THE USE OF GPS AND GIS TECHNOLOGIES TO INCREASE THE QUALITY OF THE SAMPLE AND DECREASING SUPERVISION COSTS: THE EXPERIENCE OF THE PREHCO PROJECT

Alberto García Gurutxarri [1]

Hernando Mattei [1]

Ana L. Dávila [1]

Alberto Palloni [2]

María Larriuz [1]

---

1. Graduate School of Public Health. Medical Sciences Campus. University of Puerto Rico. P.O. Box 365067, San Juan, Puerto Rico 00936-5067
2. Center for Demography and Ecology. University of Wisconsin-Madison. 4437 Social Science Bldg. 1180 Observatory Drive. Madison, WI 53706. USA.

Abstract:

Fieldwork in social sciences research requires a considerable amount of time, effort and resources. Quality control procedures and supervision are inescapable components of this process that guarantee the integrity of the sample and the accuracy of the research. GIS and GPS technologies have been used for a multiplicity of purposes. This paper describes our experience using these affordable and expanding techniques in a creative manner to decrease supervision resources and costs, and to improve the quality of the sample, in addition to facilitate data linkages and complex spatial analyses.

Keywords:

Geographic Information Systems; GIS; GPS; fieldwork; sample; quality control.

1. Introduction: the PREHCO Project

The PREHCO (Puerto Rican Elderly: Health Conditions) Project is the largest study ever about the elderly population in Puerto Rico. About four thousand older adults (60 and over) and their spouses were interviewed about demographics, health conditions, housing, marital history, income, labor history, childhood conditions, race, functional status, mistreatment, transfers, family, properties, use of health services, sexuality and other. The study was conducted on a nationally representative sample of the elderly in Puerto Rico. The project is a joint venture between the Center for Demography and Ecology of the University of Wisconsin-Madison and the Graduate School of Public Health of the University of Puerto Rico, funded by US the National Institute on Aging (NIA) and supported by the Legislature of Puerto Rico and other agencies.

The sample for the study was selected using the 2000 Census data as the framework. Two hundred and forty eight sections were selected and then grouped in five strata and twelve substrata. The sections in this study average 90 households in size and were selected within census block groups. The fieldwork of the project was divided in two different stages: the complete enumeration of the 248 sections included in the sample and the interviewing of all eligible residents in those communities. Enumerators listed all the households in the section, determined the eligibility of each household and identified the older adults and their spouses for each of them. A short term after, interviewers visited and interviewed those eligible persons.

Enumeration is a critical stage in the process. The quality of the sample is largely dependent on the quality of the enumeration process. Consequently close supervision of the enumeration share is a must.

