METHODOLOGY FOR A COMMUNITY-BASED KEY INFORMANT SURVEY[1]

Cynthia Veldman, Sara Zuckerbraun, Stefani Schneiderman, The National Opinion Research Center
Cynthia Veldman, NORC, 1155 East 60th St., Chicago, IL 60637

Key Words: Survey Research Methods, Key Informants, Community Leaders

INTRODUCTION

This paper describes the challenges involved in building a methodologically-sound sample of community-based elites by describing the multi-stepped sample construction for the Neighborhood Expert Survey. The survey interviewed key community leaders associated with eighty neighborhoods in Chicago in order to examine the institutional resources available to, and the elite social networks operating within, these communities.

SURVEY DESIGN

Definitions. The sample design for the Neighborhood Expert Survey reflects the three means by which respondents were defined to be key informants for a community. Targeted informants were defined by the nature of **who** they were, **what** they did and **where** they were located. "Who" related to a respondent's typology. Respondents were designated as *Positional* informants by virtue of the **positions** they hold in important community institutions. Or, respondents were designated as *Reputational* informants, because of their **reputations** as being very knowledgeable about specific issues important to the neighborhood or about the neighborhood itself.

"What" referred to the *Domain*, or sphere of professional or institutional activity, with which they were associated. The domains identified as important at the community level were education, religion, business, politics, law, and community-based organizations, as well as a seventh, non-professional, "lay-expert" domain. (The seventh domain was to include such lay-experts as the "revered and respected 92-year-old" resident and the gang leader nominated as Reputational informants).

"Where" linked respondents' institutions with one of the eighty *Neighborhood Clusters* in the project's geographical sample. The geographical linkage was determined either by the institution's physical location (e.g. schools, churches, businesses, community organizations) or its jurisdiction (police districts, political wards). In the former case, the institution was distinctly identified as being located in the Neighborhood Cluster itself, or in the larger Community Area within which the Neighborhood Cluster was embedded. The Neighborhood Clusters are eighty small neighborhood areas, composed of single to small groups of census tracts, preselected by the client on SES criteria. The larger Community Areas in which they are located, and about which the questionnaire focused the respondent's attention, are forty-seven of the seventy-seven *Community Areas* established by the Chicago Planning Commission in the 1930's. The names of these Community Areas are ubiquitously used, and generally well-known, by those who live and work in the city.

Dual Sample Frames. Positional respondents came from a list sample, derived from several current, non-biased, readily available professional and institutional lists, with the exception of the law domain, where the sample was obtained systematically from the twenty-five Chicago neighborhood relations sergeants. Reputational respondents came from a snowball sample generated by nominations given by respondents in response to seven questions which elicited such information in the questionnaire.

Original Design Goals & Priorities. (Table 1). NORC originally planned to interview 3,440 community leaders: seven key informants in each of the six professional domains, and at least one 'lay expert,' in each of the eighty Neighborhood Clusters. As illustrated in Table 1, these design goals were structured to meet three client priorities: even geographical distribution among the eighty Neighborhood Clusters; even domain distribution, among the six primary domains, and even respondent-type distribution. These prioritized goals translated into a targeted geographical distribution of 43 interviews in each of the eighty Neighborhood Clusters, and a targeted Domain distribution of 560 interviews within each of the six professional areas. Equal Respondent-Type distribution was an overall design goal, <u>not</u> operationalized at the domain or Neighborhood Cluster level, but aimed at completing 1,720 interviews each with both the Positional and Reputational sample members.

Strategies. In order to meet these distribution goals

[1]The authors thank Rob Sampson, Norman Bradburn, Nick Holt, Woody Carter, Jody Dougherty, Ron Dorsey, and Phil Panczuk for their valuable roles in, and contributions to, this survey. We also thank the National Institute of Justice and the John D. and Catherine T. MacArthur Foundation, sponsors of the Project on Human Development in Chicago Neigh-borhoods, a component of which is the Neighborhood Expert Survey, the focus of this paper. The project is designed by a multi-disciplinary research team from the Harvard School of Public Health, Yale University, and the University of Chicago.

