REACTIONS TO CRIME PROJECT

CENTER FOR URBAN AFFAIRS

Wesley G. Skogan[a]

Northwestern University

THE REACTIONS TO CRIME PAPERS

CENTER FOR URBAN AFFAIRS
NORTHWESTERN UNIVERSITY

WORKING PAPER: M - 31F (Revised)

THE CENTER FOR URBAN AFFAIRS RANDOM DIGIT DIALING TELEPHONE SURVEY

REACTIONS TO CRIME PROJECT

CENTER FOR URBAN AFFAIRS

Wesley G. Skogan[a]

Northwestern University

September 1978

(Revised)

NORTHWESTERN UNIVERSITY
2040 Sheridan Road
Evanston IL 60201

[a]with the assistance of Gary Jason

*This is an internal working paper of the Reactions to Crime Project and is not for quotation, citation, or general distribution without the consent of the author.

Prepared under Grant Number 77-NI-99-0018 from the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice. Points of view or opinions in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice.

CONTENTS

Page

A. THE PURPOSE OF THE SURVEY 1

B. SURVEY METHODOLOGY: GENERAL CONCERNS 2

C. RANDOM DIGIT DIALING PROCEDURES 5

D. SAMPLING FOR SEX DISTRIBUTIONS. 14

E. INTERVIEW PROCESS AND COMPLETION RATES. 17

F. INDICATORS OF SAMPLE AND DATA QUALITY 27

G. CONCLUSIONS 33

FOOTNOTES 34

CITATIONS 36

APPENDIX A: CODING OPEN-ENDED ITEMS (By Gary Jason). . . 37

APPENDIX B: NOTES ON ANALYSIS FILES. 40

APPENDIX C: THE SURVEY QUESTIONNAIRE 42

THE CENTER FOR URBAN AFFAIRS RANDOM DIGIT DIALING TELEPHONE SURVEY

A. THE PURPOSE OF THE SURVEY

This survey was conducted by the Center for Urban Affairs at Northwestern University, to gather information for two investigations of the impact of crime on the lives of city dwellers. Both research projects are concerned particularly about how individuals attempt to reduce their chances of victimization by changing their behavior, and how neighbors organized to fight crime and reduce the fear of crime. The Reactions to Crime Project ("RTC Project") is interested in the impact of crime and neighborhood conditions on these concerns, while the Rape Project is concerned specifically with sexual assault and its consequences for the lives of women. Both investigations are funded by the federal government, and the results of the survey will be included in reports to the relevant agencies about these problems. The Reactions to Crime Project is supported by the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, while the Rape Project is a program of the National Center for the Prevention and Control of Rape, a sub-division of the National Institute of Mental Health.

Northwestern's crime projects are multi-year efforts aimed at understanding how residents of urban communities cope with crime and consequences. The design and content of this survey reflected that concern. A major component of the RTC Project's effort is a study of collective responses to crime--how individuals band together to deal with crime problems. Both projects were interested in individual responses to crime (e.g., property marking, the installation of locks and bars) and the impact of fear of crime on day-to-day behavior (e.g., shopping, recreational patterns). This led to the inclusion of a number of questions in the survey calling for self-reports

of behavior. We wanted to know how people get their ideas about crime, so we asked who they talk to and what they watch on television and read in the newspapers. Because we were interested in the neighborhood as a locus of action, we asked a number of questions about events and conditions in our respondent's home areas. There were several questions about their relationship to their neighbors, and who they know and visit around their homes. The survey questionnaire included a number of questions measuring our respondent's perceptions of the extent of crime in their communities, whether they knew someone who had been a victim, and what they had done to reduce their own chances of being victimized. Finally, there were a number of specific questions about sexual assault, some of which were asked only of women.

The information collected in the survey is complemented by the notes of field observers who were stationed in the same areas in the year preceding the survey. They talked to community residents and leaders, and canvassed local organizations about anti-crime activities in their assigned neighborhoods. We also have been collecting and content-analyzing city and community newspapers which reach residents of these neighborhoods and cities. Together, these data should give us a broad picture of the impact of crime in these communities.

B. SURVEY METHODOLOGY-GENERAL CONCERNS

The sampling frame and sampling procedures employed in this survey were shaped by cost considerations and the substantive focus of the survey. While the projects share a lively interest in criminal victimization and the demographic correlates of individual victimization, these were not foci

of this survey. This was dictated in part by the relatively infrequent incidence of serious personal victimization, the only form of criminal predation which appeared--at the time we designed the survey--to have any substantial attitudinal or behavioral impact (Skogan, 1977). The victimization surveys conducted by the Census Bureau indicate that perhaps three percent of the residents 16 years of age and older of large central cities fall victim to robbery during the course of a year, and methodological research indicates that attempts to gather data over a longer recall period are fraught with difficulty. Thus, only survey samples of the magnitude employed by the Census Bureau (over 21,000 respondents per city) can gather reliable data on such events.

However, all evidence indicated that most attitudinal and behavioral responses to crime were much more normally distributed in the population. In the five large cities surveyed by the Bureau early in 1974, 52 percent of their espondents indicated that they felt "very safe" or "reasonably safe" while alone on the streets in their neighborhoods at night, while 48 percent did not. Almost the same proportion reported that they had changed their behavior "because of crime." Sample surveys are most efficiently employed to gather data on conditions of high prevalence or events of frequent incidence, and the fear of crime and actions taken to reduce the risk of victimization appeared to meet those criteria. The only exception to this expectation lay in the area of collective responses to crime. Previous research in Chicago (O'Neil, 1977) indicated that participation in anti-crime organizations is relatively infrequent.

From the beginning the RTC Project has emphasized the neighborhood basis of individual and (especially) collective action. Thus, we needed to field a survey study of individual perceptions and actions which placed respondents

within a known neighborhood nexis. Within each of the three cities under investigation--San Francisco, Philadelphia, and Chicago--the Project gathered extensive data on three or four neighborhoods. The sampling frame for the survey thus had to produce respondents who lived within the boundaries of those areas. Those boundaries were determined by the perceptions of area residents interviewed during the fieldwork phase of the project, and were not drawn to match any convenient, pre-existing geographical sub-units. Further, because we wished to use the survey data to characterize those neighborhoods, we had to gather data on large samples of respondents in each area. Finally, the neighborhoods themselves were chosen on the basis of their characteristic class and racial status, their crime rate, and upon the apparent level of organizational activity there: they are in no way representative of the cities in which they were located, or of urban neighborhoods generally. Therefore, we also fielded a modest city-wide survey of residents of each of the three communities. Those data can be utilized to place our target neighborhoods within the broader context of each city.

The Rape Project component of the enterprise also imposed an important substantive demand upon the survey: a focus upon women. While the Rape Project required comparative attitudinal data for males, many of their interests are female-specific. They are interested in the way in which women alter their life-styles to reduce their chances of victimization from rape, their perceptions of their risks under certain circumstances, and the impact of rape upon their relationships with others. Further, the Rape Project planned to conduct intensive in-person follow-up interviews with selected respondents, and the telephone survey concluded by identifying those respondents and securing their cooperation for participation in a second interview. Because of the sample sizes involved in the telephone survey, it thus was necessary

to over-sample women in order to produce enough female respondents to meet the goals of that project.

The substantive demands of the RTC and Rape Projects thus created several important methodological and procedural constraints upon the design of the survey. These included the sample sizes required, their concentration in numerous and small geographical areas, the multi-city focus of the projects, the large female contingent to be interviewed, and our interest in infrequent events, including the sensitive issue of sexual assault. Further, several of our neighborhoods housed large Spanish-speaking populations, some of whom are reputed to be undocumented aliens, and others were relatively disorganized places characterized by high residential mobility. The high crime rate in several of them also affected decisions about interviewing, for interviewer safety and interview quality both are reduced by untoward environmental conditions. Finally, our budget was (like all budgets) limited, and we could only do what we could afford.

C. RANDOM-DIGIT DIALING PROCEDURES

One of the most important decisions to be made about the survey was the mode of data collection. In practice this reduces to a choice between personal interviews and interviews gathered over the telephone (Garofalo, 1977). While there may be some dispute over the relative validity of data gathered through telephone interviews, there is firm evidence that such information is as reliable as that collected in person, and that the two methods yield data with the same marginal distributions and interrelationships between variables when used in the same sampling universe (Tuchfarber and Klecka, 1976; Groves, 1977). Data on the incidence of telephone usership

(Powell and Klecka, 1976) and the telephone and personal-interview refusal rates in big cities (Groves, 1977) indicate that telephone-based random-digit dialing sampling frames and interviewing procedures do not produce substantial unique biases if we accept in-person interviews with persons selected in more traditional ways as the criterion.

Klecka, et al. (1976) suggested that surveys conducted over the phone should cost only 30% as much as in-person interviews. More recent cost estimates have suggested somewhat less of an advantage for telephone interviews, however. Telephone interviews necessarily are substantially shorter in duration than personal interviews, thus reducing the amount of data which can be collected in them. Groves' (1977-revised) experience indicates that data collected through telephone surveys may cost about one-half as much as those collected in person.

Adopting the telephone as the interview mode solved some of the problems facing us, but exacerbated others and created several new ones. The telephone mode of interviewing lends a great deal of control over interviewer behavior and interview quality, for supervisors can conveniently monitor conversations directly and re-interviews can be conducted cheaply. Also, interviewer safety is enhanced, and it probably is more likely that interviews in unsafe neighborhoods and homes will be completed (Tuchfarber and Klecka, 1976). The reduced cost of telephone interviews also gave us some hope of conducting enough interviews within our budgetary constraints to characterize multiple cities and numerous neighborhoods.