2. The GPS and GIS technologies

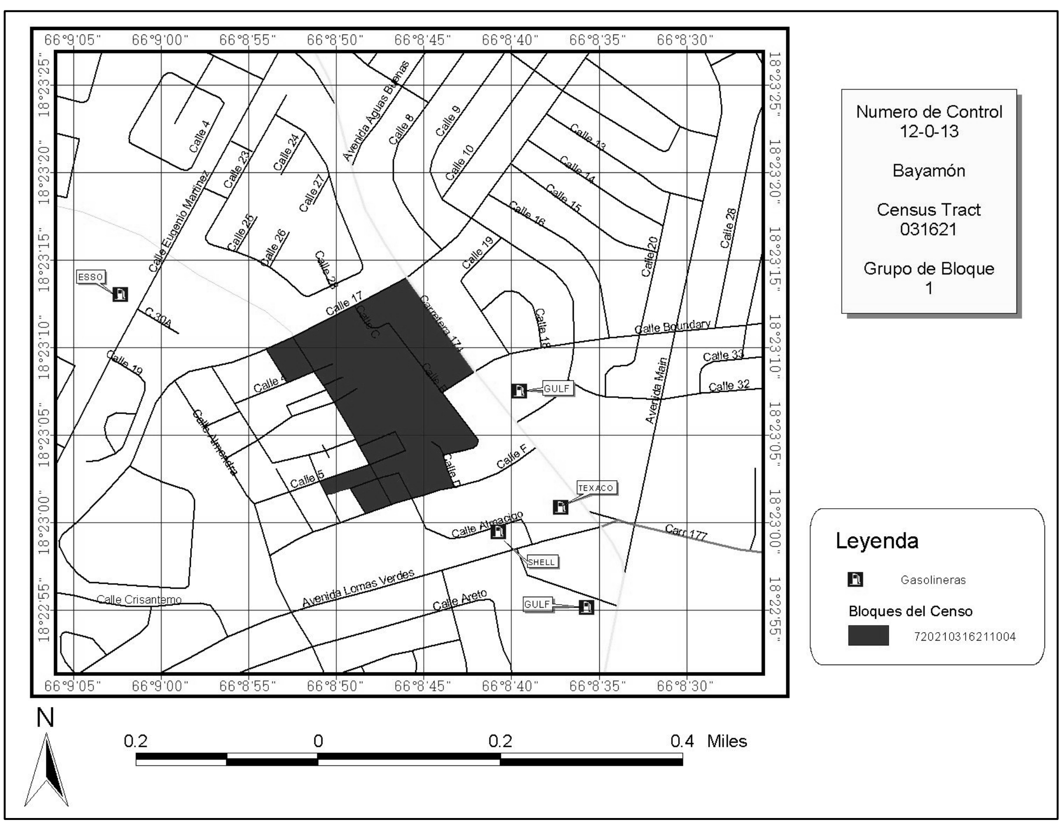
The Global Positioning System (GPS) technology is a satellite-based navigation system that has been in use for about forty years. It was designed for military purposes, it is being used for geology, navigation, farming, precision mapping, surveying, and additional applications are constantly growing. Pioneers, such as the Geographic Information and Analysis (GIA) unit at the Population Research Institute of PennState University and the University of North Carolina have promoted the use this technology in the social sciences field. GIA director, Stephen Matthews (2002) suggests that "the integrative and applied natures of GIS technologies make them attractive to many social scientists".

In simple terms, a GPS device can, in addition to many other information, give the user an exact position anywhere, anytime, and can store it for future use. Geographical Information Systems (GIS) allow the spatial integration of multiple data in a graphical environment. Even though the PREHCO study only used a small portion of the potential of this technology, we used it in a novel way to simplify the supervision process. The use of GPS and GIS provided us tools to: (a) help in the location of sample sections, (b) check that the sections enumerated are the sections in the sample, (c) determine that the location of the interview corresponds to the location of the enumeration, (d) project the sample points on electronic maps, and (e) relate the project database to existing or to future databases.

3. Requirements

In what follows we describe the requirements for using this technology. The GPS receivers: given the large number of available receivers in the market, we identified and considered a few selection criteria: precision of measurement, interface with the computer, price, simplicity of use, software compatibility, weight and waypoint storage capability. Having compared several models we chose a mid-level handheld device that complied adequately with the requirements for our study: cable connections with the computer, maximum error of measurement of 15 meters, recognized by the GIS software, affordable, lightweight (9 ounces), easy to use and with a storage capability of up to 500 waypoints. Electronic maps and software: electronic maps are available from the Census Bureau. By using the GIS software formatted hard copies of the maps were produced and the points recorded with the GPS receivers were projected onto the electronic maps. Paper maps: the enumerators used georeferenced printed maps (Figure 1). Those maps were created from the Census Tiger files and formatted using a commercial GIS software package. Maps included the identification of the census blocks in the section, the available landmarks, and the lat-long coordinates. Training: we trained fieldworkers (enumerators and interviewers) as well as supervisors. The training of fieldworkers on the use of GPS receivers took two hours whereas the training of supervisors on the use of receivers and software took.

Figure 1. Sample paper map. PREHCO Project 2003.



Census data: Census data of the sample sections is necessary, since data gathered in the field is compared with expected data as an additional criterion for quality control.