Table 1: DESIGN

Goals	Original Design 7 leaders x 6 domains x 80 neighborhoods + 1 'lay expert' x 80 neighborhoods = 3360 + 80 = 3440 interviews	Revised Design 5.5 leaders x 6 domains x 80 neighborhoods + 2 'lay experts' x 80 neighborhoods = 2640 + 160 = 2800 interviews
1: Equal Geographic distribution	Original Target Completes 43 informants per neighborhood	Revised Target Completes 35 informants per neighborhood
2: Equal Domain distribution	560 informants per domain	440 informants per domain
3: Equal Respondent-Type distribution	1720 positionals/1720 reputationals	Mixture in each neighborhood with more positionals than reputationals overall

distribution goals, and to capture more perspectives from the smaller Neighborhood Clusters than from the larger Community Areas, NORC developed and pursued two strategies for releasing cases to be interviewed.

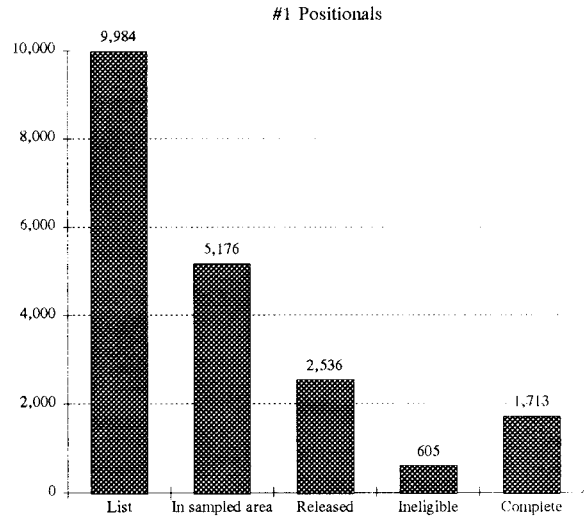
To achieve the domain and geographical distribution goals, we released only enough cases considered necessary to obtain the targeted number of completed interviews in each of the "cells;" (i.e. in each of the six domains within each of the eighty Neighborhood Clusters.) With an initial target of seven completes and an 85% response rate goal, this required the release of eight or nine cases in most cells.

To obtain perspectives bounded to the smaller Neighborhood Cluster level as far as possible, we released cases in the following order: (1) Positional sample members whose institutions were physically located in the Neighborhood Cluster; (2) Positionals located in the incubating Community Area near the Neighborhood Cluster; (3) Reputational cases whose names were generated by respondents located in the Neighborhood Cluster; and (4) Reputationals generated by respondents located in the Community Area but near the Cluster.

Revised Goals. Upon operationalizing the original sample design, we found the costs associated with sample preparation to be substantially more than anticipated due to the constraints of our sample design which were amplified by the high rate of geographical and professional mobility of our elite sample. Consequently, we were required to revise our overall goals to target 2,800 completed interviews, instead of 3,440. Despite the reduction in the overall scope, we continued to strive for even domain and geographical distribution throughout the survey. Table 1 also displays the revised targeted distributions per Neighborhood Cluster (thirty-five informants) and professional domain (440 informants) which resulted from the reduced scope.

POSITIONALS

Positional Sample Overview. Graph #1illustrates the steps in building the Positional sample. We began by making a list sample, obtaining names from fifteen publicly-available lists or publications. Examples of some of the lists we used are the Directory of Public Schools (Board of Education of Chicago, 1995), the World Chamber of Commerce Directory (1994), and the Directory of Community Organizations in Chicago (Institute of Urban Life, 1994).



We used lists that met two criteria: (1) they could be no more than 1 year old, and (2) they had to be as unbiased as possible, meaning an organization did not have to pay or be affiliated with the publisher, to be listed. All of the lists included institution name, address and phone number. When names were not available on the lists, a pre-field locating effort telephoned those institutions in the sample to obtain names.

