Using GIS To Improve Field Interviewing Efficiency: Enhanced Interviewer Selection and Sample Allocation

Ned English and Steven Pedlow
National Opinion Research Center, 55. E Monroe St, Chicago, IL, 60603

Abstract

Two tasks that must be completed in advance of a successful field-based interviewing study are the hiring of field staff and the appropriate assignment of sample to be fielded. Ideally, interviewers will be recruited that are situated geographically proximate to a convenient quantity of sample for maximum productivity. In addition, the assignment of field sample cases, usually in the form of clustered housing units, should consider two factors: the relative distance to potential field interviewers as well as the balancing of interviewer loads. Sample allocation thus becomes an optimization task in theory, a process that weighs travel distances and sample quantities for each field interviewer.

We used a programmatic GIS-based technique for field interviewer hiring and sample assignment; this paper discusses only the latter. It is novel in that it departs from the traditional, decentralized method of hiring and assignment by each field manager. GIS was utilized in order to integrate distance and proximity with traditional database processing, and a programmatic method was employed to optimally allocate clusters. The authors focus on empirical results from two NORC surveys: the *Survey of Consumer Finances 2004* and the *National Longitudinal Survey of Youth 1997 Round 8,* which demonstrated the utility of such techniques during the summer of 2004. It is hoped that the methods described herein can be used and enhanced for similar surveys and field interviewing environments.

Keywords: GIS, case assignment, field surveys

1. Introduction

Fundamentally, all face-to-face surveys require the hiring of field staff and the subsequent assignment of sample in advance of field interviewing. District-based field managers have traditionally been responsible for executing this process on their own. The manual exercise of traditional assignment-making has been challenged, however, by particular structural limitations. Such complications are due to the use of obsolete technology involving manually matching cases to appropriate interviewer by ZIP code. It would thus be desirable to develop an automated method of case assignment that could consider distance and proximity within a database structure. This type of solution can be facilitated through the use of GIS technology.

Geographic Information Systems or 'GIS' embodies a set of tools that permits the automatic linkage of data sets based on geographic variables, such as distance, adjacency, or the quality of being within given census or postal areas. Through the use of GIS, one can calculate distances and optimize sample allocation based on project-specific parameters. A necessary factor would then be to know *a priori* the addresses of potential field interviewers as well as sample address locations.

The purpose of this paper is to present a GIS-based method to facilitate the assignment of sample to staff, based on distances calculated from geocoded addresses. Data are presented from the *Survey of Consumer Finances 2004* and the *National Longitudinal Survey of Youth 1997,* two NORC field surveys. We argue that automatic case assignment using calculated distances as presented herein is preferable to the *status quo* due to demonstrated gains in efficiency. We discuss how the process should be undertaken iteratively so as to include the specialized area-specific input of field managers. In addition, we present the results of a survey of field staff who were involved as well as a summary of the outcomes from our approach. We contend that automated case assignment is an example of the kind of survey research methods that can be enhanced using geographic information systems.

2. Background and Problem

At issue are two tasks related to field survey operations, these being the hiring of field staff and the allocation of sample to the appropriate personnel. Field managers have typically been responsible for each, which have required using spreadsheets to matching candidates to sample by city and ZIP code. The primary goal of this process was to find interviewers who were within the same city as sampled segments. If a candidate was not in the same city, similar ZIP codes were used to ascertain approximately

how far they were from a case or a segment. Once interviewers were hired, a similar process was followed to assign cases to the interviewers, again by city and ZIP code using sets of spreadsheets.

While the traditional approach to interviewer hiring and sample allocation has become the de-facto method, it does carry some notable drawbacks. Firstly, the traditional method can be time-consuming and error-prone as the user iteratively manipulates the separate spreadsheets in order to distil the necessary information. It can also be difficult to visualize the geographic distribution of cases and staff in a static list format, especially if the user is not familiar with the geographic area in concern.

A second issue with the traditional method is that it introduces inaccuracies by treating cases as a list of cities and ZIP codes, and so ignores common geographic complexities. For example, the traditional method could not handle the possibility of cases in disparate cities being more proximate than those in the same city. The third drawback of the traditional method is that changes over time are not easily incorporated into the model of interviewers and cases, and require iterations of manual re-allocation. An automated method that would be capable of calculating true distances between multiple sets of addresses and could permit automatic updates would thus be beneficial. The necessary theoretical framework and set of tools to do this is GIS.