The major difficulty with the procedure was that telephone samples present many more imponderables than their in-person counterparts. In this survey we chose to employ Random Digit Dialing (RDD) techniques for selecting our respondents. We produced thousands or telephone numbers randomly, using the

computer to select three-digit prefixes serving our target areas and to generate seven-digit numbers. As discussed in detail below, this procedure does not lend itself to any certainty about what is going to happen once a survey begins. Unlike area-probability samples of physical locations, we could not know with any precision where a telephone responding to a give number would be located. We could not know whether a number was residential, commercial, or connected to a telephone booth, or to some government agency or other institution. We could not even know if it was a working number, connected to anything at all. We could learn the latter by calling each number and discovering if it was a "ringing number": however, we never could learn much about numbers which rang whenever called, but which never were answered.

Although telephone interviews thus are cheaper to conduct than face-to-face interviews, locating suitable respondents (in this case, randomly-selected adults stratified by sex and living in housing units located within the boundaries of our neighborhoods) is more expensive and complex. And, unlike personal-interview studies, telephone interviewing yields little data about nonrespondents, those who never are at home to be interviewed or refuse to cooperate with the interviewer.

This survey was carried out by the Market Opinion Research Corporation between October and December, 1977. Questionnaire preparation and initial pretesting, along with all sampling and telephone number preparation, was conducted at Northwestern. The city-wide component of the survey was designed to reach randomly-selected adults in 540 households in each city. Because a well executed random-digit dialing survey involves no clustering of sample units, the sampling variation from such surveys should approach those attributable to random chance. This sample size thus should reduce sampling

error to the 4 1/2 percent range, which we felt would enable us to speak confidently about important inter-city differences in our data. In addition, interviews were to be conducted with residents in ten selected neighborhoods, four in Chicago and three in each of the other cities. The neighborhood samples were to range in size from 200 (in four of the sites) to 450 (in six areas). The larger neighborhood samples were those in which female respondents were to be oversampled. By increasing sample sizes there we still were able to maintain an effective (weighted) sample size of about 200 respondents in each area, balanced across the sexes. In total, 1640 interviews were to be conducted in Philadelphia and San Francisco, and 1840 in Chicago.

The telephone numbers to be called were generated by a computer program. Inspection of telephone company exchange-area maps and reverse ("criss-cross") directories lising telephones by address produced a list of all three-digit prefixes operative in each target neighborhood. Lists of all prefixes operative in each city were available from their telephone companies. Some prefixes which exclusively were alloted to large institutions or reserved for commercial or telephone company use were deleted from those lists, for only residential numbers were "in scope" for this survey. Prefixes were also purged from this list if they were less than 20 percent full of listed numbers, for calling randomly in largely empty exchanges would be extremely unproductive.[1] For the city samples, this proportion was reduced to ten percent. Because telephone numbers are randomly spread by prefix within the central office area they serve (see footnote 2), we judged that this procedure did not seriously bias our data on neighborhoods as none of their sub-areas were thus excluded. However, when exchanges are only slightly filled because they have only recently been opened for new assignment, this procedure may bias the sample slightly to the disadvantage of recent movers.

Next, estimates were made of the number of telephone numbers which should be generated for each area using these prefixes. These estimates had to take into account the number of interviews we wanted to complete, our expected refusal and break-off rates, and the number of out-of-scope or non-working numbers that would remain in our telephone sample despite our best efforts to purge it of unwanted numbers. Our estimates were based upon the experience of the Behavioral Sciences Laboratory of the University of Cincinnati (Tuchfarber and Klecka, 1976) and the Survey Research Center of the University of Michigan (Groves, 1977) both of which have produced detailed reports on conducting RDD surveys. These estimates also were affected by the number of prefixes and exchange areas serving a neighborhood and the degree of correspondence between a neighborhood and the telephone company central office areas serving it. In general, the larger a target area within a central office boundary, the larger the proportion of numbers we would generate which would fall within our desired neighborhood.[2] The number of prefixes serving each of our cities and neighborhoods (less the exclusions recounted above), and the number of telephone numbers we created for each area indicated in Table One. For example, in areas in which we desired to reach 450 respondents, we usually generated 15,000 numbers. With the elimination of duplicate numbers, this

- - - - - - - - - - - - -

Table 1 goes about here

- - - - - - - - - - - - -

initial set was reduced to about 13,500. Each number was thus a unique seven-digit value created first by randomly selecting an in-scope prefix and then attaching to it a four-digit random number.

These numbers were generated by a specially-written program, BELLTEL. As it created each number, BELLTEL kept track of the order in which it was

TABLE 1

TELEPHONE SAMPLE PREPARATION

Sample	Desired N	Number of Prefixes[a]	Numbers Generated (Excluding Duplicates)	Editing-Percent Excluded	Remaining-Sent to MOR
San Francisco City	540	61	7936	9.0	7221
Visitacion Valley	450	2	10698	40.3	6386
Sunset	450	7	13442	43.8	7558
The Mission	200	10	7649	31.1	5272
Philadelphia City	540	112	7972	10.1	7154
West Phily	450	9	13777	36.0	8814
South Phily	450	9	13786	37.5	8617
Logan	200	4	9628	33.3	6425
Chicago City	540	172	6981	4.6	6675
Lincoln Park	450	12	18423	64.2[b]	6593
Wicker Park	450	9	13807	58.9[b]	5673
Woodlawn	200	9	7694	28.9	5469
Back of the Yards	200	13	7759	35.8	4984
Totals	5120	429	139552	37.8	86841

[a]Excludes prefixes estimated less than twenty percent full.

[b]Illinois Bell's name and address service was employed to screen a large proportion of the sample numbers in these areas.

born. This defined the random sequence in which they later were to be called. Then, the program sorted the telephone numbers in ascending order, to match the format of criss-cross directories, and printed them out for visual inspection by our staff.

This list of numbers was then edited by a laborous, and expensive, process designed to decrease the proportion of the final set which were commercial or institutional, not residential numbers, and numbers assigned to residences located outside of the target neighborhoods or cities.

The first stage of the cleaning process involved checking each number against a criss-cross directory for each city. Those directories include all "published" telephone numbers in a city arranged in ascending order by prefix. They do not include unpublished numbers or those assigned to coin telephones or reserved for internal telephone company use.[3]

Each computer-generated number was inspected, and its status determined. A number could be listed as assigned to a business or institution (most of whom have their numbers published), and those were deleted. Likewise, residential numbers located in the wrong area were excluded. Residential numbers located within a target area were saved. Finally, many numbers simply were not printed in the directories. These were either non-working (they did not exist), or unpublished numbers given to private subscribers, coin booths, or telephone company phones. Some also could have been assigned to any of those users since the publication of the criss-cross directory. These numbers were all retained, for unpublished residential telephones now make up 25-35 percent of the total in major cities. To exclude all numbers we could not find in the criss-cross directories would have left out this important population from our sample (Rich, 1977). In the city of Chicago about 33 percent of all residential telephone numbers currently are

unlisted. An additional 8 percent are not printed in any directories but can be accessed through directory assistance (Chicago Daily News, 3 October 1977).

The primary determinants of the proportion of numbers that could be deleted using criss-cross directories appeared to be (1) the extent to which prefixes serving an area were being utilized fully and (2) the incidence of unpublished numbers. Thus, the effects of this screening varied from area to area. In most cases it reduced the initial list of numbers for neighborhoods only about 30-40 percent. In others, with the aid of additional procedures as many as 65 percent could be eliminated. The remainder were listed in-scope residences, unpublished residential and commercial/institutional/telephone company numbers, and coin telephones, along with a substantial component of numbers which were not printed because they were not working numbers.

There was, of course, some error even in this process. Most important, the criss-cross directories available for this project were approximately nine months out-of-date. Thus, some numbers we retained as residential in-scope would be non-working at the time of the survey, for some of those families would have moved recently. Or, numbers which we deleted as out-of-scope could have been re-assigned to in-scope residences. On the other hand, some numbers which we retained because they could not be located in the criss-cross directories would have been assigned, some to businesses (bad), some to out-of-scope residences (bad), and some to in-scope residences (good). Errors in number-checking, like the proportion of numbers likely to be in-scope, vary by neighborhood, as communities vary in their rate of residential mobility and commercial expansion or contraction.

We found that approximately 290 numbers could be screened per hour through inspection in a criss-cross directory. The directories themselves were leased from private companies, Haines Directory Service and Coles Directory Service. Rental of the three city directories cost $500. In addition we spent a total of $1275 in direct labor costs for this phase of the sampling operations.

In the city of Chicago we were able to further reduce the size of our pool of random telephone numbers and update some of the information available from the criss-cross directory. In that city (but not in others), a "name and address service" will give information about specific numbers, including whether they are working numbers, published or unpublished, or if they are pay phones or internal telephone company numbers. If numbers are published, the service also supplies the name and address under which they are listed. In Chicago we were able to use this service to check approximately 70 percent of our criss-crossed numbers in one of our 450-respondent neighborhoods (Wicker Park), and 50 percent in the other (Lincoln Park). This resulted in a further reduction of the Chicago neighborhood sample by about 25 percent in Wicker Park and 30 percent in Lincoln Park. This cost us $345.

