Plotting Rural Households Where Map Details Are Insufficient: The Use of GPS in the Keokuk County Rural Health Study

ER Svendsen,* SJ Reynolds, C Zwerling, LF Burmeister, AM Stromquist, CD Taylor, JA Merchant

Keokuk County Rural Health Study, Department of Occupational and Environmental Health, University of Iowa, College of Public Health, Iowa City, IA

Abstract

The Keokuk County (Iowa) Rural Health Study (KCRHS) is a CDC-NIOSH-funded population-based, prospective cohort study, enrolling over one-fifth of the entire county population. Respiratory health and injury prevention in relation to environmental and occupational exposures are the primary focuses of this study. Health care delivery, geriatric, reproductive, and mental health are also measured. Because geographical distributions of health conditions within the study population are considered, a global positioning system (GPS) receiver has been used to geocode all rural households. The three categories of research questions that have been investigated with geographical information systems (GIS) are health care delivery, injury prevention, and health status. Within these, health care delivery questions measured the time and distance to participants' primary health care facility. Injury prevention measured crude injury rates, risk-taking behaviors, time, and distance to utilized emergency facilities. Health status measured hallmark health indicators such as tobacco use, drug and alcohol abuse, depression, and obesity. Demographic data have been collected and include age, sex, marital status, and socioeconomic status. Both crude and adjusted distributions have been performed. Medical screening and adult interviews were used as a source of GIS data. Continuous spatial distributions of the variables implicated within the study questions have been plotted in layers. Preliminary results indicate that alcohol consumption and abuse are uniformly distributed throughout the county. However, this is not the case for obesity, smoking, and reported injuries. These three seem to be clustered predominantly in the southwestern portion of the county. Further analysis is pending on the significance of these findings, and the health care delivery/injury prevention data. Through GIS analysis of the KCRHS data, the utility of GPS geocoding in rural community health and surveillance studies has been demonstrated.

Keywords: GPS, rural, Iowa

Introduction

Address matching of rural communities throughout the United States is very difficult, at best. Many addresses consist of box numbers or addresses that do not correlate with TIGER files. As a result, many rural geographic information system (GIS) applications are limited to aggregated data analyses. These methods are usually sufficient for

* Erik R Svendsen, Dept. of Occupational and Environmental Health, University of Iowa, College of Public Health, 100 Oakdale Campus, 241 IREH, Iowa City, IA 52242-5000 USA; (p) 319-335-4538; (f) 319-335-4225; E-mail: erik-svendsen@uiowa.edu

studies involving large geographic areas. For small areas such as counties, however, aggregated methods do not provide the resolution needed for small-cluster analysis. The Keokuk County (Iowa) Rural Health Study (KCRHS) has addressed this issue.

The KCRHS is a population-based, longitudinal prospective cohort study funded by the Centers for Disease Control and the National Institute for Occupational Safety and Health. The study has enrolled over one-fifth of the entire county population. Respiratory health and injury prevention are the primary focuses of this study. Health care delivery, and geriatric, reproductive, and mental health are also measured.

Method

In late 1997, the KCRHS began collecting global positioning system (GPS) coordinate data during its scheduled household site visits. This protocol was added to enable GIS analysis of the self-reported injury data collected in the study. Following its introduction, analysis of health care delivery and community health status data while controlling for confounding demographic variables including age, sex, marital status, and socioeconomic status was proposed. A nested GIS study had begun to emerge.

Within the three proposed GIS study areas of health care delivery, injury prevention, and health status, health care delivery questions measure the time and distance to participants' primary health care facility. Injury prevention questions measure crude injury rates, risk-taking behaviors, and time/distance to utilized emergency facilities. Health status measures hallmark health indicators such as tobacco use, drug and alcohol abuse, depression, and obesity. Medical screening and adult interview instruments used in the KCRHS contain all of these data.