Our insistence on high quality, unbiased lists determined, to a large extent, the types of Positional respondents in the survey. The business domain, for example, is not exhaustive. The lists of businesses we found were all unsatisfactory in some way. For instance, one list only included companies which were owned in Chicago, excluding important companies such as Dommicks and Walgreens; one list collected only corporate headquarters, and one list was too extensive by including every single storefront in the city. Rather than bias our sample by using an unsatisfactory list, we excluded some businesses from our Positional sample in favor of gathering them through the

Table 2. Positional Informant Categories

Education:
Public/Private High School Principals;
Public/Private Middle School Principal;
Public/Private Elementary School Principal;
Local School Council President;
Private Special-needs School Principal

Business:
Community Reinvestment Officer (Banking);
Realty Company Owner/Manager;
Community Newspaper Editor;
Chamber of Commerce/Business Assoc. Pres.

Religion:
Catholic or Orthodox Priest;
Protestant Pastor; Mosque Iman;
Synagogue Rabbi; Other Religious Leader;

Law Enforcement:
Beat Officer; Sector Sergeant;
Neighborhood Relations Sergeant;
District Commander

Political:
Alderman; Ward Committeeman;
State Representative; State Senator

Community Organization:
Community Development Assoc. Director;
Ethnic Organization President;
Social Service Agency Director;
Youth Agency Director;
Community Center Director;
Vocational Training Center Director;
Branch Librarian;
Neighborhood Housing Org. Pres.;
Clinic or Hospital Director

Reputational sample. The Reputational nominations in business did in fact capture the wide variety of community-anchored local businesses actually desired. Similarly, post-secondary school personnel who were not included in the Positional sample became abundant Reputational nominees. Ultimately, we obtained high quality lists for each Positional category displayed in Table 2.

Attrition through Geographical Design. Returning to Graph #1 and the Positional sample building process, the first bar indicates that the original list sample included almost 10,000 names. Since the underlying lists or publications had a city-wide focus, we next had to determine which of these 10,000 institutions were physically located in, or had jurisdiction over, the sampled neighborhoods in order to assess their eligibility for the survey. If eligible, we then had to assign each to a unique Neighborhood Cluster and/or Community Area. We used the following two locating and assignment methods.

(1) We entered the address of the institution into a CD-ROM Tiger Program called The Census Tract Locator. The program identifies the census tract associated with that address. Using a database that linked census tracts to their related Neighborhood Clusters and Community Areas, each institution was then precisely assigned geographically.

(2) For the political and law enforcement domains, we used ward and police district maps in conjunction with detailed street maps containing the neighborhood boundaries to identify the Community Areas and Neighborhood Clusters within their jurisdiction. When distinct assignments were not possible, we established a set of replicable decision rules to determine assignments.

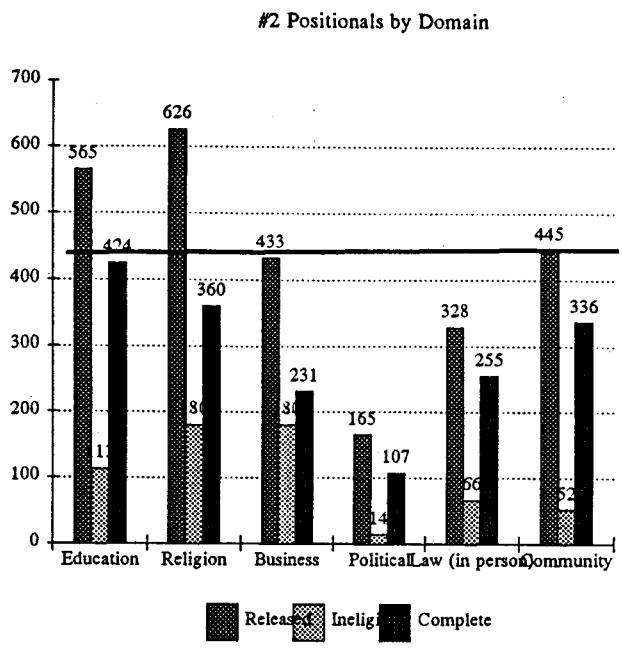
When the locating and assigning processes were complete, only 5,176 potential respondents, roughly half the original list sample, were actually located in, or had jurisdiction over, either the Neighborhood Clusters or Community Areas. This is illustrated by the second bar on Graph #1.