In all of the cities we were able to do more number-deletion based upon information available from the telephone companies or apparent upon inspection of the numbers and directories. For example, in Chicago all numbers in the "9900" range are reserved for telephone company use, as are all numbers beginning with "00" in San Francisco. They were deleted. Businesses may hold any number, but in some prefixes they tend to be clustered in the 8000 and 9000 ranges, and inspection through the criss-cross directories isolated banks of numbers within a prefix that clearly were reserved for

commercial use. In some prefixes, 9000-series numbers not listed in the directories proved to be coin phones. In Philadelphia, we were able to secure a list of all telephone numbers assigned to "semi-public" coin telephones (those located within and assigned to private establishments such as bars or restaurants), and in San Francisco, we acquired a list of all coin telephones served by prefixes operative in our target neighborhoods. All of these were deleted. Finally, we carefully inspected the city samples and the telephone numbers for each area, searching for large sequential banks of numbers which were not traceable. If a range of 100 numbers or more was found in which no listings were available, it was checked to validate that it was a working bank of numbers. In all of the cities we called telephone company Service Representatives responsible for suspicious prefixes, explained what we were about, and asked if there were any residential subscriptions active within that bank. In most cases we were able to secure this information, although Service Representatives for Bell Telephone in Chicago were less cooperative than those in other cities. This enabled us to delete blocks of non-residential or non-working numbers. This procedure is useful because telephone companies open new numbers for assignment in banks of 1000, as demand requires. It is also inexpensive, for researchers may call telephone company employees anywhere in the country "collect" in order to inquire about their service.

After each number was checked against the criss-cross directories, screened through coin-phone lists, checked for commercial sequences and dead banks, and (for some numbers in Chicago) checked through the name and address service, all out-of-scope numbers were deleted from their area files using a text-editing program. Then, the remainder were re-sorted using the original sequence number, returning them to their random order. These numbers were

then printed on pressure-sensitive labels (along with a city and neighborhood identifier and a new continuous sequence number), and sent to MOR.

Altogether, we utilized $2,666 worth of computer time and file storage charges on Northwestern's CDC 6400 processing these numbers.

The original, random order defined the calling sequence for the numbers in each sample. This calling sequence is illustrated in Figure 1. Each number for an area or city was called in turn. For numbers which appear to

- - - - - - - - - - - - - - -

Figure 1 goes about here

- - - - - - - - - - - - - - -

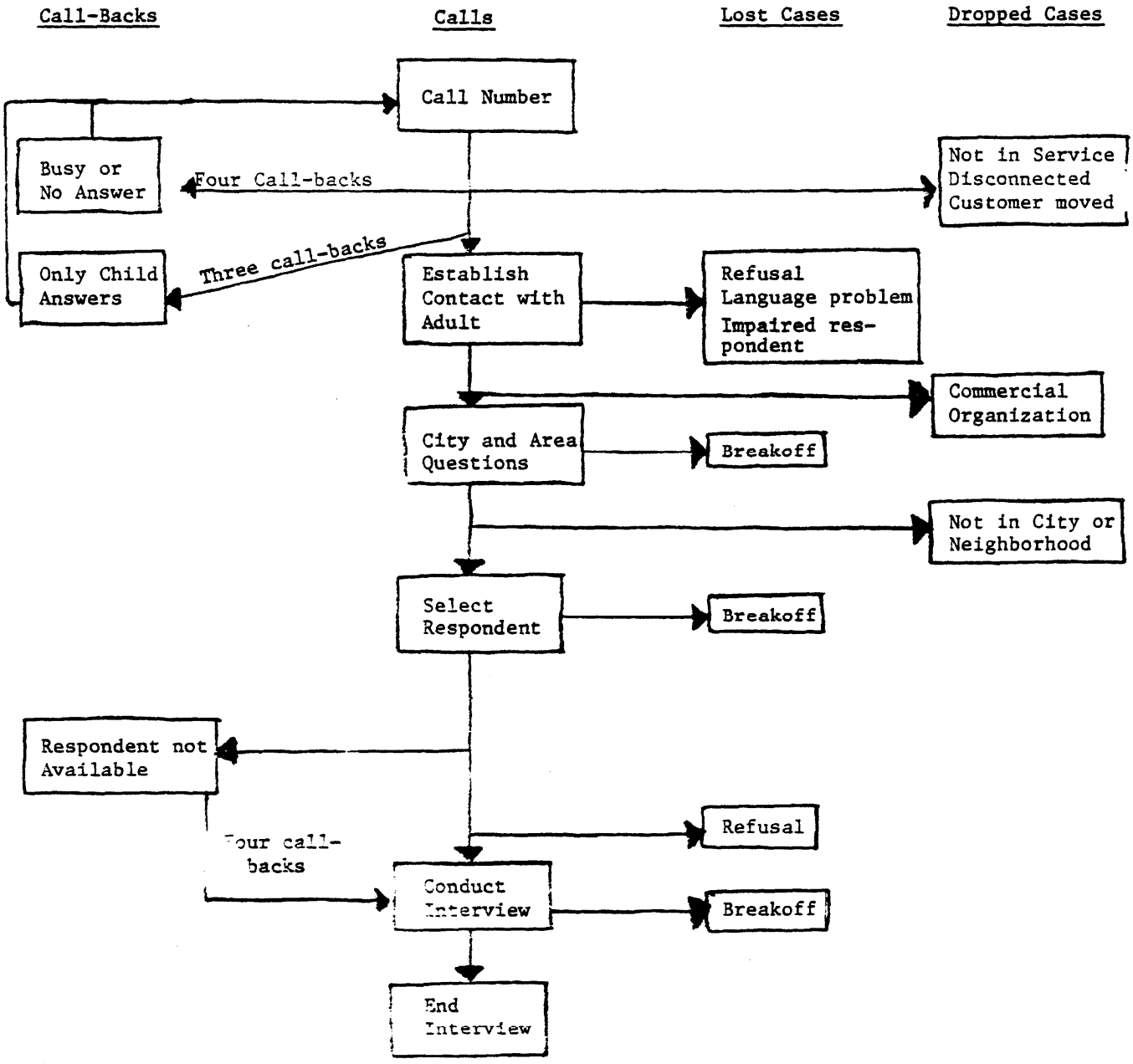
be operating, a total of five calls were made, spread over days and shifts, to reach a responsible adult.[4] An early screen question took out commercial or institutional phones which slipped through our number-checking process. Another checked each household in a neighborhood sample to make sure it lay within the specified area boundaries.[5] A total of 3 call-backs were made to find an adult at home to serve as a household informant. This informant was quizzed to establish the composition of the household, and a respondent (18 or older) then was randomly selected using a Trodahl-Carter-Bryant selection matrix. As many as four call-backs could be made to arrange an interview with this respondent. Thus, no number was substituted for another; rather, interviewers worked numbers in batches of 1,000, making the requisite call-backs or eliminating numbers as out-of-scope roughly in sequence until the respondent quota (specified in Table 1) was reached in each city and neighborhood.

D. SAMPLING FOR SEX DISTRIBUTIONS

Because of the substantive interests of the Rape Project, female respondents were to be oversampled in several of the neighborhood surveys.

FIGURE I

RDD SURVEY CALL SEQUENCE



Oversampling of females was accomplished by manipulating the use of the Trodahl-Carter-Bryant respondent selection matrix so that they were more likely to be randomly selected. Figure 2 presents an example of a respondent selection matrix which oversamples females.

- - - - - - - - - - - - - - - - -
Figure 2 goes about here
- - - - - - - - - - - - - - - - -

The T-C-B respondent selection procedure involves the use of several different versions of a grid for selecting respondents. The grid is formed by the number of adults and the number of males in a household. Those figures identify a unique household respondent (see Figure 2 below). The sex proportions of the resulting sample can be manipulated by the mixture of male and female respondents identified in a grid, and by the random rotation of selection matrices favoring various classes of respondents.

In the analyses of the data conducted by the RTC Project, female respondents are under-counted to reflect their true proportion in the population. While this may present some difficulty in interpreting tests of significance calculated from the data, it will not affect the reliability of the findings. In our analysis of the data we assume that the effect of down-weighting is to make tests of significance more <u>conservative</u> (there are more sample cases than assumed in the calculations), and thus we often continue to employ them. Table 2 (below) reports the final distribution by sex of respondents in each of the city and neighborhood samples. In order to adjust these samples, the 1970 Census estimate of the proportion of females in the resident population of the cities (about 53 percent of each) was used as the criterion. In addition to the areas in whcih we deliberately over-sampled females, several samples (notably Chicago and Philadelphia City-wide, Back of the Yards, and Woodlawn) included somewhat too many women. We therefore re-weighted every sample using the appropriate city-wide criterion, for sex is the strongest

FIGURE 2

RESPONDENT SELECTION GRID

Row B	Col. A Number of Adults in Household			
Number of Men in Household	1	2	3	4
0	Woman	Oldest Woman	Oldest Woman	Youngest Woman
1	Man	Woman	Youngest Woman	Man
2		Youngest Man	Woman	Woman/ Youngest Woman
3			Oldest Man	Youngest Woman
4 or more				Youngest Man

Version 4

NOTE: The intersection of Col. A and Row B determines the sex and relative age of the respondent to be interviewed.

individual-level predictor of both victimization and fear, and weighting appeared to be a necessary step if we were to make meaningful estimates of the level and salience of each at the city and neighborhood level.

