U.S. Department of Justice
Bureau of Justice Statistics

The District of Columbia Crime Victimization

Study Implementation

by

Brenda G. Cox

Danny R. Allen

Jane W. Bergsten

James J. Collins

Dale S. DeWitt

Research Triangle Institute

Research Triangle Park, N.C.

December 1983

ABSTRACT

The 1982 Supplemental Appropriations Bill allocated funds for a crime victimization study in the District of Columbia. The primary objective of the study was the measurement of the extent of crime in the District of Columbia and the impact of crime on the quality of life in the District. Of secondary interest was the degree to which Congressional employees working in the Capitol Hill area are subject to victimization and the extent to which victimization and the fear of victimization have decreased their work productivity. The District of Columbia Crime Victimization Study was conducted by the Research Triangle Institute under a contract from the Bureau of Justice Statistics. This final report summarizes the results of Phase II of the study. Phase I involved the design of survey procedures and instruments and the specification of methods for sample selection, data collection, data processing, and statistical analysis for the study. These specifications were implemented in Phase II. The data collected in the study were used to prepare a Report to Congress and the District of Columbia Government on crime victimization in the District of Columbia.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The 1982 Supplemental Appropriations Bill allocated funds for a crime victimization study in the District of Columbia. Under contract to the Bureau of Justice Statistics (BJS), the Research Triangle Institute (RTI) designed and implemented the District of Columbia Crime Victimization Study. The primary objective of the study was to determine the extent of crime in the District of Columbia Standard Metropolitan Statistical Area (DC-SMSA) and the impact of crime on the quality of life in the District. A secondary objective was to determine the degree to which Capitol Hill employees are subjected to victimization and the extent to which victimization and the fear of victimization have decreased work productivity. A major focus of both objectives was the measurement of crime victimization in the work place and in travel to and from work.

To implement these objectives, RTI conducted two surveys: the District of Columbia Household Victimization Survey (DCHVS), which measured crime victimization occurring to residents of the DC-SMSA, and the Capitol Hill Employees Victimization Survey (CHEVS), which measured crime victimization occurring to Capitol Hill employees. The relevant information about how RTI conducted these surveys includes:

- From what groups were the samples selected?
- What questions were asked in the interview?
- How was the interview conducted?
- What information resulted from the study? and
- How are the data being reported?

Before answering these questions, it is important to note that telephone interviewing was used in both surveys.

The sample of households to participate in the DCHVS was selected by first creating a list of all telephone exchange codes used in the DC-SMSA. This exchange code is the area code and the first three digits of the seven digit telephone number. All possible four digits were added to these exchange codes to create a list of all telephone numbers allocated to the DC-SMSA by the local telephone companies. A sample of numbers from this list was randomly selected and telephone interviewers dialed these numbers and determined whether the number was associated with a residence or not. For residential telephones, the interviewer individually surveyed each household member who was 14 or older, beginning first with adult members of the household. Responses for 12 and 13 year olds were obtained from their parents. This procedure is similar to that used by the National Crime Survey, from which RTI borrowed many procedures for this study.

Using the random digit dialing procedure, all households with telephones had an opportunity for inclusion in the study. Unavoidably, the DCHVS is subject to undercoverage of nontelephone homes. Survey resources precluded the use of personal interviews for nontelephone households which would have been the only way to obtain their response. However, census data were used in the estimation process to compensate for these "lost" households by weighting the data prior to analysis. The distribution of the weighted data was made similar to that of the general population for factors such as age, race, and sex, which are correlated with telephone ownership and with crime victimization as well.

The sample of Capitol Hill employees was selected from employee lists of the Senate, the House of Representatives, the Library of Congress, the Congressional Budget Office, the Office of Technology Assessment, and the Architect of the Capitol. Prior to the interview, RTI mailed each sample

employee a letter describing the survey and included a postcard which the employee was asked to mark with the most convenient time to call.

