Herman A. Taylor

ABSTRACT *The increasing use of geographic information systems (GIS) in epidemiological population studies requires careful attention to the methods employed in accomplishing geocoding and creating a GIS. Studies have provided limited details, hampering the ability to assess validity of spatial data. The purpose of this paper is to describe the multiphase geocoding methods used to retrospectively create a GIS in the Jackson Heart Study (JHS). We used baseline data from 5,302 participants enrolled in the JHS between 2000 and 2004 in a multiphase process to accomplish geocoding 2 years after participant enrollment. After initial deletion of ungeocodable addresses (n=52), 96% were geocoded using ArcGIS. An interactive method using data abstraction from participant records, use of additional maps and street reference files, and verification of existence of address, yielded successful geocoding of all but 13 addresses. Overall, nearly 99% (n=5,237) of the JHS cohort was geocoded retrospectively using the multiple strategies for improving and locating geocodable addresses. Geocoding validation procedures revealed highly accurate and reliable geographic data. Using the methods and protocol developed provided a reliable spatial database that can be used for further investigation of spatial epidemiology. Baseline results were used to describe participants by select geographic indicators, including residence in urban or rural areas, as well as to validate the effectiveness of the study's sampling plan. Further, our results indicate that retrospectively developing a reliable GIS for a large, epidemiological study is feasible. This paper describes some of the challenges in retrospectively creating a GIS and provides practical tips that enhanced the success.*

KEYWORDS *African Americans, Jackson Heart Study, Cohort studies, Geocoding, Geographic information systems, Spatial distribution*

BACKGROUND

Environment and neighborhood context are increasingly recognized as having a crucial link to individual's and communities' health.¹⁻⁴ Geocoding and the use of a

Robinson, Wyatt, and Faruque are with the School of Nursing, University of Mississippi Medical Center, Jackson, MS, USA; Wyatt, Hickson, Sims, and Taylor are with the School of Medicine, University of Mississippi Medical Center, Jackson, MS, USA; Wyatt, Hickson, Sims, and Taylor are with the Jackson Heart Study Examination Center, University of Mississippi Medical Center, Jackson, MS, USA; Hickson, Sarpong, and Taylor are with the Jackson Heart Study Coordinating Center, Jackson State University, Jackson, MS, USA; Gwinn is with the Center for Statistical Consultation and Research, Department of Spatial Analysis & GIS, University of Michigan, Ann Arbor, MI, USA.

Correspondence: Jennifer C. Robinson, School of Nursing, University of Mississippi Medical Center, 2500 North State Street, Jackson, MS 39216-4505, USA. (E-mail: jcrobinson@son.umsmed.edu)

GIS are essential tools being utilized to investigate the spatial relationships between area context and health.⁵ Key ecological and contextual factors that contribute to health outcomes can now be collected, managed, and assessed as a result of GIS technological advances. Many population studies have embraced the use of mapping as a method to investigate spatial clustering and the role of determinants of disease and health outcomes,^{3,6-8} and others whose resources or study design features precluded such data collection at baseline may have interest in retrospective geographical analyses. Few studies employing geocoding and a GIS have described the detailed methods. The lack of methodological detail limits the use and evaluation of reliability and validity of GIS data and hampers the multidimensional examination of contextual factors on health. In response to widening discussions of validity of spatial data, researchers are beginning to report more detail about geocoding match results⁹ or accuracy.^{10,11}

This paper describes the multiphase geocoding methods and resulting GIS used to assess baseline JHS participant characteristics and determine urban and rural geographic location in a large population study without geographic data collected at baseline. The protocol provides direction that can be utilized in other studies to increase the spatial sample size and improve the reliability of the geographic data. The discussion also addresses the challenges encountered while conducting this protocol. A detailed protocol for geocoding an existing dataset will become increasingly valuable as advancements in GIS methods and technology are applied to epidemiological and other research.