Operationally, this was accomplished in the following manner: a weighting variable called SEXWT was created which had a value 1.0 for all males, while females in each sample were given weights calculated using the following formula:

$$\text{SEXWT} = \frac{\text{\# of females in city census}}{\text{\# of males in city census}} \times \frac{\text{\# of males in sample}}{\text{\# of females in sample}}$$

- - - - - - - - - - - - - - -

Table 2 goes about here

- - - - - - - - - - - - - - -

In addition to its primary data-gathering function, the telephone survey also was a vehicle for securing the cooperation of selected individuals for further, intensive follow-up interviews, to be conducted in-person. Those interviews focused upon sexual assault and self-protective measures taken by women. In selected areas, female respondents were to be asked--at the conclusion of the regular interview--if they would be willing to cooperate in such a study. A modest financial incentive for doing so was offered. This is illustrative of one important use of telephone surveys, as a pre-screening device. Our experience indicates that such a sampling strategy might be of some utility when sensitive topics requiring some rapport and trust are involved Table 3 indicates the proportion of women indicating that they would be willing

- - - - - - - - - - - - -

Table 3 goes about here

- - - - - - - - - - - - - -

to be interviewed in person by area.

Table 2

Sample	Telephone Numbers Sent to MOR	Telephone Numbers Used by MOR	Completed Interviews	Percent in Spanish	Percent Female
San Francisco City	7221	2721	539	7.1	52.3
Visitacion Valley	6386	4401	448	6.5	67.4
Sunset	7558	3372	453	5.1	62.9
The Mission	2572	1722	201	13.9	46.3
Philadelphia City	7154	2249	540	1.7	58.1
West Phily	8814	2689	450	1.1	72.7
South Phily	8617	2163	449	4.0	68.6
Logan	6425	1271	201	4.0	51.7
Chicago City	6675	1785	539	6.5	59.0
Lincoln	6593	2933	450	11.1	58.9
Wicker Park	5673	4014	451	6.9	64.1
Woodlawn	5469	1403	200	1.0	68.0
Back of the Yards	4984	1396	200	14.0	61.0
Totals	86841	32119	5121	5.9	61.4

Table 3

RESPONSES TO SCREEN QUESTION ASKING FEMALE RESPONDENTS TO PARTICIPATE IN RAPE PROJECT FOLLOW-UP IN-PERSON INTERVIEWS ABOUT SEXUAL ASSAULT

Area Sample	Percent saying: YES	NO	(N)*
West Philadelphia	39	61	(306)
South Philadelphia	27	73	(289)
Lincoln Park	37	63	(241)
Wicker Park	22	78	(257)
Sunset	26	74	(280)
Visitacion Valley	32	68	(288)
TOTAL	30	70	1661

*Unweighted number of females asked to participate.

E. INTERVIEW PROCESS AND COMPLETION RATES

Table 2 also presents summary information describing the use of the sample telephone numbers, the number of completed interviews, and their distribution by language. In all, almost 87,000 pre-screened sample numbers were forwarded to Market Opinion Research. Of those, 32,000 (37%) were used in various ways, following the call sequence described in Figure 1. As this indicates, our rules of thumb for estimating the number of telephone numbers which would be required for each sample led us to produce and process for too many of them. A total of 5121 interviews were completed, spread across the cities and neighborhoods as specified.

Almost six percent of all interviews for the survey were conducted in Spanish rather than English. Each of the city field offices was staffed with at least one Spanish-language interviewer. They generallly "worked" the Spanish-speaking samples in each city, and in addition handled all cases identified by other interviewers as requiring questioning in Spanish. The Spanish-language version of the questionnaire was developed by our field staff, in consultation with OMAR, Incorp., a Chicago marketing firm. That interview form was used most extensively in Chicago (Back of the Yards and Lincoln Park), and in the Mission district in San Francisco.

As outlined in Section C and Figure 1 above, our respondents were reached via computer-generated random telephone numbers. Each number was called in succession from a randomly-ordered list, and was re-called a number of times if necessary. Some could be dropped from the sample immediately, for they proved to be nonworking numbers; others had to be dialed several times before anyone answered, and even then the household member selected for interviewing often had to be called again. Table 4 documents the magnitude of this task. It indicates the number of telephone numbers which had to be called once, twice, or as many twelve times before

- - - - - - - - - - - - - - - -
Table 4 goes about here
- - - - - - - - - - - - - - - -

ultimately they could be "disposed of." About two-thirds of the sample numbers were called only once, while up to five calls led to the ultimate disposition of over 90 percent of the numbers. If every unlikely contingency in the interviewing process illustrated in Figure 1 occurred--if a household were reached only on the fifth call, if it then took three calls to reach a qualified adult informant, and if it finally took four additional calls to complete an interview with the selected respondent--a total of twelve calls could be made to a sample number. As Table 4 indicates, this occurred only once in over 32,000 cases. The data in Table 4 indicate that random digit dialing using computer generated numbers can be a relatively efficient sampling design,for a large number of non-productive sample numbers can be disposed of very early in the process.

Table 5 details the disposition of each of the 56,000 telephone calls made to the 32,000 numbers for this study. As it indicates, the most common result of a call was that it rang, but that no one answered. The next most common outcome was for the interviewer to discover that the computer had

- - - - - - - - - - - - - - - - - -
Table 5 goes about here
- - - - - - - - - - - - - - - - - -

generated a non-working number. About nine percent of all calls resulted in a completed interview, while refusals accounted for twelve percent of them. About nine percent of all calls reached households located outside of city boundaries or outside of the target neighborhoods which we were attempting to sample.

Our use of random digit dialing in conjunction with geographical screening questions to reach households in such selected areas was one of the major features of this survey. The first responsible person reached by each call (the "household informant") was asked a brief series of screening questions

TABLE 4

NUMBER OF CALLS REQUIRED TO DISPOSE
OF A SAMPLE TELEPHONE NUMBER*

Number of Calls	Telephone Numbers Requiring This Number of Calls to Reach Final Disposition		
	Number	Percent	Cumulative
1	21555	67.4	67.4
2	4374	13.7	81.0
3	2207	6.9	87.9
4	1230	3.8	91.8
5	1948	6.1	97.8
6	428	1.3	99.2
7	197	0.6	99.8
8	43	0.1	99.9
9	16	0.05	99.9+
10	4	0.01	99.9+
11	2	0.01	99.9+
12	1	0.00	100.0
Total	32205	100.0	

*Computed from call records supplied by Market Opinion Research

TABLE 5

DISPOSITIONS OF TELEPHONE CALLS*

Call Disposition	Percent of Calls Made
Number not in service	15.6
No answer	38.2
Business number	4.2
Location not in city	0.5
Location not in neighborhood	8.8
Need a Spanish interviewer	0.8
Household respondent not available	5.9
Refusal by household respondent	12.4
Selected respondent not available	2.0
Refusal by selected respondent	1.2
Breakoff during interview	0.2
Other disposition	1.2
Completed interview	9.1
	100.1%
(N)	56093

*Computed from call records supplied by Market Opinion Research

to assure that the number served a residence, and that the household was located in the central city (for the city-wide samples) or in the proper neighborhood. Because these neighborhoods usually were smaller than telephone company central office areas, and often lay astride two or more of them, we knew that a considerable proportion of the households we reached would not be "in scope" for this study. Table 6 details the magnitude of this sampling

- - - - - - - - - - - - - - -

Table 6 goes about here

- - - - - - - - - - - - - - -

problem for each area in the survey.

As Table 6 indicates, sampling cities for respondents using random digit dialing presented few difficulties. In these samples few of those answering fell outside of city boundaries. The bulk of those who were outside the city lived in San Francisco, which is served by one telephone central office area which also includes Daley City to the South. The proportion of city-sample respondents ruled "out-of-scope" for geographical reasons averaged only 3.3 percent in this survey. The ten neighborhood telephone number samples, on the other hand, contained an ample supply of out-of-scope numbers. The least productive number set was that for Lincoln Park in Chicago; there, one-half of all the household informants contacted by telephone said the residence was outside of the boundaries of our study area. The South Philadelphia area, on the other hand, was extremely large, and lay within one telephone exchange area. There only 13 percent of all calls reached households outside our neighborhood lines. On the average, 33 percent of all household informants we contacted reported that they lived beyond the borders of our localities, ten times the fraction for the city-wide samples.

TABLE 6

RESULTS OF SCREENING NUMBERS FOR CITY AND NEIGHBORHOOD RESIDENCE*

Sample	Contacts with Residences[a]	Proportion out of Study Area	Average Number of Calls per Completion
San Francisco City	1472	5.8	8.9
Sunset	2076	26.9	12.6
Visitacion Valley	2176	28.4	17.8
Mission	844	34.6	17.1
Philadelphia City	1310	1.4	8.0
West Philadelphia	1576	27.9	11.7
South Philadelphia	1316	12.9	8.9
Logan	704	21.3	10.7
Chicago City	1073	2.7	6.3
Lincoln Park	1945	50.1	12.5
Wicker Park	2515	45.6	12.3
Woodlawn	747	46.6	9.7
Back of the Yards	848	38.9	11.7
TOTAL	18746	27.5	11.6

[a]Excludes a few interviews terminated for lack of a Spanish-language interviewer.

*Computed from call records supplied by Market Opinion Research.

These proportions have substantial cost implications for those considering random digit dialing surveys of cities and communities. Screening households for locational or other selection criteria is expensive. It is difficult enough to locate adult informants in households, beginning with a set of computer-generated numbers, without adding factors further reducing the productivity of a set of numbers. Our experience indicates that the cost of such screening mounts rapidly when the scope of target areas is reduced, or when they do not match telephone company exchange areas well. In our least productive sample, Visitacion Valley in San Francisco, interviewers averaged only one completed interview for every eighteen dialings. In South Philapdelphia, on the other hand, one dialing in nine resulted in a completed interview, and the Chicago city-wide sample produced one completion for every six calls. Table 6 reports these ratios for each sample in the survey.