GIS is defined as a combination of spatial database management and spatial analytical tools, along with computerized cartography, used to facilitate the accumulation and manipulation of geocoded objects (Goss 1995, Bond and Devine 1991). GIS has commonly been employed in social science research to link attribute information to spatial or geographic information, and has considerably increased the capacity for data processing in recent years (English and Murphy 2004). If GIS software could be programmed to ascertain the locations of interviewers and segments and then perform the necessary distance calculations, it could represent the necessary solution to automated case assignment. Essentially, we propose using distance as an analog or replacement for the traditional method.

The field of automated case assignment is not well-represented in the literature. While there has been considerable research in computer assisted personal interviewing (CAPI) technology, there hasn't been in optimizing case assignments. We feel this is a new avenue with considerable research potential.

3. Methodology

Two NORC surveys fielded in 2004, the *Survey of Consumer Finances 2004 (SCF '04)* and Round 8 of the *National Longitudinal Survey of Youth 1997 (NLSY '97),* provided the opportunity to develop a new case assignment approach. The *SCF '04* was our first attempt at using the automated case assignment procedure to assign area probability cases. This paper will focus on the case assignment process rather than that for interviewer hiring in both surveys.

3.1 The Survey of Consumer Finances

The first step in the process was to acquire two address lists, these being a list of 190 field interviewer addresses and a list of the sampled case addresses. *SCF* '04 was a dual-frame survey, with a traditional area-probability (AP) sample of 5042 cases within 877 clusters that was supplemented by 5047 non-clustered "list sample" cases. We assigned list cases separately to interviewers but assigned AP clusters to the same interviewer. Therefore, our list of "cases" to be assigned was 877 + 5047 or 5924 entries.

We then geocoded both the interviewer and case addresses using the software package *MapMarker Plus*. "Geocoding" is a process by which geographic coordinates in the form of longitude and latitude are appended to each address, and so places the addresses in geographic space.

Once we had the address lists geocoded, we used a program written in the *MapInfo Professional*–based scripting language *MapBasic* to calculate the ten closest field interviewers from each case, "as the crow flies". Cases were then assigned to the closest interviewer in this construction as a "first shot" at case assignment. We also included project-specific decision rules in the program's logic, which entailed not having a maximum distance tolerance, not having a maximum number of cases per field interviewer, and not constraining cases by region.

Essentially, for our first attempt at case assignment we gave the closest case to the closest field interviewer with no region constraint and no attempt to balance the sample a-priori. The second key step was to have a feedback loop where the field managers could suggest changes based on their local knowledge. This step was important in that it permitted the input of traditional knowledge, such as what specific field interviewers were most appropriate for a specific city, or which interviewers were specifically hired for list-sample cases, to override the distance-based assignment. A

final edited version was then generated and sent to the field for production.

3.2 National Longitudinal Survey of Youth 1997

Soon after the *SCF '04,* we implemented automated case assignment in a second field survey, this being round 8 of the *National Longitudinal Survey of Youth 1997* (*NLSY '97)*. Our methodology was similar to the *SCF '04* in that we began with lists of field interviewers and case addresses. Since *NLSY '97* was the seventh round of a longitudinal sample, it only had a list sample of *8913* cases which were geocoded in the same manner as the *SCF '04*.

Automated case assignment for the *NLSY '97* differed in that stricter requirements were placed on the programming logic. Specifically, cases were restricted to be assigned to an interviewer within the same field region within 200 miles of the assigned field interviewer.

3.3 Survey of Field Managers

While the project requirements were met in both the *SCF '04* and *NLSY '97* with regard to case assignment, we wanted to learn more about how useful the automated case assignment actually was from the perspective of field managers, those traditionally responsible for assignments. Moreover, we wanted to learn what changes and improvements should be made in the future. To these ends, we designed a survey and sent it to all *SCF '04* and *NLSY '97* field managers. Survey questions were as follows:

1. *Were you given an automated case assignment for the (SCF '04, NLSY'97)?*
2. *If q1 = yes, how many hours would you estimate you spent revising the automated case assignment?*
3. *If q1 = yes, how many hours would you have spent without the automated case assignment?*
4. *If q1 = no, how many hours did you spend making case assignments by hand?*
5. *How would you describe your experience with the new automated assignment information 1- Very Positive to 5- Very Negative?*
6. *How easy was it to understand and implement 1- Very easy to 5- Very Difficult ?*
7. *Would you support future automated case assignment 1- Completely Support to 5- Not Support at all*
8. *What changes would you make to the procedure for automated case assignment?*

4. Results and Discussion

Figure 1 in Appendix A shows the number of cases assigned per field interviewer in the *SCF '04* using the automated case assignment prior to any input from the field. It is clear that most interviewers were assigned a reasonable caseload, e.g. fewer than 100 cases, with the mode being approximately 30 cases. Some interviewers were assigned more than this number, which was the generally result of being the only interviewer in a geographic area with a large number of cases.