METHODS

The JHS is the largest single-site, population-based, all-African American longitudinal study ever conducted in the US. Designed to investigate the causes of cardiovascular disease using traditional and novel approaches,¹² the JHS included establishing a GIS for determining the adequacy of the study sampling plan, exploring spatial relationships among variables, and describing the geographic and social characteristics of the spatial data. Between 2000 and 2004, the JHS recruited and examined 5,302 African Americans from Hinds, Madison, and Rankin counties comprising the Jackson, MS metropolitan statistical area (MSA). The MSA includes the urban areas of Jackson and surrounding towns as well as rural areas across the three counties. Details of the study design, recruitment, informed consent procedures, and data collection protocols have been published elsewhere.¹³⁻¹⁶ Briefly, adults between the ages of 35 and 84 were enrolled via four sampling frames: the Jackson, MS site of the Atherosclerosis Risk in Communities (ARIC) study, random, volunteer, and family subsamples.^{14,17} Younger or older family members living outside the tricounty study area were eligible for enrollment in the family substudy. All participants responded to detailed home induction and clinic interviews conducted by trained interviewers and participated in a clinic examination. Mailing address information was collected during these interviews.

Institutional Review Board approval for geocoding and creation of a geographic database using JHS participant data was obtained from the University of Mississippi Medical Center (UMMC) and ratified by Jackson State University and Tougaloo College. Assurances to protect participants' confidentiality were also obtained from the JHS Geographic Data Subcommittee. The official disclaimer on the use of the GIS is posted on the JHS website (<http://jhs.jsums.edu/jhsinfo>).

Creating a GIS Database

Creating a reliable geodatabase (JHS GIS) that could be used in future studies to examine spatial and contextual relationships was a priority. Our goals were: (1) to successfully geocode 95% or greater of the addresses to the correct census block group, (2) assign latitude and longitude coordinates to each geocode, (3) avoid recontacting participants which would increase participant burden, and (4) minimize positional error without increasing resources and costs significantly.

We used a multiphase process that included: assuring complete participant addresses; geocoding addresses to longitudinal and latitudinal coordinates; validation procedures; classification of and georeferencing participants to census block groups; and linking census sociodemographic characteristics to each block group.

Phase 1: Assuring Complete Participant Addresses Participant's residential mailing address was obtained by trained interviewers during the home induction interview and subsequently entered into the JHS ClinTrial® database by trained data-entry personnel. Because the JHS was not designed specifically for geocoding, mailing addresses rather than physical street addresses were collected at baseline, presenting challenges to geocoding. Addresses that contained post office box numbers or rural route addresses which could not be improved were considered to have the potential for incorrect positioning and error in the geocode. In accordance with results of prior research documenting geocoding bias, geocoding by zip codes if no street address was available was also deemed to have unacceptable locational accuracy and was not conducted.^{18–20}

Prior to geocoding, addresses were sorted and searched for obvious data entry and spelling errors. Missing, incomplete, and inaccurate addresses were identified from the database and paper forms were retrieved to abstract the available contact information. If needed, an online telephone directory (www.anywho.com) was used to obtain matching street address information. Only addresses that matched the participant's name were included. An *a priori* decision was made to exclude geocoding of the family substudy participants living outside the tricounty area for the baseline GIS because the numbers were few ($n=13$) and locations were widely distributed across the state and US limiting the power to detect spatial neighborhood effects on individual health. Post office and rural route addresses that could not be improved were excluded due to likely positional error, as were addresses outside the tricounty study area.

Phase 2: Geocoding Street Addresses Geocoding was performed in ArcGIS 9.1²¹ because of its utility, ability to query data and create maps, and wide use by researchers and a number of industries, including environmental, governmental, and public health agencies. Topologically Integrated Geographic Encoding and Referencing (TIGER©) geographic map files²² and 2000 US Census data were projected using USA Contiguous Equidistant Conic with North American Datum 1983 (NAD83) geographic coordinate system projection.

The ArcGIS StreetMap USA “US Streets with Zone” address locator, a geocoding engine to perform address standardization and matching based on set parameters, and ESRI 2005 street maps were used to assign each address a geographic location. ArcGIS offers two geocoding methods: automatic matching (batch matching) based on the address match parameters, and interactive matching, where addresses can be reviewed and addresses corrected as needed.