A completed interview constituted only one of several possible final dispositions for each sample telephone number, however. The dialings and re-dialings documented in Table 4 also led us to telephones serving commercial establishments or organizations rather than residences, and to households where no adult ever could be found. Table 7 reports the distribution of the ultimate disposition of each sample telephone number. It is from this data that the completion rate for the survey can be estimated.

Table 7 Goes About Here

As Table 7 indicates, the most frequent disposition of a sample number was that it was "not in service." Only 6.5 percent of all numbers, on the other hand, rang on five different occasions without someone answering. Our judgement is that a substantial proportion of these serve pay telephones

TABLE 7

FINAL DISPOSITION OF ALL SAMPLE TELEPHONE NUMBERS*

Final Disposition	Number	Percent of All Sample Numbers
Numbers not in service	8670	27.1
No answers after 5 calls	2091	6.5
Business numbers screened out	2364	7.4
Locations not in City	279	0.9
Locations not in neighborhood	4884	15.3
Needed Spanish interviewers	134	0.4
No household respondents reached	171	0.5
Refusal by household respondents	6867	21.5
Selected respondents never reached	63	0.2
Refused by selected respondents	665	2.1
Breakoffs during interview	88	0.3
Completed interviews	5085	15.9
Other final dispositions	644	2.0
Total	32005	100.1%

*Computed from call records supplied by Market Opinion Research. Excludes a very small number of faulty, mispunched, or blank records.

and other non-residential locations, for we were not calling during a peak vacation period. About seven percent of the computer-generated numbers reached businesses or organizations, and over sixteen percent yielded residences which lay outside our study-area boundaries. All of these numbers, which constituted fifty-seven percent of the total called, were "ineligible" to produce respondents, and are excluded from our computation of completion rates.

The remaining dispositions include some more troublesome figures, however. About 130 households were abandoned by the organization conducting the survey for lack of a Spanish interviewer. The bulk of these were reached by numbers aimed at the Wicker Park neighborhood in Chicago, a community with a substantial number of Spanish-speaking residents. The final sample of respondents in that area was 32 percent Spanish-speaking; following procedures like those below for estimating the proportion of those which would have been in-scope geographically, this figure could have approached 50 percent if those abandoned households had been interviewed. Our conservations with Market Opinion Research on this matter indicate that they had difficulty locating Spanish-language interviewers in Chicago, and that their administrative procedures led them to continue to log in completed English-language interviews in that area until their respondents quota was met.

In an additional 171 cases, 0.5 percent of all numbers, a household apparently was reached, but no suitable responsible informant ever was located. Up to three call-backs were to be used to reach such an individual, but we still must count these numbers as "eligible" for interviewing and debit our completion rate by this (small) total.

The most serious difficulty with the survey is to be found in the number of persons who refused to cooperate in the enterprise. Over 6,800 numbers,

about 22 percent of the total, reached immediately non-cooperating householders. A much smaller number--665--of our randomly-selected respondents refused to be interviewed; as in most surveys, our major problem was "getting in the door" in the first place. Only in 63 cases were we unable to reach a randomly-selected respondent, and once interviews began only rarely were they terminated. Only in 88 cases did a respondent decide to terminate an interview once it had begun, perhaps testimony to the generally interesting issues covered by the questions.

The aggregate impact of these break-offs, refusals, and other interviewing failures are captured in the survey's "completion rate," the proportion of eligible respondents who refused to participate in the study. Table 8 illustrates our procedures for calculating various completion rates for this project. Each is increasingly "less conservative," making more restrictive assumptions about which numbers were eligible to produce respondents.

- - - - - - - - - - - - - - - -

Table 8 goes about here

- - - - - - - - - - - - - - - -

The "gross rate" presented in Table 8 is simply the total number of completed interviews divided by the total number of sample telephone numbers used in the survey. By this count, the completion rate for the survey was about 16 percent. However, it is clear that this is not the appropriate way of calculating such a rate for a random digit dialing survey, for the procedure demands the generation of a great number of non-working telephone numbers and the completion of a number of calls to businesses, hospitals, university centrix systems, and other non-residential establishments. This is the price paid for reaching unlisted telephone numbers. Further, in

TABLE 8

CALCULATION OF AGGREGATE COMPLETION RATE*

Type of Rate	Denominator of Rate	Resulting Completion Rate
"Gross Rate"	Total Sample numbers 32005	15.9%
"Most Conservative"	Subtract ineligibles	
	Not in service (8670) Business (2364)	
	Not in areas (5163)	
	Leaves 15808	32.2%
"Still Conservative"	Subtract numbers	
	Never answered (2091)	
	Leaves 13717	37.1%
"Most Reasonable"	Subtract 44.1% of Spanish, failures, household refusals and not availables, as estimated "out of area" (3163)	
	Leaves 10554	48.2%
"Best that can be said"	Subtract "other dispositions" (644)	
	Leaves 9910	51.3%

*Computed from call records supplied by Market Opinion Research

this survey we were bound to reach a large number of households which were not located in our target neighborhoods, and a somewhat smaller number which lay outside of the cities we were surveying. They also were not eligible to participate in this study. Thus the next and "most conservative" completion rate for the survey presented in Table 8 excludes these ineligible numbers from its denominator. This more than doubles the rate.

A "still conservative" approach to the completion rate then excludes from the denominator of eligible numbers those which never were answered despite our elaborate call-back procedures. As indicated above, we suspect that the bulk of these also were not residential numbers. This placed out estimated completion rate for the survey at 37 percent.

The "most reasonable" completion rate calculated in Table 8 makes an important correction for the estimated proportion of certain numbers--those which were terminated for want of a Spanish-language interviewer, those in which a responsible informant could not be found, and household refusals--which would have been outside of our city and neighborhood lines. In Lincoln Park, for example, over fifty percent of the households we did screen proved to lie outside those boundaries; this proportion (see Table 6) is used here as an estimate of the proportion of households we could not screen that similarly would have been excluded. We are convinced that this is a conservative procedure, for hearing in an interviewer's introduction that we desired to speak only to residents of a specified area certainly would have encouraged out-of-scope respondents to hang up more quickly.

The resulting "most reasonable" completion rate for the survey as a whole was 47 percent. This is substantially below completion rates reported for most house-to-house surveys, which average now about 75 percent, and is less than rates reported by Tuchfarber and Klecka (1976), O'Neil (1976),

and Groves (1977) for their random digit dialing surveys. However, Market Opinion Research indicates that it is quite in line with the current experience of commercial firms.

The least conservative estimate of our completion rate, the "best than can be said" in Table 8, further reduces the denominator of eligible households by those in which "other" dispositions were made of the case. The bulk of these may have involved respondents who were not eligible for questioning. According to our interviewers, many of these sample numbers led to households in which neither English or Spanish was spoken; in San Francisco this included a large number of Chinese-speaking households, while in South Philadelphia Italian speakers predominated. Some randomly-selected respondents proved to be deaf, physically incapacitated, or mentally too disturbed to participate in the survey, and their cases are included in this category as well. While we have included them in the "failure" column in this report, these are all respondents who would have been missed in any standard survey.

Table 9 presents a detailed analysis of all reasons for non-completions in this survey. It is clear that the bulk of them were initial refusals by household informants; only about 12 percent of these failures can be traced to refusals to cooperate by selected respondents, and only 2 percent to break-offs once interviews began.

- - - - - - - - - - - - - -

Table 9 goes about here

- - - - - - - - - - - - - -

One reason for the relatively high proportion of household refusals in this as opposed to other surveys may have been our lack of any follow-up

TABLE 9

ANALYSIS OF NONCOMPLETIONS*

Coded Source of Noncompletion	Percent of Noncompletions		
	Total Sample	Citywide Samples	Neighborhood Samples
Needed a Spanish interviewer (est)[a]	1.4	0.6	1.6
Selected respondent never located	1.1	1.0	1.2
Selected respondent refused	12.0	9.4	13.0
Breakoff of interview	1.6	1.2	1.8
Household respondent never located (est)[a]	1.8	1.8	1.7
Household respondent refused (est)[a]	70.5	80.5	66.3
Other Disposition	11.6	5.6	14.5
TOTAL	5533	2032	3657

[a]Estimates for noncompletions in the sample areas. Estimate is based on an "out of scope" proportion of 44.1% for the total sample, 6.5% for the citywide samples, and 52.6% for the neighborhood samples, based on area screening results for completed screenings.

*Computed from call records supplied by Market Opinion Research

attempt to convert such refusals to completions. For example, those who refuse to participate in surveys conducted by the U.S. Census Bureau are recontacted by crew chiefs and other supervisors; failing that, they may receive a "personal" letter from the Director of the Census Bureau soliciting their urgently-needed participation. However, it is the experience of some survey firms that such attempts to secure the cooperation of those initially refusing to participate in a telephone survey are extremely expensive, and we choose to rely upon other randomly-selected respondents from the same sample area to "substitute" for non-cooperators.

Table 10 presents these "most reasonable" completion rates for each of the thirteen samples generated for the survey. In general, the city-wide samples produced a lower completion rate--45 percent--than the 50 percent success rate characterizing the neighborhoods. We speculate that indicating that we wished to talk to residents of their specific area encouraged respondents in our neighborhoods to participate in the study. Completion rates were highest in two Chicago neighborhoods, Lincoln Park and Woodlawn. One being a white and middle-class area and the other a poor and black community tends to discount any simple demographic explanation for these completion rates. The rate in Wicker Park in Chicago was depressed considerably by our Spanish-language interviewing problem there. The average completion rate was lowest for samples in San Francisco, and the San Francisco city-wide sample produced the lowest completion rate of all.