De-duplication. Following locating and assigning, we instituted a process called *de-duplication* -- a process we grew very familiar with on this survey. There were many names in our database that appeared more than once. Very frequently, elected city and state government officials (Chicago alderman, Illinois state representatives and senators, Congressmen) also hold the position of their respective party Ward Committeeman. In addition, many principals were principals of more than one school, CRA officers often functioned as the CRA officer for several branches of the same bank, and newspaper editors sometimes edited numerous papers in different communities. It was not uncommon for domains to overlap as well. For example, a pastor (in the religion domain) could also hold the position of principal in a religious school (in the education domain), and/or Executive Director in a Community Service Organization (community organization domain). Since each person could only be interviewed once, it was necessary to identify and delete such duplicate cases from the database.

Because of spelling and typing variation, automated computer searches for duplicate entries were not possible. Instead, we developed a computer driven, yet manual process of de-duplication which identified repeated last names, organization names, addresses, and phone numbers.

Random selection. Given our strategy to release only nine cases per cell [domain x cluster], random selection operating on the cell level was sometimes necessary. While most neighborhoods did not have an overabundance of cases in any of their domains, some did. For example, one neighborhood had nearly 100 realtors in their business domain; another contained almost twenty churches. By randomly selecting nine initial cases in cells such as these, we were able to uphold the design goals of the sample in as unbiased a way as possible.

As Graph #1 shows, after full implementation of our de-duplication and random selection activities, we could ultimately release only 2,536 Positional cases. This was less than half of the number of cases we had identified as meeting the survey's geographical design and represented



roughly a 4:1 ratio from list sample to released cases. This dramatic and surprising dropoff accounted for the much higher than initially-expected time and costs required for sample preparation.

Attrition through Post-Release Ineligibility. The fourth bar in Graph #1 illustrates the additional sample attrition which occurred, even after cases were released, due to the design constraints, with 605 cases eventually characterized as ineligible. The largest number of cases, 248, became ineligible because the organizations had either closed or moved to an unsampled area in Chicago since the original lists had been compiled. Another 116 ineligibles represented sample members who strongly identified themselves as completely unfamiliar with the neighborhood, most often due to recent acquisition of their position. Positions that were not currently filled or had become defunct were also identified as ineligible cases.

One hundred fourteen cases turned out to be duplicates, having slipped through our careful de-duplication process. Thirty seven cases could not be completed due to a language barrier. An additional 219 cases were characterized as ineligible because the cell was determined to be full. As a consequence of the reduction in scope to 2,800, certain Neighborhood Clusters and/or Informant Categories became *saturated*, containing enough completed cases to meet or, in many cases, exceed the reduced design goals. It was decided to instead expend remaining resources toward obtaining interviews in those cells which were farthest from meeting the target number of completes.

Domain Distribution of Positionals. Graph #2, Positionals by Domain, summarizes the successes and challenges of the Positional sample. The pattern that developed, where only a small percentage of the listed sample was actually released, held true across all domains with the exception of the law and political domains. As a result of the distinctive jurisdictional nature of law and political institutions, as well as our source for Positional names for the police officers, we released a much higher proportion of the listed sample in these domains (Table 3).

Table 3. Listed and Released Positional Cases, by Domain

Domain	#Listed	# Released	% Released
Education	1987	565	28%
Religion	3367	626	19%
Business	2646	433	16%
Political	327	165	50%
Law Enf.	328	328	100%
Com. Org.	1329	415	31%

In Graph #2, the first bar within each domain displays the cases released for interviewing. The second bar shows the number of cases that subsequently became ineligible. Domains differed distinctively. The education and community organization domains were primarily comprised of stable, well-established institutions, rooted in the community. The religious domain contained many "transient" churches that opened, closed, moved, merged and changed names frequently. Similarly, many realtors in the business domain had relocated or gone out of business, while the business organizations and editors remained quite stably rooted in the community.