JHS geocoding was done using batch matching followed by interactive rematch-ing.^{2,3} Participants were geocoded and assigned x - (longitude) and y - (latitude) coordinates in decimal degrees using spelling sensitivity of 70%, minimum candidate score of 10%, and minimum match score of 80% as the set parameters. Match scores are assigned based on how well the address attribute matches the address locator in the geocoding engine and are affected by spelling, house number, directional attributes such as north or south and street versus avenue, street name, city, state, and zip code differences. A perfect match has a match score of 100%. Coordinates were assigned during geocoding. Briefly, during geocoding, the ESRI street map uses an address locator with address attributes and indexes to translate nonspatial addresses (e.g., those addresses in a table that contain house number, prefix direction, street name, street type, city, state, etc.) into locations on a map. The points with corresponding coordinates are extrapolated from the spatial data that exists in the address locator on the known locations, including x - and y -coordinates of street intersections and end points of street segments and located spatially on the map. Thus, each geocoded address is a point consisting of a pair of x,y coordinates.²¹

Unmatched addresses were reviewed individually and matched if deemed to be acceptable (e.g., if the street name was misspelled). Match scores below 80% were assessed during the interactive rematch procedure. Unmatched files were exported and rematched using a street map locator from the US Census Bureau Mississippi Data Center (www.olemiss.edu/depts/sdc/). Further attempts to locate unmatched addresses included manually geocoding each address using: (1) purchased paper street maps of the tricounty area with temporal copyrights for the JHS baseline examination,^{24–27} and (2) Mapquest (www.mapquest.com)²⁸ and the US Postal Service (www.usps.com)²⁹ websites to verify that there was an existing address matching the one recorded in the paper file.

Phase 3: Validation A second geocoding process was performed by an independent researcher without record abstraction for missing or incomplete values. The second geocoding process was performed using ESRI 2000 and 2002 street maps to assign each address a geographic location. This allowed for comparison of the two resulting GIS, enabling the validation and accuracy of the first geocoding process. A 25% quality control (QC) check was conducted by comparing the available x - and y -coordinates and match scores from each process. Euclidean distances were computed using the x - and y -coordinates to determine the accuracy in using the 2005 street maps as compared to the 2000 and 2002 street maps.

Phase 4: Classification of and Georeferencing of Participants to Block Groups Each block was classified as urban or non-urban based on the 2000 US Census Bureau's characterization of an urban cluster or urban area.³⁰ Any block not classified as urban was considered rural. Because our unit of analysis was block group, a block group was classified as urban or rural if all blocks within the block group were similarly classified. For block groups containing both urban and rural blocks, mixed block groups were defined based on the preponderance of urban or rural blocks: mixed urban block groups contained a greater number of urban block groups; mixed rural block groups contained a greater number of rural blocks. Finally, each participant, along with their baseline data, was georeferenced to a census-defined block group.

Phase 5: Linking Neighborhood Sociodemographic Characteristics Select block group sociodemographic characteristics (e.g., median block group income; mean block group education) obtained from the 2000 US Census were linked to the GIS by specific block group codes. A series of maps were produced to geographically illustrate the distribution of sociodemographic characteristics across the study region.

RESULTS

Participants with a missing mailing address ($n=10$), post office box address that could not be improved by record abstraction or reverse lookup using the telephone number ($n=29$), or an address outside the tricounty area ($n=13$) were excluded during phase one, leaving 5,250 for geocoding in phase two. The majority of the deleted post office box addresses were located in urban area post offices ($n=16$). After implementing both the batch and interactive matches, 4,524 addresses (86%) were matched with a score of 80% or higher, indicating how well the participant address matched the address locator, and 558 (10.6%) were matched with a score of less than 80%; 168 addresses (3.2%) were unmatched. Of the 168 unmatched records, 25 (15.0%) additional addresses were matched using the US Census Bureau Mississippi Data Center street map files, leaving 143 to be manually geocoded. Of these, 130 were manually plotted inside the tricounty area, seven were plotted outside of the tricounty area, and six could not be plotted. The latter two groups ($n=13$) were removed from the GIS. The final GIS included 5,237 (98.8%) of the 5,302 JHS participants (Fig. 1).

Of the addresses that were not matched using batch and interactive matches and either set of map files and were subsequently manually plotted, 60% were located in urban block groups, 15% were in rural, 15% were in mixed rural, and less than 10% were in mixed urban block groups. Urban addresses were easier to locate using Mapquest and US Postal Service addresses. Mixed urban and mixed rural areas were slightly more difficult especially if new subdivisions with new streets had been added. Rural addresses, which comprised only 15% of those plotted, required the use of the purchased maps and Mapquest.