- - - - - - - - - - - - - -

Table 10 goes about here

- - - - - - - - - - - - - -

One of the major disadvantages of random digit dialing telephone surveys is that we know little about those who did not participate in the survey.

TABLE 10

MOST REASONABLE COMPLETION RATES FOR SAMPLES AREAS*

Sample	Completion Rate
San Francisco City	40.5
Sunset	42.7
Visitaction Valley	40.6
Mission	52.6
Philadelphia City	41.7
West Philadelphia	52.1
South Philadelphia	45.4
Logan	45.6
Chicago City	51.3
Lincoln Park	62.9
Wicker Park	42.0
Woodlawn	61.9
Back of the Yards	49.9
TOTAL	48.2

*Computed from call records supplied by Market Opinion Research

In house-to-house surveys, interviewers can glean a great deal of information about those who refuse to participate in them, and estimates even can be made of the race and class status of householders who are never found at home. Telephone interviewing procedures have a distinct disadvantage when they fail, for we do not even know where those noncompletions occur. Thus, we cannot characterize respondents and non-respondents to this survey, nor examine the distinctive characteristics which seem to predict non-cooperation.

This limitation of telephone surveys lends special importance to more indirect and inferential evaluations of the quality of the data when non-cooperation is frequent. The problem is that low completion rates may signal difficulties with the representativeness and analytic utility of the data. We are concerned about the representativeness of data when we wish to use a sample to make estimates of the distribution of something--like levels of fear--in a city or neighborhood. We are concerned about the analytic utility of data when we wish to investigate the relationship between variables measured in the survey and generalize about their co-variation in the population. These are somewhat different issues, and problems with the representativeness of a sample do not necessarily degrade the analytic utility of the data. Often, for example, we deal with data which purposively overrepresents population groups (e.g. high-income blacks, Spanish-speaking women) in order to generalize more accurately about them. On the other hand, high refusal rates suggest that people who did agree to be interviewed are perhaps systematically different, or unusual, or represent distinctive clusters of personal attributes. Thus, the low completion rates achieved by this survey forces us to pay careful attention to both of these issues, and to document as fully as possible the extent to which the resulting data reflect the populations from which they were drawn.

F. INDICATORS OF SAMPLE AND DATA QUALITY

In survey research one is always interested in the extent to which samples accurately reflect, or "represent," the population from which they were drawn. However, reliable criteria on which to judge the representativeness of a sample usually do not exist. We do surveys because things of interest are unknown. In addition, comparative measures of the attributes of populations are subject to errors which are both similar to and different from our own. Finally, Americans are an extraordinarily mobile people. Approximately twenty percent of the American population moves each year, rendering any criterion describing what a sample "ought to look like" suspicious if it was not itself determined in a timely fashion.

In this case, our problem is one of estimating the representativeness of the thirteen independent city and neighborhood samples of respondents we assembled through our telephone interviews. The only available and reliable descriptions of the city populations from which they were drawn, those derived from the U.S. Census, were fully seven years out of date when our interviews were conducted. However, this Census data still is of some value in assessing the quality of our sampling and interviewing procedures at the city level. It will be less useful in the case of our neighborhood samples. Neighborhood boundaries were defined after extensive interviews with area residents, and do not correspond closely to official geographical subdivisions of the cities. Further, we chose many of our neighborhoods for study because they were known to be areas undergoing rapid social and economic change. In some (e.g. Wicker Park), Latino populations are growing, while in others (e.g. Lincoln Park) white middle-class residents are beginning to predominate.

Table 11 presents several indicators comparing the city-wide samples of

- - - - - - - - - - - - - - - -

Table 11 goes about here

- - - - - - - - - - - - - - - -

respondents we interviewed in 1977 with the characteristics in 1970 of the populations (18 years of age and older) of the three cities from which they were drawn. Several notable features of the samples are apparent in Table 11. First, our respondents and the city censuses are broadly comparable on two dimensions--the proportions of the populations that are foreign-born, and who own their own homes. The city surveys slightly but consistently uncovered somewhat fewer elderly respondents than lived in these cities in 1970. Our San Francisco sample in particular seems to be a bit young. The Philadelphia sample appears to overrepresent home owners, but our 1977 survey figure for that is much closer to the Census Bureau's home-owner estimate for their 1974 victimization survey in that city (Law Enforcement Assistance Administration, 1977).

Those are variables for which we would expect no <u>substantial</u> change to have taken place during the 1970-77 period. The same is not true of the racial composition of the cities, and racial changes widely attributed to the cities of Philadelphia and Chicago are reflected in Table 11. We are most knowledgeable about estimates of the population of Chicago; our survey in that city set the community's black population at 42 percent of the total, which is exactly on the most popular local mark. The Chicago Urban League (1978) estimates that the city was 38.5 percent black in 1975, up from 32.8 percent in 1970. Projecting that rate of population change forward into 1977 yields a population estimate of 41 percent black, just one percent short of our figure for the telephone sample. No similar data are available for

Table 11

1977 SURVEY AND 1970 CENSUS DATA FOR CITIES*

Variable	Chicago		Philadelphia		San Francisco	
	Survey	Census	Survey	Census	Survey	Census
Percent White	56	71	61	70	77	76
Percent Own Home	36	35	53	35	33	33
Percent Family Income Over $15,000	37	17	28	13	38	15
Percent U.S. Born	87	85	94	91	82	76
Percent Over 65 Yrs.	12	16	12	17	9	18
Percent High School Graduates[A]	76	52	80	47	92	78

*Base for census data on persons is population 18 years of age and older. Base for home ownership is number of households.
Data drawn from: U.S. Bureau of the Census. Characteristics of the Population, 1970 Census of Population, and Housing Characteristics for States, Cities, and Counties, 1970 Census of Housing.

[A]of those 25 years of age and older. Survey respondents indicating they completed "technical or vocational" school as their highest level of educational achievement are excluded to facilitate the comparison of survey with census figures.

Philadelphia, but the Census Bureau's estimate for 1974 of the size of the white population in that city lay just midway between the 1970 and 1977 figures given in Table 11, 66 percent (Law Enforcement Assistance Administration, 1976: Table 12). The fact that our survey samples were somewhat younger than the 1970 Census count for these cities is in accord with these figures on racial change, for urban blacks as a whole are somewhat younger than their white counterparts.

There is apparent disagreement between the two data sources about two other key population figures, income and education levels for the cities. The income differences apparent in Table 11 can be attributed to inflation during the 1970-77 period, however. In each city the proportion of respondents indicating yearly family incomes exceeding $15,000 was slightly more than double the 1970 figure in the 1977 survey. During that time, however, the proportion of American families reporting incomes over $15,000 rose from 22 to 50 percent nationally, a 125 percent increase (U.S. Census Bureau, 1977: Table 708). In our city surveys, in comparison to census counts in 1970, the average rise was 129 percent. Thus, we judge the samples interviewed over the telephone in 1977 to represent satisfactorially high and low income groups in the populations of the three cities.

We are less certain of the representativeness of the samples with regard to education. Table 11 indicates substantial differences in the 1970 census and 1977 sample estimates of the proportion of city residents (twenty-five years of age and older) who were at least high school graduates. Substantially larger proportions of our respondents claimed high school diplomas, and we are not able to discount the observed differences. There is an upward secular trend in the proportion of high school graduates in the population. Between 1970 and 1977 the proportion of American population at least graduating from high school increased by 16 percent (U.S. Bureau of the Census, 1977). That trend cannot account for all of the differences

between the two observations documented in Table 11, however. In Philadelphia the 1970 Census and 1977 survey differences would indicate a 70 percent rise in the proportion of high school graduates, while in Chicago it would indicate a 46 percent rise. The difference between the 1970 Census in San Francisco and our 1977 survey there is only 18 percent, however, a figure in line with national trends.

Table 12 assesses the quality of the data in a somewhat different fashion. At the conclusion of each interview, interviewers were asked to

- - - - - - - - - - - - - - - -

Table 12 goes about here

- - - - - - - - - - - - - - - -

rate the process they just had completed along several dimensions. Table 12 reports, first, the proportion of respondents whose English seemed "poor." Those constituted relatively few of our cases, only 1.7 percent. Somewhat more (2.7 percent) were judged "uncooperative" by their interviewer, and an equal number were suspected by the interviewers of giving information during the interview which was "inaccurate." About one in twenty were judged "uninterested" in the interview.

These proportions, which may signal difficulties in the validity of the data collected, are relatively small. They do not seem to point to data problems in any particular sample: only the Visitacion Valley sample scores over the mean on all four dimensions, while the remainder are mixed or (in Logan and for San Francisco City) fall below the mean for all respondents.