4. Procedure

The following tasks were carried out or facilitated using the GPS technology:

1. The enumerator located the section by using the printed map and the GPS receiver. When necessary, and to verify that the section to be enumerated is indeed the sample section, the position reported by the receiver was compared with the map coordinates for a given point. For hard to find sections, the receiver was used to locate the section (Figure 2). This feature was already highlighted by Nusser and Fox (2001).
2. The enumerator marked a waypoint for each house in the sample section.
3. Those stored points were transferred from the GPS receptor to the main computer by the supervisor, projected on a digital map and stored (Figure 3).
4. The interviewer marked a new point for each eligible house, interviewed or not.

5. These new points can then be projected on the same electronic map to verify that the enumerated and the interviewed houses are the same.

Figure 2. Sample section is located using the GPS and the printed map

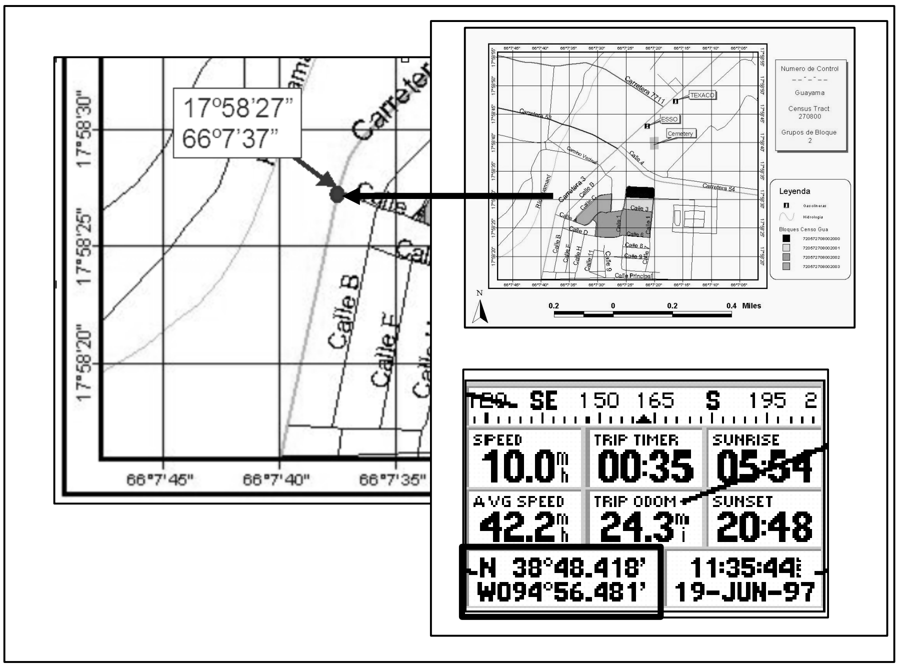
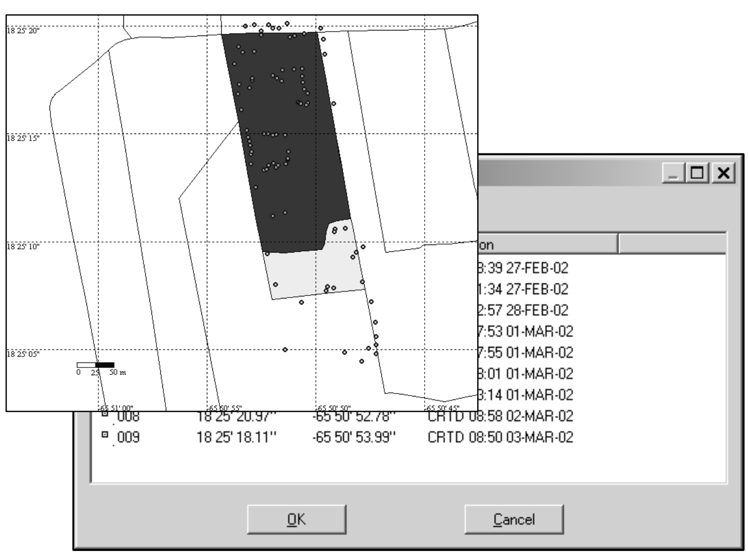