The independent geocoding process without record abstraction yielded a much lower success rate: 4,524 (85.3%) of the 5,302 were successfully geocoded with a match of 80% or higher, 259 (4.9%) were matched with scores less than 80%. The remaining records ($n=519$) were unmatched and not included in the GIS of the independent geocoding process. The 25% QC check resulted in a 100% match of scores and Euclidean distances equal to zero (i.e., geographic location of the geocoded addresses were the same).

As depicted in the study area map (Fig. 2) JHS participants were georeferenced to 288 (91%) of the 317 block groups. According to the 2000 US Census, the percentage of African American population in the 29 block groups without JHS participants varied: 25 contained 30% or less African Americans with seven containing no African Americans; four contained more than 30% African American population. One block group with no JHS participants contained 99% African American population, but was small in both area and total population ($n=559$) compared to the other block groups.

The majority of the 317 block groups within the study area were classified urban ($n=213$, 67.2%), with the remainder almost equally classified as rural ($n=40$,

FIGURE 1. Algorithm of the steps used in Phases 1, 2, and 3 of the geocoding process for the JHS.

12.6%), mixed rural ($n=35$, 11%), and mixed urban ($n=29$, 9.1%; Fig. 3). Similarly, 4,033 (77.0%) JHS participants resided in urban block groups. Over 500 participants resided in rural block groups ($n=504$), followed by 446 in mixed rural, and 254 in mixed urban block groups.

The JHS sample was comprised of a higher percentage of African American females compared to the 2000 Census (which included females of all races) and a higher percentage of participants with a high school diploma or above (Table 1).

FIGURE 2. Distribution of JHS participants per block group.

Other participant demographic data have been previously reported (13) and are not included in this description of the GIS methods utilized.

DISCUSSION

The increasing focus on the effects of environment and place on health necessitates reliable and valid geographical datasets. This detailed description of the methods

FIGURE 3. Type of block group.

and outcomes of creating the JHS GIS provides a resource for accomplishing retrospective geocoding of large epidemiologic studies where spatially linked participant data were not obtained at baseline. Such methodological detail is essential for evaluating the quality of the geographic data to be used in further contextual and spatial analyses.

The detailed descriptions of methods used in creating a GIS, including geocoding addresses, in population studies to date have been minimal.^{9,31,32} Other

TABLE 1 Select baseline characteristics for the Jackson Heart Study (JHS) cohort and 2000 US Census stratified by type of block group within tricounty study area

	Urban block groups <i>n</i> = 213		Rural block groups <i>n</i> = 40		Mixed urban block groups <i>n</i> = 29		Mixed rural block groups <i>n</i> = 35	
	JHS participants <i>n</i> = 4033	Census population <i>n</i> = 251,978	JHS participants <i>n</i> = 504	Census population <i>n</i> = 69,297	JHS participants <i>n</i> = 254	Census population <i>n</i> = 58,566	JHS participants <i>n</i> = 446	Census population <i>n</i> = 60,960
Sex								
Female	65%	56%	61%	53%	63%	53%	58%	52%
Male	35%	44%	39%	47%	37%	47%	42%	48%
Education								
High School or Above (Range)	79.9%	77.4% (33–100%)	83.5%	74.8% (53–99%)	86.6%	85% (44–100%)	90.8%	84.9% (64–98%)

Proxy for neighborhood is defined as Census block group. The JHS study area consists of 317 block groups

large, cardiovascular epidemiological studies, including ARIC,^{3,6} the Cardiovascular Health Study,⁸ and the Coronary Artery Disease Risk Development in Young Adults,³³ have utilized geocoded data, but did not describe their GIS methods in detail, limiting the ability to assess the reliability of the geocoded data. Several reported the use of private firms to geocode their data without providing the level of methodological detail presented here.^{33–36} Whether geocoding is done in-house or by a commercial firm, reporting the detailed methods and geocoding results are important for assessing validity as accuracy and error varies using either method.^{11,37} The description of the phases, processes, and decisions involved in retrospectively creating a GIS in the JHS provides a basis for future studies to include in study protocols. Such GIS data offer the opportunity to improve the understanding of place and its contribution in differential health exposures, access to health services, and health outcomes for African American and other minority ethnic groups.