Figure 2 shows the same distribution, but for the *NLSY '97*. The mode on figure two is shifted to the right when compared to the *SCF '04*, being approximately 70 cases instead of 30. The difference is that there were fewer interviewers and more cases on *NLSY '97* compared with the *SCF '04*. Nonetheless, assigned cases in NLSY '97 were still reasonable when considering distances.

Figure 3 shows the distances between field cases and their assigned interviewers in miles for the *SCF '04*, and Figure 4 shows the same for the *NLSY '97*. The *SCF '04* data has a median of 11.05 miles, and flattens considerably after 75 miles. Because of its longitudinal nature, the *NLSY* '97 had higher distances between cases, with an approximate median of 17.43 miles. This is a result of the longitudinal survey design, as cases were assigned the previous interviewer even if either party moved, as illustrated graphically in figure 4.

Considered together, Figures 2-4 show that the automated case assignment procedure can assign reasonable caseloads at distances that are considered feasible. Rural or suburban areas were simpler to assign, as there were fewer cases or interviewers and less clustering of each. Urban areas sometimes had instances were one interviewer was assigned too many cases and other nearby interviewers were assigned too few. This effect notwithstanding, automatic assignment is more useful in urban areas than suburban or rural areas due to the larger quantities of cases and interviewing resources to deal with. The question is, then, how did the field staff, specifically the field managers, feel about the process.

Table 1 summarizes the major survey results regarding the questions 'overall experience', 'ease of understanding', and 'would use in the future' . The response rate for this survey was 85%, as 17 out of 20 interviewers that were still on staff responded. Median values for the 'overall experience' and 'ease of understanding' were 2 meaning 'support', whereas for

'support in future' was 1 meaning 'strongly support'. So, the survey results demonstrate a generally positive experience and strong support for the future. The open-ended questions pointed to a desire for more iterations between the field and the central office to permit more input.

One telling outcome from the survey was an estimate for time savings for the field managers using the automated case assignment vs. implementing the traditional approach. The median hours spent with the automatic assignment was 4.0 with and estimated 9.5 without, which translates to 5.5 hours of savings on average across all field managers. If one looks at the time savings at the aggregate level, counting the time spent automating the process against the field manager savings, it was 2.1 hours per manager on the *SCF '04*, or 25.2 hours total for 12 managers. For the *NLSY '97*, it was 6.3 hours per manager or 63 total hours for 10 managers. These savings were greater for the *NLSY '97* because of the experience in improving the process with the *SCF '04,* and the fact that we had already developed most of the necessary computer code.

5. Conclusions

We argue that our experience with automated case assignment shows it is an effective analog or replacement for the traditional method, with the advantage of added efficiency. Consequently, we feel automated case assignment is worth continuing.

There are a number of improvements, however, that should be pursued. Firstly, it would be valuable to calculate distance by routing along a street network instead of using spherical distance "as the crow flies". Secondly, developing automated caseload balancing as an addition would save time. A third major improvement would be to include variables of interest besides distance, such as language, race/ethnicity, or experience. One would then conduct a fuzzy match across all variables to match the best interviewer to a case.

References

Bond, Derek and Paula Devine (1991), "The role of geographic information systems in survey analysis," *The Statistician,* 40, pp. 209-215.

English, Ned and Whitney Murphy (2004), "Discovering Rare Urban Populations for Community Health Surveys: Area Probability Vs. GIS-Based Telephone Sampling Approaches," 2004 Proceedings of the American Statistical Association, AAPOR Survey Research Methods Section [CD ROM], Alexandria, VA: American Statistical Association: pp 4773-4780.

Goss, John (1995), "We know who you are and we know where you live: The instrumental reality of geodemographic systems, *Economic Geography*, 71 (2): pp. 171-198.