Both surveys used the same questionnaire. The interviewer began by asking a set of lead-in questions about crime and participation in community programs to combat crime. Next the individual was asked to report crimes that had occurred to him/her since January 1982. Example crimes and example crime locations were read by the interviewer to jog the memory of the respondent. Next, the interviewer obtained details about each crime that the respondent mentioned. The interviewer closed by asking general information questions such as age, race, and sex, and the characteristics of the dwelling in which the respondent lived.

The interview was conducted using a computer assisted telephone interview procedure. Rather than using a printed questionnaire, the interviewer read the questions as they were displayed on a computer viewing screen and simultaneously recorded the respondent's answers. This process gives greater control over the interview and reduces the length of time required to complete the interview.

The sample data were analyzed to describe the characteristics of victims and the effect of victimization on their lives. The impact of crime was evaluated for the various types of crimes as well. Examples of the kinds of questions for which answers were sought include:

- What types of individuals tend to be victimized?

- What percent of crime victimizations result in injury?

- To what extent do economic losses result from crime victimization?

- How frequently do work place victimizations occur? and

- Are certain categories of employees (such as women for instance) more likely to experience work-related victimizations?

To answer questions such as these, data analysts examined tabular summaries of the data. These tables were created for population aggregations of sufficient size so that the information could not be linked to particular individuals. Comparisons were made between Capitol Hill employees' victimization and that of the DC-SMSA population. In addition, the victimization experience for DC-SMSA residents was compared to that of the entire nation. This later comparison was made using data collected as a part of the National Crime Survey.

RTI prepared a report to Congress and the District of Columbia Government describing the results of these analyses. In addition, a public use data file was developed for the DCHVS data. To preserve the confidentiality of the respondent data, all identifying information was removed or encrypted prior to delivery of the data.

CHAPTER 1. INTRODUCTION

The 1982 Supplemental Appropriations Bill allocated funds for a crime victimization study in the District of Columbia. Of primary concern was the extent of crime in the District of Columbia and the impact of crime on the quality of life in the District. A secondary concern was the degree to which Congressional employees working in the Capitol Hill area are subject to victimization and the extent to which victimization and the fear of victimization decrease their work productivity. The legislation specified that the study would be conducted by the Bureau of Justice Statistics (BJS) in conjunction with the Bureau of the Census and in consultation with Congress. Under contract to BJS, the Research Triangle Institute (RTI) designed and implemented the District of Columbia Crime Victimization Study. The instrument for the study was developed by the Bureau of Social Science Research (BSSR). As a part of the Crime Survey Redesign consortium, BSSR has been investigating alternative instrument designs and data collection procedures.

The District of Columbia Crime Victimization Study had two phases. Phase I involved the design of survey procedures and instruments for use in the study, which included determining the study goals, developing the survey instrument, and specifying methods for sample selection, data collection, data processing, and data analysis for the study. These specifications were implemented in Phase II of the study. The data collected in the study were used to prepare a Report to Congress and the District of Columbia Government on crime victimization.

The District of Columbia Crime Victimization Study contained two survey components: the District of Columbia Household Victimization Survey (DCHVS), which measured crime victimization for residents of the District

of Columbia Standard Metropolitan Statistical Area (DC-SMSA), and the Capitol Hill Employees Victimization Survey (CHEVS), which measured crime victimization for Capitol Hill employees. The objectives of the study were to measure crime victimization for the DC-SMSA and for Capitol Hill employees and to make comparisons between the two groups. Within the DC-SMSA, separate estimation was required for the District proper and the outlying suburbs. Because of differences in the instruments and the survey design procedures used in obtaining victimization data, it should be emphasized that <u>comparisons of DCHVS or CHEVS crime victimization rates with national rates are not appropriate</u>.

The target population for the DCHVS was the civilian, noninstitutionalized residents age 12 and older of the DC-SMSA and those residents of adjacent areas that share telephone exchange codes with the DC-SMSA. The 1980 Census definition of the DC-SMSA was used in the study. Under this definition, the DC-SMSA includes the District of Columbia; Charles County, Montgomery County, and Prince George's County in Maryland; Arlington County, Fairfax County, Loudoun County, and Prince William County in Virginia; and the independent cities of Alexandria, Fairfax, Falls Church, Manassas, and Manassas Park in Virginia.