In the JHS after initial deletion of addresses deemed either ungeocodable or unacceptable for geocoding by our protocol ($n=52$), 96% were geocoded using the automatic batch match and default setting in ArcGIS. Following the multistage methods utilized and described, nearly 99% of the enrolled JHS cohort was geocoded. Several other studies have reported similar¹⁰ or slightly lower results^{38,39} using a multistage interactive approach or further data collection. McElroy et al.³⁹ recontacted almost 600 participants, but were only successful in matching an additional 276 participants, increasing both personnel and study costs. The JHS used record abstraction and a variety of other methods to improve address specificity and to assign a geographic location to each address, without recontacting participants. Two street map sources with separate address locators were used to improve matching which may not be typical of most other studies that used geocoding.^{8,31,32,34,36} Combining multiple map sources improved the ability to identify street locations and match results without significant financial or time costs. Verification of the existence of addresses using www.usps.com and subsequent location of unmatched addresses using www.mapquest.com were useful tools and improved our results. While these resources are readily available for no cost, both methods increased the personnel hours on the project.

Georeferencing the participants' geocoded data to block groups provided the opportunity to develop maps to assess the distribution of JHS participants across the tricounty study area. While it was not the aim of the JHS to have a balanced sample from each block group but rather to obtain a representative population-based sample with the primary objective of investigating the causes of high prevalence of cardiovascular disease, the GIS gave a snapshot of the success of recruitment strategies that were employed. Classifying each block group as urban, rural, or mixed urban/rural allowed comparing sociodemographic characteristics of the JHS participants with census information for the region. JHS participants were georeferenced to nearly 93% of the block groups in the study area that contained any African American residents and 98% with 5% or more African American residents. The map dispersion provides a means of illustrating the unique elements of the JHS cohort and evaluating the sampling plan for the JHS. The majority of the cohort resided in areas with >30% African American population by design. Racial and ethnic identification was not available in the sampling frame used for the JHS random sample. Due to the high cost of determining eligibility and anticipated low numbers of eligible residents, as lists of potential participants were generated for recruitment in the random sample, persons living in neighborhoods with less than 30% African American residents were excluded from contact.¹⁴ Thus, participants

living in neighborhoods with 0 to 29% African Americans were not enrolled through the random sample. The ARIC sample, all of which were originally recruited from within the Jackson city limits, represented a large proportion of the JHS participants that resided in urban block groups. The family and volunteer samples were enrolled if participants met eligibility criteria with no specific geographic target.

Creating the JHS GIS revealed several key lessons. Complete, accurate baseline address data were extremely important to the success of the project. Because the JHS was not designed specifically for geocoding or for future spatial and contextual analyses, mailing addresses rather than physical street addresses were collected at baseline, presenting challenges for geocoding. Address matching requires complete and accurate addresses, especially in urban areas where a single street name may have several directional designations (e.g., N, S, NW, etc.) or types (street, place, circle, etc.). The extensive time required for record abstraction and address improvement (approximately 200 h) and location of map addresses for subsequent manual plotting on the map (an additional 80 h), coupled with the support of GIS experts, were critical to the success of geocoding the JHS cohort. Goldberg et al.³⁸ investigated the time and effort to improve geocodes in five existing datasets with an interactive web mapping system and concluded that the time involved was substantial, but overall was cost effective. Their time results per record were shorter than in the JHS as they did not attempt to improve addresses through further data abstraction in existing records or other methods used here. Including collection of geocodable addresses in the study design could have reduced much of the time involved in improving addresses.

Additionally, geocoding results varied by urban, rural, mixed urban, and mixed rural location. Generally, urban addresses were easier to locate using the batch and interactive rematch techniques as street names were included in our two map files used. Of the urban addresses manually plotted, the location was effectively located using Mapquest or our reference maps and plotted within a confined area. Addresses located in mixed urban and mixed rural areas frequently involved the addition of new streets that were either not found on the reference maps or via Mapquest. We were able to locate the address using one or the other source or additional reference maps. Rural addresses were located and plotted but generally took more time to locate and necessitated the use of all of our resources.