In addition to interviewer judgments, it is possible to assess the quality of a data set by examining the extent to which missing information will constitute a problem at the analysis stage. There are several ways that missing data for variables can occur in a survey. Respondents may

Table 12

INTERVIEWER RATINGS OF DIFFICULTIES IN THE INTERVIEWING PROCESS*

Sample	Percent--Respondent's English "Poor"	Percent--Respondent Judged "Not Very Cooperative"	Percent--Information Given by Respondent Judged "Inaccurate"	Percent--Respondent Judged "Not Interested"
San Francisco City	0.7	1.5	1.7	3.0
Visitacion Valley	2.5	2.7	5.6	5.8
Sunset	2.4	2.9	3.8	4.9
Mission	1.0	1.5	4.0	5.5
Philadelphia	0.7	3.9	3.3	6.7
West Philadelphia	1.6	2.9	2.4	6.7
South Philadelphia	2.4	2.9	1.8	5.3
Logan	1.0	2.0	1.0	2.0
Chicago City	1.5	3.0	2.6	5.4
Lincoln Park	2.0	2.0	1.8	4.2
Wicker Park	2.9	2.4	4.0	4.0
Woodlawn	1.0	3.0	4.5	5.0
Back of the Yards	1.0	4.5	1.0	4.5
Totals	1.7	2.7	2.9	5.0

*Base all unweighted interviews (N = 5121)

legitimately answer "don't know" to a particular item, or think that it is inappropriate to their case. One duty of the interviewer in most instances is to discourage the selection of don't know responses, and to re-prompt respondents using the desired response categories whenever this occurs. However, in some cases respondents may in fact "not know," or may continue to adhere to their initial response, and in those situations their honest answers are properly recorded. Missing data also will result when interviewers fail to ask a particular question, or to record a response, or when respondents insist on some response which in no way can be accommodated in the pre-printed categories available for a closed-response question. Finally, parts of a questionnaire may be void of all responses because a "breakoff" occurred at the insistance of the respondent.

Figure 3 illustrates the extent to which missing data haunts the analysis of our telephone survey. It charts the proportion of responses for whom data is missing on fourteen selected attitudinal items and fourteen demographic questions. The attitudinal items were scattered systematically throughout the questionnaire, while the demographic questions all were concentrated at the end of the instrument. As Figure 3 indicates, in almost

- - - - - - - - - - - - - - - -

Figure 3 goes about here

- - - - - - - - - - - - - - - -

two-thirds of all cases there were <u>no</u> missing values recorded either for the demographic or attitudinal items, and that very few respondents were coded as missing on more than two or three of the items in each set. About 1.4 percent of the respondents were missing <u>all</u> fourteen demographic measures; were those who terminated the interview. In no case was a respondent coded as missing on more than ten of the attitudinal items, some of which also fell toward the end of the instrument.

FIGURE 3

SUMMARY OF MISSING DATA IN THE TELEPHONE SURVEY:
NUMBER OF RESPONDENTS WHO WERE MISSING DATA ON
FOURTEEN SELECTED DEMOGRAPHIC AND ATTITUDE ITEMS

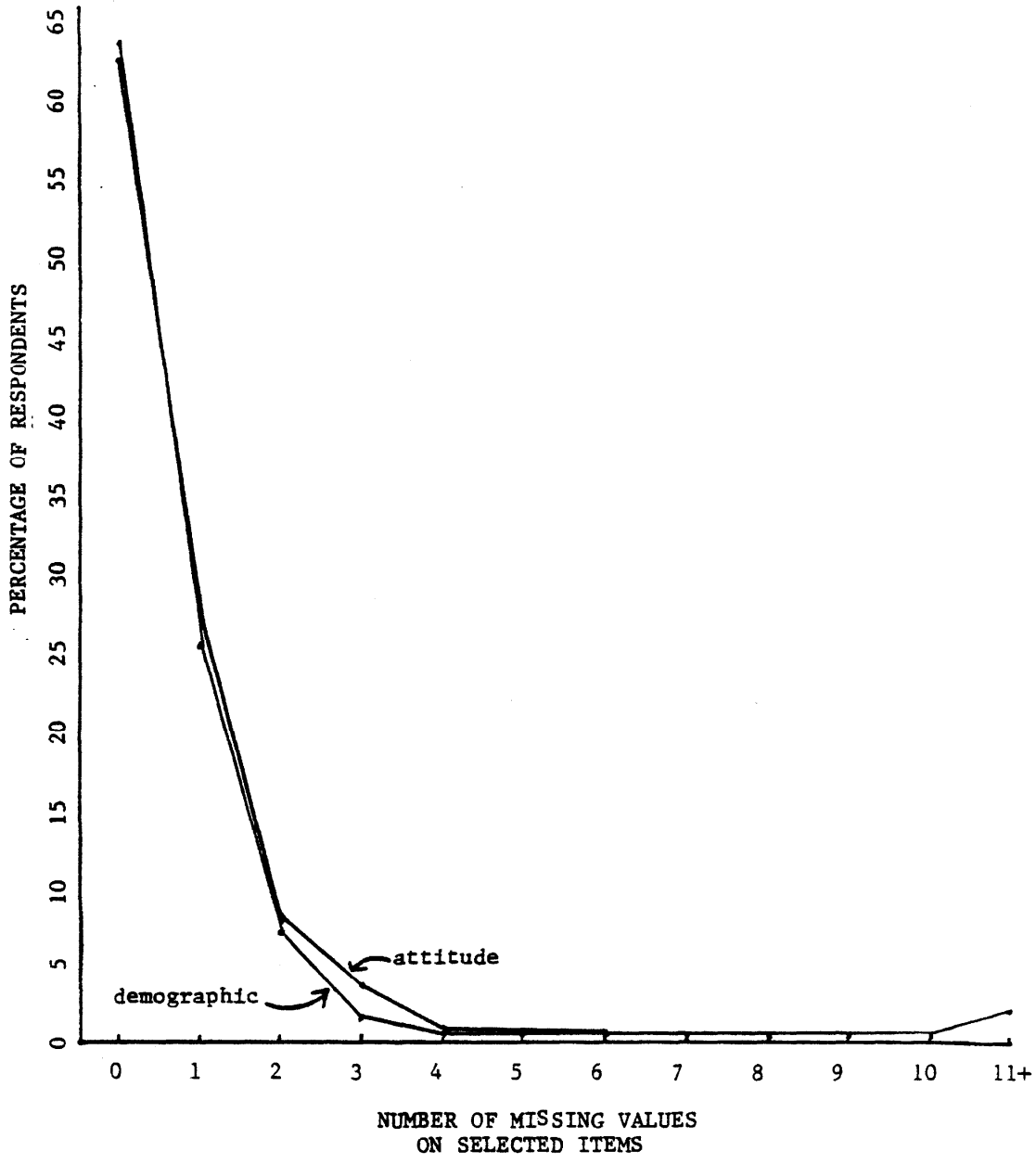


Table 13 presents a break-down of missing data cases by sample, for the three cities. It details the average number of missing-data variables for each respondent in each of the thirteen samples. Over the entire group, responses to an average of 0.8 of the fourteen demographic and 0.6 of the fourteen selected attitudinal items were coded as missing. There appears to be a slight tendency for respondents in Philadelphia to have missed items in the demographic section of the questionnaire, or to have broken off questioning before that point. However, this concentration of missing data is not to be found among the attitudinal items; in those cases, Philadelphia seems to have the best item-completion record of the three cities.

- - - - - - - - - - - - - - - -

Table 13 goes about here

- - - - - - - - - - - - - - - -

In addition to these judgments of data quality and counts of missing data, it is possible to make a systematic assessment of the quality of one piece of data collected in the survey. In the course of validating for a ten-percent sample of respondents that interviews were conducted as specified, MOR supervisors asked respondents in their re-interviews, "How many years have you personally lived in your present neighborhood?" This duplicated a question asked on the first call, and gives us a more precise estimate of the test-retest reliability of this variable.

Table 14 presents a cross-tabulation of the responses to this item, grouped in five categories. In all, 8.6 percent of respondents in the same

- - - - - - - - - - - - - - - -

Table 14 goes about here

- - - - - - - - - - - - - - - -

Table 13

MISSING DATA FOR AREA SAMPLES

Sample	Average Number of Responses Missing—of Fourteen Items in Each Category: Demographic	Attitudinal
San Francisco City	.54	.62
Mission	.49	.74
Visitacion Valley	.74	.50
Sunset	.63	.62
Chicago City	.83	.63
Back of the Yards	.93	.63
Woodlawn	.84	.52
Wicker Park	.95	.69
Lincoln Park	.61	.67
Philadelphia City	1.03	.45
Logan	.77	.55
South Philadelphia	1.09	.35
West Philadelphia	1.12	.49
Total	.80	.58

Table 14

TEST-RETEST OF LENGTH OF RESIDENCE MEASURE,

USING THE TEN-PERCENT VALIDATION SAMPLE*

Original Interview: Number of Years	Validation Interview: Number of Years					
	0-1	2-3	4-5	6-10	11+	(Total)
0-1	62	10	2	1	0	(75)
2-3	3	73	1	2	2	(81)
4-5	4	1	49	3	0	(57)
6-10	1	0	2	85	4	(92)
11+	3	3	0	3	210	(219)
(Total)	(73)	(87)	(54)	(94)	(216)	(524)

Total Nonagreements = 45/524 (8.6%)

Nonagreements of More Than One Category = 18/524 (3.4%)

*Total validations in all three cities

households gave different answers to this question. Only 3.4 percent of all respondent-pairs gave us answers that were discrepant by more than one category. We judge this to be evidence of acceptable test-retest reliability for this item and, by inference, for at least similar demographic items in the questionnaire.