The target population of the CHEVS was the employees of the Congress, specifically employees of the Senate, the House of Representatives, the Architect of the Capitol, the Library of Congress, the Congressional Budget Office (CBO), and the Office of Technology Assessment (OTA). Elected members of Congress were not included in the CHEVS.

In both surveys, the respondents were asked to report victimizations that happened to them during the period from January 1, 1982 to the date of the interview. Since data collection occurred from May through August of 1983, sample individuals reported victimization data for a minimum of 16

2

months and a maximum of 19½ months. For analysis purposes, a common time period was needed. Therefore, it was decided that only victimizations occurring in the fixed time period from May 1, 1982 to April 30, 1983 would be included. Victimization data were included or excluded from the analysis based upon the date of occurrence; hence, it was important to obtain the month in which victimizations occurred.

The early 1982 months were included in the reference period since May 1 has few salient features as a reference point. A second reason for including January to April 1982 was the presumption that telescoping into the reference period might occur; including these four months for data collection but excluding these data from analysis would remove some of this telescoping bias.

The end of the reference period was set at the interview date since it was thought intuitively appealing to allow respondents to discuss their more recent experiences. It should be noted that forward telescoping of events from the analysis time period into months after April 1983 may have occurred for respondents who were interviewed during the latter part of the data collection period. Although such forward telescoping would result in underestimation of crime during the analysis time period, it should not affect the comparisons between population subgroups since the sample was evenly distributed over the data collection period.

Except for screening questions needed for data collection purposes, the DCHVS and the CHEVS used the same data collection instrument. This instrument was a streamlined version of an experimental instrument that was developed and tested in a pilot study in Peoria, Illinois by the Crime Survey Redesign consortium. This experimental instrument differs from that used in the National Crime Survey (to which it was compared in the pilot study) in that the screener questions cover more types of incidents and all

3

respondents within the household are screened for household crimes. The additional screener questions are intended to jog the respondent's memory about the NCS crime types while discussing the new crime types. Since household crimes may have differing degrees of saliency for household members, asking household screening questions of all respondents should improve reporting but at the expense of duplicate reporting. In the Peoria Study, the effect of this duplication was removed by weighting the incident data based upon the number of reports of the incident.

The questionnaire used for the District of Columbia Crime Victimization Study had similar features, as well as a cueing approach to obtain victimization reports. Originally developed for use with a mail questionnaire, this cueing approach first lists various types of crimes and asks the respondent, "Right off, can you think of a time during 1982 or 1983 when any of these things happened to you?" After recording the immediate responses, the interviewer then reads a list of example crimes and example crime locations. The respondent is instructed to stop the interviewer whenever he/she thinks of a crime that has not been previously mentioned. Each time a cue provokes a response, the respondent's description of the incident is entered into the list of events. The interviewer then probes for other similar events. From initial pretests, this cueing approach appears to elicit more reports of criminal incidents than the NCS screener.

A modified version of the NCS incident form was also developed for use in this study. The modified incident form is divided into several sections. The first section serves a "verification" purpose in the sense that it determines the date of the incident, the type of crime that occurred (including non-crimes), and the person or persons involved. Only for crimes that occurred within the analysis time period of May 1, 1982 to

April 30, 1983 were the remaining sections of the incident form completed. These sections obtained information about the characteristics of the criminal incident and the associated offenders.

Data collection for both surveys was by telephone using computer assisted telephone interviewing (CATI) procedures. RTI developed CATI versions of the instruments that BSSR provided and developed the household roster and screening portion of the instrument. The DCHVS data collection procedures were similar to the National Crime Survey except that it was conducted via a random digit dialed telephone survey. That is, victimizations were obtained for all individuals 12 years old and older within sample households, with the data for 12 and 13 year olds obtained by proxy and 14 year olds and up interviewed individually to obtain their victimization data. In the CHEVS, only the sampled employees were interviewed.