The JHS GIS is not without limitations. Many of these limitations may similarly exist in other studies using retrospective geocoding of existing participant data. The self-reported participant residential addresses obtained at one point in time incurred all the usual biases of using cross-sectional data. Geocoding did not occur contemporaneously with the collection of baseline data. To minimize potential temporal biases between baseline data collection and geocoding, maps that corresponded to the time period of JHS participant enrollment were used and nearly half of the problematic addresses were located using www.mapquest.com soon after participant baseline data collection. However, the remaining were geocoded 2–4 years after baseline data collection. One suggestion for researchers is to geocode as soon as possible after initial participant data collection and if not possible, to obtain maps of the area that correspond temporally. Libraries and bookstores may be excellent places to begin to search for such maps.

In the JHS, 86% of the addresses were matched at 80% or higher. However, match scores reflect the agreement between the address that is being geocoded and the address locator of the geocoding engine and do not provide information on the

accuracy of the geocoded position. Although the geocoding results revealed very high success in locating addresses, there is the possibility that urban address locations may be more precise than rural because of shorter distances between street segments. Rural street segments are longer and although rural addresses were successfully located, the potential for location error must be acknowledged. Several researchers have noted that location accuracy is better in urban areas, while error increases in suburban areas and is the greatest in rural areas.^{37,40,41} Differences in geocoded locations generally have been less than 100 m for urban areas and can range from 52 to greater than 1,000 m in rural areas.^{40,41} Although we did not compare our geocoded results to positions obtained either via satellite or by using a global positioning system (GPS), we believe the dataset is accurate based on the multiple methods and maps used and has the level of precision necessary for the types of spatial analysis planned in future studies using JHS data. As in any research, researchers must determine the needs of the project and how the positional error inherent in each geocoding method may affect the validity of results.⁴⁰ For the JHS geodatabase, exact GPS coordinates would have been expensive and time-consuming to obtain after enrollment on 5,302 participants in three counties.

The reverse telephone lookup procedures could not be employed for cellular telephones, which will present continuing challenges for future research, as they increasingly become the primary telephone. Strategies that could have prevented most of the issues of address location encountered in geocoding a large, prospective study include using a field GPS unit to create *x*- and *y*- position coordinates, obtaining a 911 system address at the time of the home interview, obtaining the street address in addition to the mailing address, and documenting the house location with major cross streets. The cost of GPS units has substantially declined since the study planning for the JHS and new studies should be able to incorporate this more accurate method into research designs. Future studies should consider this relatively inexpensive addition at baseline as part of home-based recruitment and data collection.

Sophisticated analyses of contextual effects on health outcomes will be enhanced by GIS across the life course. Tracking participant mobility as well as obtaining actual addresses at specific life points could strengthen longitudinal study designs. Such data collection is difficult, often relying on recall of past addresses and length of residence. The JHS will be able to geocode parental address at time of birth from birth certificates; however, this process is likely to be fraught with methodological difficulties. As well, ongoing annual follow-up and surveillance of JHS participants will allow for ongoing address geocoding as participants move within or outside of the study area.

CONCLUSION

This paper described many of the challenges in creating a GIS in one large population study and offered practical solutions that enhanced success. Using these multiple strategies, almost 99% of the enrolled participants' addresses were geocoded. Other researchers can utilize many of these solutions either during the research planning stages or retrospectively using existing data in order to create geographic data in large population-based studies without large fiscal costs. The JHS GIS allows analyses and mapping of chronic disease patterns and investigation of the impact of neighborhood residence and neighborhood characteristics with different chronic diseases and health behaviors. Further, it allows analyses of the

impact of the built environment on health and health behavior, and assessments of access to health resources. Many of these studies are currently underway. The JHS and its newly created GIS provide a unique opportunity to begin to identify potential reasons for the health disparities that continue to exist for African Americans in the US.