Appendix A Tables

Figure 1- Frequency distribution of SCF cases per field interviewer

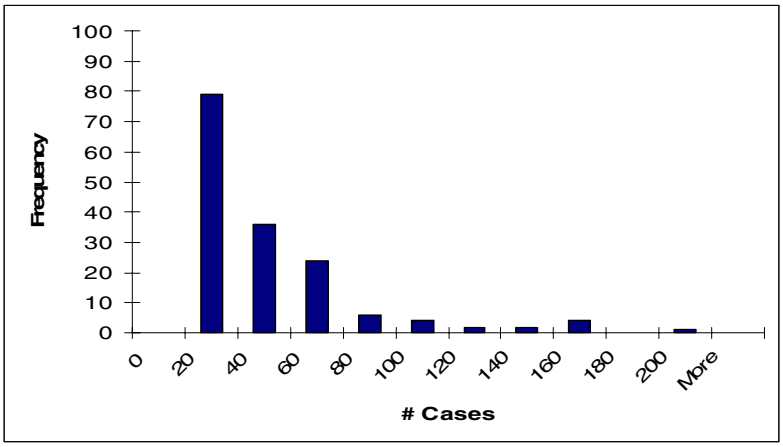

Figure 2- Frequency distribution of NLSY cases per field interviewer

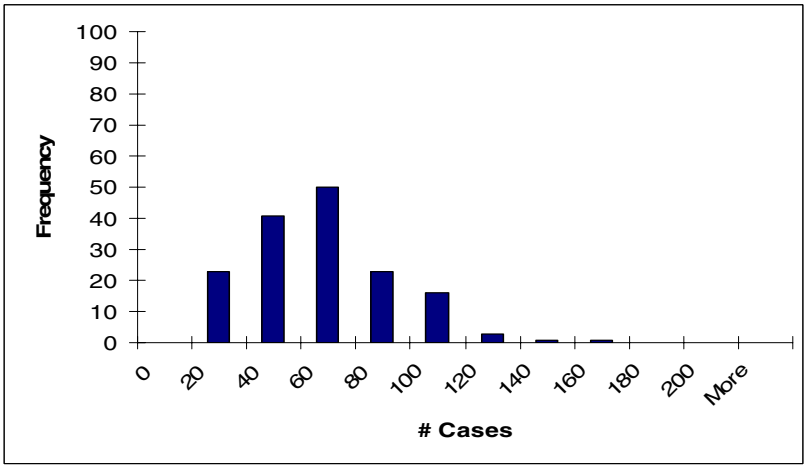

Figure 3- SCF Distances per Case

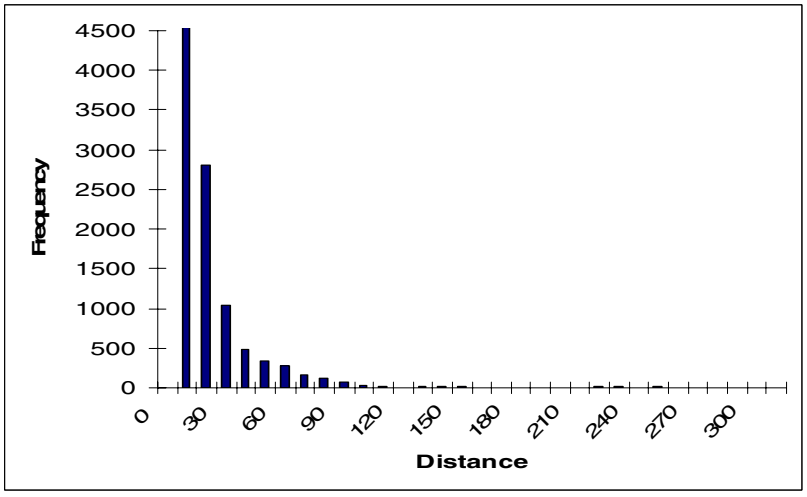

Figure 4- NLSY Distances per Case

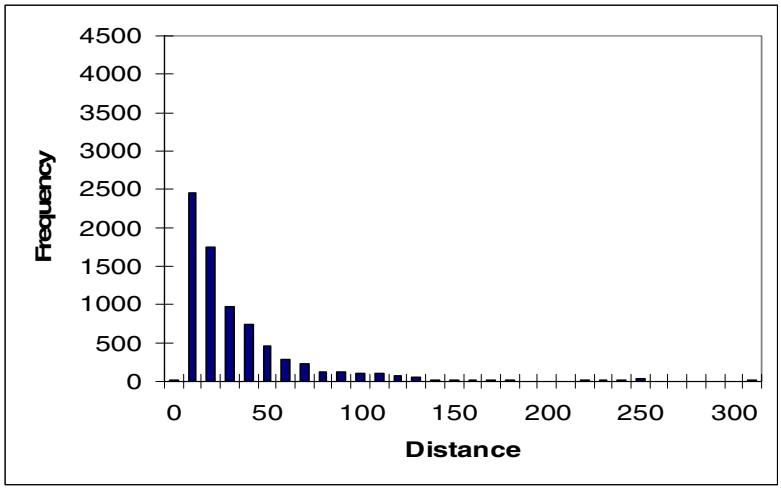

Table 5- Survey results

	Degree of Support						
	Highest				*Lowest*	*Median*	*Mean*
Question	***1***	***2***	***3***	***4***	***5***		
Overall Experience (16 responses)	5	5	3	3	0	*2*	*2.3*
Ease of Understanding (17 responses)	7	7	2	1	0	*2*	*1.8*
Support in future (17 responses)	9	2	3	3	0	*1*	*2.0*