Figure 3. Waypoints are projected on a digital map and stored



5. Results

To realize the advantages of this approach in the fieldwork, we considered the following indicators: those related to the implementation process (the complexity of training and use, and the supervision process), those related to the quality of the sample (the number of out-of-the sample targets identified, sample points not enumerated and identified, and the percent of not found sample sections), those related to cost control (the savings in out-of-the-sample points not interviewed and the savings in supervision costs) and, finally, those related to the use of data (the process of electronic filing of sample points and location, the possibility to relate our database with other databases, the possibility of organizing the progress of the fieldwork, and the possibility of tracing the progress of the fieldwork).

5.1. Implementation

The process mainly involved the acquisition and setup of the GPS receivers, half day training for the supervisors and a two hours training for both enumerators and interviewers. Even though the receiver used had many capabilities, for the purpose of the project the procedure is straight forward. No enumerator or interviewer reported significant difficulties using the device. The problems reported usually were related to the involuntary repetition of a waypoint. Since the enumerator and the interviewer had to write down the waypoint number in the Household Control Card (HCC), repeated waypoints are easily identified and deleted in the supervision. The process of supervision included the transfer and verification of the waypoints from the enumerator or interviewer to the supervisor's computer. Because of the use of the GPS receivers no section sketches were needed for the enumeration, except for those covering a part of a block, which simplified the process significantly. Although the supervision process was very fast and reliable, throughout the fieldwork process we identified and implemented improvements regarding a more efficient filing system (see below).

5.2. Quality of the sample

The identification of households out-of-the-sample units is one of the most remarkable advantages from using GPS technology. The inclusion of households not in the sample implies: a decrease in the quality and the power of the sample and an increase of costs associated with interviewing those households. The waypoints recorded by the enumerator were projected on an electronic map of the section (Figure 3). The supervisor discussed with the enumerator the inclusion of those points outside the section borders, taking into account their position and perimeter, address information, and the description given by the enumerator concerning the topography and other details. Those factors were evaluated by the supervisor and a decision was made about including or excluding those points. Because of the precision of the receiver and other error sources, a buffer zone was considered. Deleted points were identified with a different symbol and kept on file for analysis purposes. Twenty five out of 248 (10%) sample sections were chosen to estimate the proportion of out-of-the-sample targets and spouses identified. An average of 7.65 households was identified to be out of the sections. The estimated average of targets and spouses for those households was 2.29 and 0.95 respectively. Even with an expensive supervision process, without the use of GPS technology most of those points would not have been identified, since their inclusion in the sample is usually related to a difficult identification of the sections boundaries. These results suggest a substantial increase in the integrity of the sample.

The identification of sample points excluded during the enumeration process complements the identification of points that are out of the sample sections, as explained above. The number of households and eligible households in the sample section was compared with the census data. Sections having a significant lower or higher number of households when compared with the census data and having an abnormal household distribution were reviewed and revisited. At this point, the results are difficult to evaluate. Our procedure for this task did not take full advantage of GPS technology, since waypoints were projected on an electronic map of census block polygons. In the future, the combination of the GPS information with geocoded satellite photography (IKONOS) for Puerto Rico and GIS software will permit easy identification of the households missed in the enumeration process by superimposing the photography, the section blocks boundaries, the roads and the waypoints. The households that have not been enumerated should appear on the image as not having a corresponding waypoint, along with their position and road access. The only limitation will be the updating of satellite images, a problem similar to that encountered when using censal materials during the intercensal period.

An additional difficulty in ensuring the integrity of the sample is the number of sample sections not found, especially in rural areas. This difficulty usually requires the substitution or deletion of those sections, which theoretically could be different from those easy located, in terms of important variables such as the accessibility to services, housing characteristics and economic indicators. As explained in the fieldwork procedure, the enumerator or interviewer used the GPS in combination with geocoded maps. That combination allowed the enumerator/interviewer to use the GPS to locate the section and confirm his position. This procedure virtually eliminated the possibility of not being able to locate sample sections. All the 248 sections in our study were located by the enumerators. Although the number of not located sections is usually low, this is an additional benefit derived from the use of GPS technology. Also, by using this strategy, 100% percent of the sections were to some extent checked for quality control.