ACKNOWLEDGMENTS

The authors gratefully thank the participants and staff of the Jackson Heart Study for their contributions and commitment. We also thank Gloria Miller for her contribution to the geocoding used in validating the results. This work was supported by National Institute of Health contracts NO1-HC-95170, NO1-HC-95171, and NO1-HC-95172 provided by the National Heart, Lung, and Blood Institute and the National Center for Minority Health and Health Disparities, National Institutes of Health and by funding to Jennifer C. Robinson, contracts F31-NR0008460 from the National Institute of Nursing Research and the National Center for Minority Health and Health Disparities and T32-NR07073 (University of Michigan School of Nursing) from the National Institute of Nursing Research.

REFERENCES

1. Augustin T, Glass TA, James BD, Schwartz BS. Neighborhood psychosocial hazards and cardiovascular disease: the Baltimore Memory Study. *Am J Public Health*. 2008; 98(9): 1664-1670.
2. Chaix B, Rosvall M, Merlo J. Neighborhood socioeconomic deprivation and residential instability: effects on incidence of ischemic heart disease and survival after myocardial infarction. *Epidemiology*. 2007; 18(1): 104-111.
3. Diez-Roux AV, Nieto FJ, Muntaner C, Tyroler HA, Comstock GW, Shahar E. Neighborhood environments and coronary heart disease: a multilevel analysis. *Am J Epidemiol*. 1997; 146(1): 48-63.
4. Winkleby M, Sundquist K, Cubbin C. Inequities in CHD incidence and case fatality by neighborhood deprivation. *Am J Prev Med*. 2007; 32(2): 97-106.
5. Kazda MJ, Beel ER, Villegas D, Martinez JG, Patel N, Migala W. Methodological complexities and the use of GIS in conducting a community needs assessment of a large U. S. municipality. *J Community Health*. 2009; 34: 210-215.
6. Diez-Roux AV, Merkin SS, Arnett D, et al. Neighborhood of residence and incidence of coronary heart disease. *N Engl J Med*. 2001; 345(2): 99-106.
7. Diez-Roux AV, Jacobs DR, Kiefe CI. Neighborhood characteristics and components of the insulin resistance syndrome in young adults: the Coronary Artery Risk Development in Young Adults (CARDIA) Study. *Diabetes Care*. 2002; 25(11): 1976-1982.
8. Diez-Roux AV, Kiefe CI, Jacobs DR, et al. Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiologic studies. *Ann Epidemiol*. 2001; 11(6): 395-405.
9. Gilboa SM, Mendola P, Olshan AF, et al. Comparison of residential geocoding methods in population-based study of air quality and birth defects. *Environ Res*. 2006; 101: 256-262.
10. Lovasi GS, Weiss JC, Hoskins RE, et al. Comparing a single-stage geocoding method to a multi-stage geocoding method: how much and where do they disagree? *Int J Health Geogr*. 2007; 6: 12.
11. Whitsel EA, Quibrera PM, Smith RL, et al. Accuracy of commercial geocoding: assessment and implications. *Epidemiol Perspect Innov*. 2006; 3: 8.