G. CONCLUSION

In summary, it appears that the Center for Urban Affairs' telephone survey was a successful experiment. Several aspects of the survey were pioneering: to our knowledge no one before has attempted to use random digit dialing techniques to sample community areas, and there have been few surveys like ours which have been of comparable magnitude. Both of these aspects of the survey were responses to the substantive demands of the problem at hand, and the resulting data appears to be useful in shedding light upon those problems. A combination of our use of the telephone to gather the data and our need to screen households for geographical location appears to have reduced the completion rate for the survey. However, the resulting data match reasonably well our best estimates of what it "should" look like in demographic profile. Interviewer's ratings of respondent cooperation and truthfulness indicate that those we reached were engaged by the questioning, and this analysis of the quality of the resulting data suggests that it is quite high. Further, our efforts to generate multi-item scales from items designed to tap the central concepts which lay behind the survey instrument have been quite successful. Our data scaling activities will be detailed in another report; however, the high reliability of the measures produced from this survey data reinforces our conviction that the survey was successful indeed.

FOOTNOTES

1. Telephone companies generally let prefixes become approximately 75 percent full (45-55 percent with listed numbers, 20-25 percent with unpublished numbers), whereupon "relief demand" leads them to open a new prefix. This has been made much simpler by the abandonment of alphabetic prefix names and the isolation of calling areas from one another in area code regions.
2. A central office area is a geographical region served by a telephone company (area) office within a city. In Chicago there are, for example, 30 central office areas, while in San Francisco there are 12. In general, all telephones physically connected within a central office area must use a number prefix uniquely associated with that area; no telephones outside of an area can employ its prefixes, and numbers within it must utilize one of its prefixes. This is a mechanical and electronic consideration, determined by telephone company switching systems. In the areas we studied, prefixes serving a central office area seemed to be scattered throughout it, not geographically concentrated within the exchange area. Thus, if a researcher is attempting to dial randomly into an area smaller than a central office area, some of the numbers generated will reach telephones outside of the target area. The smaller the target area is in relation to the central office area (for prefixes appear to scatter randomly), the greater this problem will be. Target areas that span central office areas greatly magnify the problem, and are to be avoided if possible.
3. For example, these include "test numbers", some of which merely ring, enabling company personnel to test telephones.

4. In general, non-working numbers ring either a recording or an operato who passes along a message to that effect. Occasionally, there are malfunctions in this procedure. If one is calling long distance, there is no charge for reaching a non-working number. This makes it relatively inexpensive to use a telephone to test hypotheses about the existence of banks of non-working numbers.

5. A note about recent movers. The sampling frame for this survey is telephone numbers. Thus, if a call reached a recording which indicated that the former subscriber to that number now could be found at a new number (probably because the household had moved), we did not follow-up that suggestion. This has practical advantages for neighborhood surveys, for movers who did not "take their telephone number with them" probably moved out of their old central office area, and thus out of our target area.

CITATIONS

Chicago Daily News. 1977. "Unlisted Phones: It's Privacy by Numbers." October 3.

Chicago Urban League. 1978. Where Blacks Live: Race and Residence in Chicago in the 1970's. Chicago: Spring, 1978.

Furstenberg, Frank. 1971. "Public Reaction to Crime in the Streets," The American Scholar 40 (Autumn): 601-610.

Garofalo, James. 1977. Local Victim Surveys. Washington DC: National Criminal Justice Information and Statistics Service, Law Enforcement Assistance Administration.

Groves, Robert M. 1977. "A Comparison of National Telephone and Personal Interviews." Paper presented at the 1977 meeting of the American Association of Public Opinion Research.

Klecka, William R., et al. 1976. "Reducing the Cost of Victim Surveys," in Wesley G. Skogan (ed.) Sample Surveys of the Victims of Crime. Ballinger, pp. 207-222.

Law Enforcement Assistance Administration. 1976. Criminal Victimization Surveys in Chicago, Detroit, Los Angeles, New York and Philadelphia. Washington DC: U.S. Government Printing Office.

O'Neil, Michael J. 1976. "The Impact of Community & Individual Characterization on Citizen Corporation with the Police," Evanston IL: Ph.D. Dissertation, Northwestern University.

Powell, A. Elizabeth, and William R. Klecka. 1976. "Does It Matter Who Is Missed in Telephone Surveys." Paper presented at the 1977 meeting of the American Association of Public Opinion Research.

Rich, Clyde L. 1977. "Is Random Digit Dialing Really Necessary." Journal of Marketing Research 14 (August): 300-305.

Skogan, Wesley G. 1977. "Public Policy and the Fear of Crime in Large American Cities," in John A. Gardiner (ed.) Public Law and Public Policy. New York: Praeger, Chapter 1.

Tuchfarber, Alfred, and William R. Klecka. 1976. Random Digit Dialing: Lowering the Cost of Victimization Surveys. Washington DC: The Police Foundation.

U.S. Bureau of the Census. 1977. Statistical Abstract of the United States. Washington DC: U.S. Government Printing Office.

Warner, W. Lloyd, Marchia Meeker, and Kenneth Eells. 1949. Social Class in America. Chicago: Science Research Associates.

APPENDIX A

CODING OPEN-ENDED ITEMS
(by Gary Jason)

The RTC/FOR telephone survey contained seven questions which were "open ended." That is, there were no predetermined categories printed on the survey instrument for use by the interviewer. Responses to these open-ended questions were written in full on the questionnaire and left for post-interview coding. The coding was done on 80-column coding sheets, which later were keypunched and merged into the closed-ended data files.

The first items coded were the community organizations to which the respondent belonged. The first step in coding community organizations was to make up a master list of all named organizations in each community. These lists were then alphabetized. Community organizations which were spelled incorrectly along with organizations which were miss-named but identifiable were given the same identification number as the "proper" organization. Coding allowed for up to four organizations.

The kind of crime activity dealt with by the organization was coded from a list of forty-nine possible crime activities. Each organization was given up to two codes for the activity. This was the final phase of the telephone survey coding. All codes were validated by establishing agreement on them by two different coders.

As the coding of the first city (Philadelphia) progressed, the list of crimes coded originally as "other" burgeoned. As was the procedure throughout the coding process, index cards were made on all not immediately-codable responses. The coders later decided upon which codes would have to be added to the original list(s) based upon the frequency of "other" responses.

The follow-up question, "What did you read or hear about it?" (crime mentioned), was only coded for the presence or absence of details. This provided a list of all questionnaires where details were mentioned, for possible inspection in the future.

A list of rape prevention strategies was employed to code the questions: "Is there anything else you can think of that would help prevent rape?" (up to two responses coded), and, "From all the things you can think of, which one do you feel would work best to prevent rape?" (one response coded). The original list contained twenty-one prevention strategies including an "other" and "not-ascertained" category.

The final list, which was completed by the end of the Philadelphia coding, included fifteen additional responses, plus changes in several on the first list. Most of these changes were expansions in the wording of the code. Again the added codes were based upon the response frequency in the phone survey. When the final coding categories for the rape question were complete, all prior "questionable" codes were rechecked, and coded appropriately.

All coding of the respondents' occupations was based upon the seven point scale for measuring status characteristics developed by Warner, et al. (1949). Additional occupations were added to the Warner scale only after they had been agreed upon by at least two different coders. The primary questions in assigning an occupation to a given category were: 1) How much education does the occupation require? 2) How much income is involved? 3) Is the occupation prestigious? 4) Is the occupation social-service related? In addition to specific occupations, a number of responses fell into the categories: 1) corporation or industry, 2) can't tell; not ascertained and 3) refused.

Ten percent of all interviews were coded a second time in order to test the reliability of the coding. Data on coding errors detected in this re-check are found in Table A-1.

Table A-1 goes about here

The total amount of disagreement between the first and second coding was 1.8 percent for the 10 percent sample. That is, there was 98.2% agreement between all pairs of codes. All validating was done "blindly": i.e., the first coding was not examined before the second coding was completed.

There was little discrepancy between the "best" and the "worst" coders. The first-ranking coder had an error rate of 1.4%, whereas the sixth ranking coder had an error rate of 2.6%. Much of this cohesiveness in coding was due to the constant consultation between coders on ambiguous coding judgments.

Error rates for individual questions reflected the difficulties inherent in various types of coding. That is, whereas the coding of organizations was relatively straightforward (hence yielding only a 0.1 percent error), the coding of occupation required more subjective interpretations (hence a larger "error" term: 5.9 percent). The standard deviation for discrepant occupation codes (eliminating "other", "non-existant", "corporation or industry" and "can't tell; not ascertained" because of their nominal--not ordinal--meanings) was 1.5. This means that on the 5.9 percent of the occupation codes that coders differed upon, that difference averaged only one and one-half scale points.

The breakdown of percentages of individual coders by individual questions bears out the notion that the unambiguous questions (e.g. organization, crime type) gave coders less trouble than the "rape" and "occupation" questions, which often required more judgment.

Table A-1

OPEN-ENDED CODING ERROR ANALYSIS

PERCENT ERROR BY INDIVIDUAL CODERS

Coder	Total Questionnaires Coded	Total Questionnaires Validated	Total Codes Validated	Percent of Codes in Error
1	731	101	1212	1.4%
2	1227	107	1284	1.6
3	316	35	420	1.7
4	1565	153	1836	1.8
5	825	84	1008	2.1
6	451	42	504	2.6
Total	5115	522	6264	1.8

Total percent error for 522 questionnaires and 6264 codes = 1.8%

PERCENT ERROR BY INDIVIDUAL CATEGORIES

Question	Percent
Identification number	0.0%
Organizations (up to four)	0.1%
Crime Listed (up to two)	2.4%
Crime Details Mentioned (yes or no)	1.0%
Other Rape Strategies (up to two)	2.4%
Best Rape Strategy	4.4%
Occupation	5.9%