In summary, the quality of the sample was increased mainly by identifying households not in the sample and locating difficult sections and, additionally, by identifying households in the sample not enumerated in the first attempt. No section was completed using the traditional strategy for comparison, although previous experiences suggest a significant improvement in the control on the quality of the sample.

5.3. Costs

Another clear advantage of using this technology is the savings involved. The cost analysis was performed by comparing the GPS strategy with the traditional strategy. The results are presented in Table 1 and explained below. The comparison was made for both the enumeration and the interviewing processes. The costs associated with the use of GPS technology were calculated adding the costs of receivers, training, batteries, software, quality control and enumerators fees. These were compared to the costs of a traditional strategy, in terms of supervision costs and enumerators fees, including the sections sketches. Since all fifty three receivers were used for both enumeration and interviewing, this cost was considered just once and it was applied to the enumeration costs. GPS devices are becoming not only more portable but also more affordable. A mid-level receiver, as utilized in our project, ranges from $275 to $300. Fifty three devices were used for a total of 248 sections and about 4,000 interviews. Electronic maps can be obtained for free from the US Census Bureau, and some of the software packages utilized are shareware or freeware (Trackmaker and Waypoint). A commercial GIS software single user license (ArcView) costs about $700. The cost of training was calculated by adding the stipends paid to the trainer and the trainees, corresponding to a half a day workshop for the supervisors and another two hours workshop for the enumerators. Given the simplicity of use of the selected GPS receiver, the training of enumerators was offered by PREHCO project personnel at the PREHCO facilities. An estimated average of one package of four AA alkaline batteries was used for each section. In the field quality control was performed for sections sharing two characteristics: a significant discrepancy in the number of households and older adults with the census data, and an abnormal spatial distribution of waypoints around the section.

Table 1. Comparison of fieldwork costs using the GPS strategy and a traditional strategy. PREHCO Project 2003.

ENUMERATION	
Estimated cost of the GPS strategy	**$43,400**
GPS receivers (53 receivers)	$15,000
Enumerators training on GPS	$400
Batteries	$1,500
Software	$700
Quality control (100% of sections quality controlled in some extent)	$1,000
Enumerators fees (no sketches needed)	$24,800
Estimated cost of the traditional strategy	**$49,760**
Full time supervisor (10 months)	$20,000
Enumerators fees (including sections sketches)	$29,760
Savings in the enumeration process	**$6,360**
Percent of savings	**12.78%**
INTERVIEWING	
Estimated cost of the GPS strategy	**$278,004**
Interviewers training on GPS	$400
Batteries	$1,000
Interviewers fees for targets (4,000 targets)	$220,000
Interviewers fees for spouses (41.51% spouses)	$16,604
Incentives for targets	$40,000
Estimated cost of the traditional strategy	**$315,875**
Interviewers fees for targets (4,000 targets)	$220,000
Interviewers fees for spouses (41.51% spouses)	$16,604
Incentives for interviewees	$40,000
Out-of-the-sample target interviewed (2.29 average per section)	$31,236
Out-of-the-sample spouses interviewed (0.95 average per section)	$2,356
Out-of-the-sample incentives for targets interviewed	$5,679
Savings in the interviewing process	**$37,871**
Percent of savings	**11.99%**
TOTAL SAVINGS	**$44,231**
PERCENT OF TOTAL SAVINGS	**12.10%**

Using these criteria, only ten out of 248 sections were identified as needing additional quality control in the field. For those sections this implied a duplication of the original enumeration costs. Without using GPS technology, at least one full time supervisor would have been hired during the enumeration process and a minimum of 38 sections (about 15%) should have received quality control for enumeration. In addition, sketches for every section would be necessary (an estimated increase of 20% in the enumeration fees). As shown in Table 1, a 12.78% savings in the enumeration costs was obtained. However the main savings come from the interviewing process, since it is the largest part of the fieldwork budget. The identification of the targets and the spouses out-of-the-sample led to savings of 11.99% of the total interviewing costs. The savings from the enumeration and the interviewing processes add up to a 12.10% of the total fieldwork costs. Although the supervision of interviewing techniques and procedures cannot be replaced using this technology, additional savings related to the supervision of interviewing have to be also considered.