12. Taylor HA. The Jackson Heart Study: an overview. *Ethn Dis.* 2005; 15(Suppl 6): 1-3.
13. Carpenter MA, Crow R, Steffes M, et al. Laboratory, reading center, and coordinating center data management methods in the Jackson Heart Study. *Am J Med Sci.* 2004; 328(3): 131-144.
14. Fuqua SR, Wyatt SB, Andrew ME, et al. Recruiting African-American research participation in The Jackson Heart Study: methods, response rates, and sample description. *Ethn Dis.* 2005; 15(Suppl 6): 18-29.
15. Payne TJ, Wyatt SB, Mosley TH, et al. Sociocultural methods in The Jackson Heart Study: conceptual and descriptive overview. *Ethn Dis.* 2005; 15(Suppl 6): 38-48.
16. Taylor HA, Wilson JG, Jones DW, et al. Toward resolution of cardiovascular health disparities in African Americans: design and methods of The Jackson Heart Study. *Ethn Dis.* 2005; 15(Suppl 6): 4-17.
17. Wilson JG, Rotimi CN, Ekunwe L, et al. Study design for genetic analysis in The Jackson Heart Study. *Ethn Dis.* 2005; 15(Suppl 6): 30-37.
18. Krieger N, Waterman PD, Chen JT, Soobader M, Subramanian SV, Carson R. Zip code caveat: bias due to spatiotemporal mismatches between zip codes and US Census-defined geographic areas—The Public Health Disparities Geocoding Project. *Am J Public Health.* 2002; 92(7): 1100-1102.
19. Krieger N, Waterman PD, Lemieux K, Zierler S, Hogan JW. On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. *Am J Public Health.* 2001; 91(7): 1114-1116.
20. Hurley SE, Saunders TM, Nivas R, Hertz A, Reynolds P. Post office box addresses: a challenge for geographic information systems-based studies. *Epidemiology.* 2003; 14(4): 386-391.
21. ArcGIS [computer program]. Version 9.1. Redlands, CA: Environmental Systems Research Institute, Inc.; 2005.
22. US Census Bureau. TIGER[®] overview. <http://www.census.gov/geo/www/tiger/overview.html>. Accessed April 24, 2003.
23. Gorr WL, Kurland KS. *Learning and using geographic information systems: ArcGIS edition*. Boston, MA: Thomson Learning; 2007.
24. The Locator Map Company. *Jackson and surrounding areas*. 1st ed. Baton Rouge, LA: The Locator Map Company; 2004.
25. Certified Map Corporation. *Madison county map atlas*. Brandon, MS: Certified Map Corporation; 2000.
26. Certified Map Corporation. *Metro Jackson: Hinds & Rankin counties map atlas*. Brandon, MS: Certified Map Corporation; 2000.
27. Certified Map Corporation. *Metro Jackson map atlas*. Flowood, MS: Certified Map Corporation; 2004.
28. MapQuest Inc. Mapquest Web site. www.mapquest.com. Accessed Dec 10, 2006.
29. United States Postal Service. United States Postal Service Web site. www.usps.com. 1999-2007. Accessed Dec 10, 2006.
30. Barron WG Jr. Urban area criteria for Census 2000. *Fed Regist.* 2002; 67(51): 11663-11670.
31. Klassen AC, Curriero FC, Hong JH, et al. The role of area-level influences on prostate cancer grade and stage at diagnosis. *Prev Med.* 2004; 39: 441-448.
32. Vieira V, Webster T, Aschengrau A, Ozonoff D. A method for spatial analysis of risk in a population-based case-control study. *Int J Hyg Environ Health.* 2002; 205: 115-120.
33. Diez-Roux AV, Merkin SS, Hannan P, Jacobs DR, Kiefe CI. Area characteristics, individual-level socioeconomic indicators, and smoking in young adults: The Coronary Artery Disease Risk Development in Young Adults Study. *Am J Epidemiol.* 2003; 157(4): 315-326.
34. Krieger N, Chen JT, Waterman PD, Soobader M, Subramanian SV, Carson R. Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and

- childhood lead poisoning: The Public Health Disparities Geocoding Project (US). *J Epidemiol Community Health*. 2003; 57: 186-199.
35. Siffel C, Strickland MJ, Gardner BR, Kirby RS, Correa A. Role of geographic information systems in birth defects surveillance and research. *Birth Defects Res A Clin Mol Teratol*. 2006; 76: 825-833.
 36. Xiao H, Gwede CK, Kiros G, Milla K. Analysis of prostate cancer incidence using geographic information system and multilevel modeling. *J Natl Med Assoc*. 2007; 99(3): 218-225.
 37. Ward MH, Nuckols JR, Giglierano J, et al. Positional accuracy of two methods of geocoding. *Epidemiology*. 2005; 16(4): 542-547.
 38. Goldberg DW, Wilson JP, Knoblock CA, Ritz B, Cockburn MG. An effective and efficient approach for manually improving geocoded data. *Int J Health Geogr*. 2008; 7: 60.
 39. McElroy JA, Remington PL, Trentham-Dietz A, Robert SA, Newcomb PA. Geocoding addresses from a large population-based study: lessons learned. *Epidemiology*. 2003; 14(4): 399-407.
 40. Cayo MR, Talbot TO. Positional error in automated geocoding of residential addresses. *Int J Health Geogr*. 2003;2:10. www.ij-healthgeographics.com/content/2/1/10.
 41. Bonner MR, Han D, Nie J, Rogerson P, Vena JE, Freudenheim JL. Positional accuracy of geocoded addresses in epidemiologic research. *Epidemiology*. 2003; 14(4): 408-412.