The first round of data collection was completed in February 1998. As of this writing, only a small fraction of the homes within the study have been geocoded. The uncoded homes were scheduled for geocoding by GPS during the second round of data collection set for a two-year duration beginning September 1998. Due to the novelty of using GPS in geocoding rural households for GIS analysis on the county level, a preliminary study was undertaken for the purpose of presenting the technique at the third GIS in Public Health Conference in August 1998.

The preliminary study design incorporated a variety of plotting techniques to supplement the lack of GPS data. First, all households that were coded as residing within a town were geocoded within their town. Of the 454 town households, 326 were address matched. One hundred forty-nine (149) matches were performed directly with the Maptitude 4.03 GIS (Geonomics, Inc., Boston, MA) and 177 required manual matching assistance within the GIS. The remaining 128 were geocoded to the centroid of the town by using the locate-by-town function in Maptitude. Because the zip codes were not centered about a single town, plotting town residences at the zip code centroid was inadequate for this exercise. Rural and farm homes were not successfully address matched this way. Instead, GPS coordinates were taken by supplemental site visits for 440 households. These measures were taken at the mailbox for all measurements so that privacy was maintained. Fifty-eight (58) additional homes were plotted manually using the locate-by-pointing tool in Maptitude based on the known location. This was done using a locally produced basemap of the county homes and plats as a guide. Plotting homes manually may also have been performed by comparison with digital orthophotos now available nationwide at Microsoft's collaborative Web site, TerraServer

(http://terraserver.microsoft.com). Together, 498 (90%) of the 553 rural/farm households were plotted by GPS. When participants are considered rather than households, 904 (92%) of 987 rural/farm participants were plotted. In total, 1,563 (95%) of the 1,646 participants were plotted.

Accuracy of geocoded data is always a concern, especially when the study intends to perform continuous distribution analyses. It was approximated that the variation between the GPS recorded data and the actual household location data was within a quarter of a mile because no driveway exceeded that value. Data replication was not performed, so precision was not measured. GPS data were not updated by spatial correction factors, so the programmed error was still present. The Magellan 2000 GPS has an inherent error of within 50–100 yards, according to the manufacturer. Though not yet validated, it was assumed that the locate-by-pointing method was within the same quarter-mile accuracy constraints. Most towns were much less than a half-mile in diameter. Therefore, the majority of the town box households that were coded to the centroid are assumed to be well within the same quarter-mile accuracy range. The significance of these accuracy measures has not yet been evaluated.

Results

Once geocoded, the geographic data were analyzed using Distance Mapping and Analysis Program (DMAP) software (freeware available at http://www.uiowa.edu/~geog/health/index11.html). Obesity rates were presented for demonstration. A two-mile spatial filter was used to calculate the rates of obesity and the significance of the observed distribution. The denominator used was the observed county probability (0.42) of being obese within positive respondents. Preliminary results showed statistically significant clusters of obesity in the western portion of the county. Bands were overlaid to exhibit the two-mile spatial filter region of significance detected about the grid points. These were raw data that did not control for any confounding variables such as socioeconomic status. The intent of the demonstrated analysis was the illustration of the proposed GIS analysis of the final adjusted GPS database.

Conclusion

This initial study was not intended to produce rates or graphs for peer review publication. Rather, this study was intended to introduce the techniques available for plotting rural homes when address-matching capabilities are not available. Once a corrected and complete dataset of GPS coordinates is collected by this study in the next two years, further GIS epidemiologic studies will follow.

Preliminary results indicate that alcohol consumption and abuse are uniformly distributed throughout the county. This is not the case, however, for obesity, smoking, and reported injuries. Further analysis of the obesity data using DMAP statistical testing software has provided three regions of highly significant increased obesity rates. These three regions seem to be clustered predominantly in the western portion of the county. Through GIS analysis of the KCRHS data, the utility of GPS geocoding in rural community health and surveillance studies has been demonstrated.

Acknowledgments

G Rushton and the co-authors of the "GIS in Public Health" CD, The University of Iowa, Department of Geography