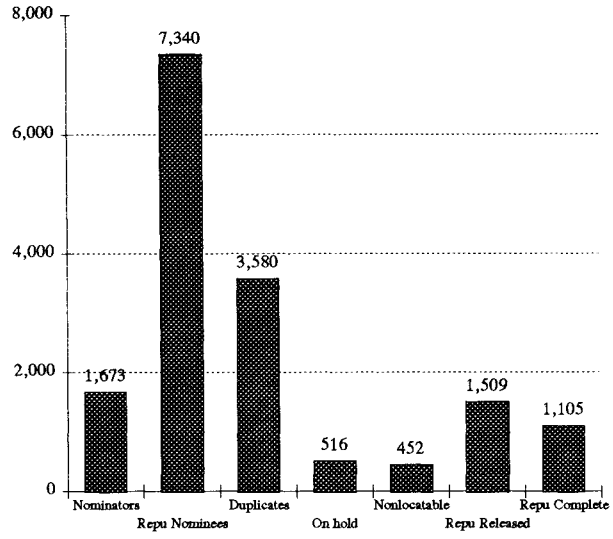
The third bar in each domain represents completed interviews. The horizontal line across the top of the graph represents the revised target of 440 completes per domain for *both* Positional and Reputational-type respondents. Positional interviews not only comprised more than half of the completes in five of the six domains but actually exceeded 75% of the total completes in the Education, Religion, and Community Organization domains. The graph also indicates why the goal of even distribution for the Political domain had to be adjusted to more accurately reflect the institutional realities of that domain.

REPUTATIONALS

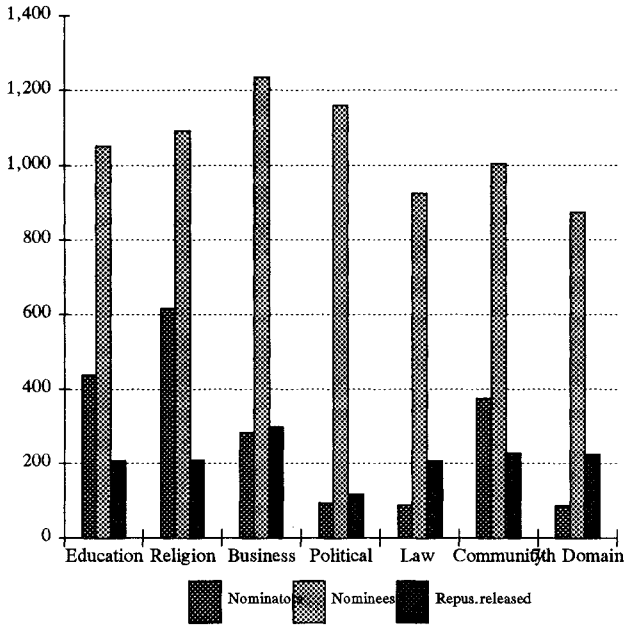
Reputational Sample Overview. The Reputational sample was generated from responses to seven questions asked during the interview. The questions asked respondents to nominate individuals in the community who were knowledgeable about either neighborhood issues related to a particular domain or about the neighborhood in general. (Questions available upon request).

Graph #3 represents the process of developing the Reputational sample. The initial wave of the Reputational sample was based on data collected from the first 817 completed Positional interviews. The data generated 3,140 Reputational nominations, for an average of 3.8 nominations per Positional respondent. From those 3,140

#3 Reputationals



#4 Reputationals by Domain



nominations, the number of useable Reputational cases dropped to 874. In other words, 817 Positional nominators yielded 874 selected, non-duplicated, locatable Reputational cases, approximately a 1 to 1 ratio. After those 874 Reputationals were released to work, we also used their responses to the seven questions to generate additional Reputational nominations, although the vast majority of reputational nominations were generated by the initially-interviewed Positional Respondents grounded in the Neighborhood Clusters. All told, we analyzed 7,340 nominations from 1,673 nominators (both Positional and Reputational) and released 1,485 cases to work.