5.4. Use of data

The sample waypoints were stored both on GIS type files and spreadsheets. This allowed a more effective organization of the interviewers work, easy tracking of the fieldwork progress and, when finished, the possibility of linking our database with any other geocoded database. The geocoding of our database will permit the spatial analysis of virtually any variable in the PREHCO survey, and the geographical relationship of our variables with other external geocoded variables. A good filing procedure is necessary to avoid the problems related with the proper identification of waypoints and interviews. A careful plan has to be established in advance including the details for setting up receivers, the units of measurement and geographical datum, files structure and labeling, waypoint symbols to be used, etc.

6. Conclusions

The results of our experience suggest clear benefits derived from the use of GPS technology: a decrease in sample losses and an increase in the quality of the sample, a reduction in the supervision costs, as well as expansion of basic quality control to the whole sample. Simplicity of training and use are additional advantages. The results also confirmed the reliability of this technology.

Special attention should be given to the potential associated with availability of geocoded data. According to Yasnoff and Sondik (1999) geocoding is a “near-universal link for sorting and integrating records from multiple information systems into a more coherent whole”. The number of geocoded databases is growing rapidly. The objective 23-3 of the of the US Department of Health and Human Services Healthy People 2010 project (2000) indicates that the proportion of public health data systems that use geocoding should be increased to 90% by 2010. This means that data from PREHCO and other health related surveys could be matched with other information and analyzed in terms what has been called “spatial patterns of disease” (Bui, 2002). This feature not only simplifies existing analyses, but the linked data sets make possible analyses that previously were unfeasible (Yasnoff and Sondik, 1999).

The coordination of geocoding efforts within a geographical area will multiply the usefulness and applications derived from the individual survey results.

When using the technology issues related to confidentiality should be considered. This concern was highlighted in the Healthy People 2010 objectives as the challenge of increasing public access to GIS information systems without compromising confidentiality. Richards et al. (1999) pointed out that “statistical and epidemiological methods need to be developed to protect individual and household confidentiality”, especially when linking several databases with apparent individual safeguards for confidentiality.

In summary, basic use of GPS and GIS technology, with low training requirements and a modest investment, can serve to improve the integrity of the sample, to reinforce the quality control procedures, and to reduce the supervision costs while providing a valuable tool for relating multiple databases spatially.

7. References

Bui, E. (2002). The potential of linking environmental, socio-economic and health data for health policy development. Abstract submitted to the Symposium on Health Data Linkage, March 20-21, 2002. Sydney, Australia.

Matthews, S.A. (2002). Introduction to Geographic Information Systems (GIS). GIS Resource Document 02-01 (GIS-RD-02-01) from the GIS Service Core, Social Science Research Institute, Pennsylvania State University Web site: www.pop.psu.edu/gia-core.

Nusser, S.M., and Jean Fox (2001). Using digital geospatial information to locate sample units in the field. Proceedings of the Federal Committee on Statistical Methods 2001 Research Conference, Washington, D.C.

Richards, T.B., Charles M. Croner, Gerard Rushton, Carol K. Brown, and Littleton Fowler (1999). Geographic Information Systems and Public Health: Mapping the future [electronic version]. Public Health Reports 1999, 114: 359-373.

US Department of Health and Human Services (2000). Healthy People 2010 Objectives. Vol. II, Second Edition, Objective 23-3. Office of Disease Prevention and Health Promotion. Washington, DC.

Yasnoff, W.A., & Edward J. Sondik (1999). Geographic information systems (GIS) in public health practice in the new millennium [electronic version]. Journal of Public Health Management and Practice, July 1999, Vol. 5, No. 4.