The dramatic drop from the number of Reputational nominations to the number of useable Reputational cases can be explained largely by the amount of duplicative nominations received. Of the 7,340 nominations, 3,580 (49%) were found to be duplicate names. This graph includes duplicates defined as either the same individual nominated more than once, or a nominee who was already released from our sample. Both types of duplicates, in a sense, validated our sample design. For example, when an individual was recommended by several different Positional respondents, it was more likely that the nominee was indeed an important member of the community. Similarly, when a nominee repeatedly turned out to be in our Positional sample, it supported both the accuracy of our Positional lists and the validity of their inclusion as key community informants. The duplication rate rose as more nominations were generated indicating we had reached institutional saturation in many of the communities with our sample.

Other nominations that had to be eliminated were those that were not locatable. There were 452 un-locatable cases, representing a 94% locating rate for non-duplicated nominations. This graph represents an enormous investment in time and labor, as the nomination data was often incomplete or misleading, and involved the use of multiple locating tactics to obtain sufficient locating information.

An additional 516 nominations were not useable Reputational cases because they had been placed on hold by the client. Either these cases fell into saturated geographical areas or domains, from which we did not want to release any more cases, or they lacked sufficient information for us to be certain that they were indeed key community members for the Neighborhood Clusters or domains of interest.

Domain Distribution of Reputationals. Graph #4 illustrates how the nominators, nominees, and useable Reputational cases are distributed by domain. The solid bar represents the number of nominators in each domain. The striped bar represents the total number of individuals nominated in each domain. The dotted bar represents the number of Reputational cases released to be interviewed in each domain.

It is important to note that nominators were asked to recommend individuals not only within their own domain but for each of the other domains as well. This is significant because it was thought that nominators may recommend individuals within their own domains only. In other words, there was a concern that the unequal distribution of respondents per domain may result in an unequal distribution of Reputational nominees, and ultimately cases, per domain. As illustrated in Graph #4, this was clearly not a problem. For example, the law domain included only 88 nominators, yet yielded 206 Reputational cases.

In addition, the data provide evidence that nominators recommended individuals outside of the original informant categories that generated the Positional sample. This was

particularly evident in the business domain, where nominations captured a far broader range of businesses than the original lists included. As discussed in the Positionals section and illustrated in Table 2, the only Positionals in the business domain were business associations, banks, realtors, and editors. In contrast, the Reputational nominees in the business domain included such categories as grocery store owner, florist, restaurant manager, and hardware store owner. This broader scope of business informants was desirable, as it provides a clearer picture of the business domain within each community.

Geographic Distribution by Respondent Type. Graph #5 depicts the total number of completed interviews with both Positional and Reputational respondents in each of the eighty Neighborhood Clusters. The solid bar represents the number of completed Positional interviews; the dotted bar the number of completed Reputational interviews. The horizontal line at the top of the graph represents the revised targeted number of interviews (35) in each Neighborhood Cluster. This graph illustrates that the targetednumber of completes was met or exceeded in more than half of the neighborhood clusters and that each cluster generated data collected from a substantial number of interviews with both Reputational and Positional informants, meeting a primary survey objective.

FINAL RECOMMENDATIONS

Recommendations for subsequent community-based, elite surveys can be derived from our experiences in constructing the sample for the Neighborhood Experts Survey. In conducting a community-based, elite survey equal reliance upon both a Positional list sample and a Reputational snowball sample appears to be the best approach for achieving a balance between a replicable methodology and a reliable reflection of institutional and geographical reality. However, one must be prepared for a high level of pre-field or pre-release effort concentrated upon case preparation, and post-release effort concentrated upon gaining cooperation with, as well as tracking, the sample. Dependent upon the number and kinds of constraints imposed by one's survey design, this very well might involve 85% potential case attrition rates (from original lists) and tracking literally hundreds of cells. For example, the geographical and domain constraints of the Neighborhood Expert Survey translated into our frequent review of 560 different cells (each cell displaying released and completed cases for a single domain in a specific Neighborhood Cluster).

Similarly, these constraints meant that while preparation time was spent on more than 17,000 potential cases (from both the original lists of Positionals and the nominations of Reputationals), only 2,800, or 16%, of those became completed interviews, even with the attainment of an 85% response rate. This high attrition rate, of course, was not only due to the demands of the survey design but also reflects the overlapping nature of elite social networks