Much of the analysis focused on simple descriptive statistics, such as the victimization rates per population subgroup. Results for the two surveys were compared and tested. In addition, substantive issues were investigated regarding the differential effect of victimization for D.C. city residents versus D.C. suburban residents and DC-SMSA employees versus Capitol Hill employees. These analyses required the production of a type-of-crime recode and the determination of whether or not each reported incident fell within the analysis period of May 1, 1982 to April 30, 1983. Comparisons of the DC-SMSA to the nation were made using NCS data. The results of these analyses of D.C. crime data and NCS national comparison data formed the basis for the Report to Congress and the District of Columbia Government.

CHAPTER 2. SAMPLE DESIGN FOR THE DISTRICT OF COLUMBIA CRIME SURVEYS

The sample designs for the two survey components of the District of Columbia Crime Victimization Study were straight forward applications of standard sampling methodology. The most difficult aspect of the design was obtaining estimates for the parameters that affected data collection costs. An example of such a parameter is the expected number of victimizations per 1,000 persons that would be reported in the study. In deriving sample sizes for the surveys, values were estimated for these parameters. The number of assumptions needed to produce these estimates introduced uncertainty into the expected yield of completed interviews and victimization reports that would be obtained for the two surveys. For this reason, the sampling was set up in three waves so that early results could be used to obtain survey estimates for the parameters that pertained to yields. Using these estimates, the sample size specifications were reevaluated and the proposed sample sizes for the two surveys revised downward to reflect increased survey costs. Specific details of the sampling and weighting are provided by memoranda contained in Appendix A. The remainder of this chapter summarizes the general features of the selection and weighting plan.

A.   The DCHVS Sample Design

The District of Columbia Household Victimization Survey (DCHVS) measured crime victimization for residents of the DC-SMSA. Separate estimation capability was desired for the District of Columbia proper and the suburban areas. Initially, the target population of the DCHVS was defined to be civilian, noninstitutionalized residents of the DC-SMSA age 12 and

older. In addition to the DC-SMSA household population, residents of noninstitutional, civilian group quarters are included under this definition. Based upon the 1980 Census of Population, this initial target population definition would include approximately 3.1 million persons of which 2.6 million will be 12 years or older. Since the DCHVS was to be a telephone survey, two problems existed with this target population definition. These problems resulted in a revised definition of the target population.

The first problem related to the fact that telephone exchanges frequently cross county boundaries. Hence, a sample of telephone numbers would reach households that lived on the border of the DC-SMSA and share exchange codes with the DC-SMSA. One solution to this problem would be to consider these individuals as ineligible and screen them out of the sample early in the interview by determining county of residence. Another solution is to redefine the geographical basis of the target population to be the geographical areas served by the DC-SMSA telephone exchanges. The latter solution was chosen since (1) the DC-SMSA victimizations were to be used to make comparisons with the Capitol Hill employees and (2) the victimization experience of individuals who lived across county boundaries but were served by DC-SMSA telephone exchanges should be similar to individuals inside the boundaries. Since the area outside of the DC-SMSA served by DC-SMSA telephone exchanges is minimal, the target population was defined to be the civilian, noninstitutionalized residents age 12 and older of the DC-SMSA and those adjacent areas that are served by DC-SMSA telephone exchanges.

The second telephone survey related problem was that 2.6 percent of the occupied housing units in the DC-SMSA do not have telephone service.*

---

*U.S. Bureau of the Census (1982). Provisional Estimates of Social, Economic and Housing Characteristics (PHC 80-S1-1), Table H1, page 79.

Non-telephone residences could not be surveyed and hence are a source of frame undercoverage. The target population definition could have been revised to exclude these non-telephone residences. However, we chose not to do so. The reasoning behind this decision was that survey data users will tend to ignore the non-telephone exclusion clause and use the data as though they represent the entire DC-SMSA. For this reason, instead of revising the target population definition to exclude non-telephone residences, a post-stratification adjustment was made to the sample weights to reduce the undercoverage bias in survey estimates.

An unclustered random digit dialing approach was used in interviewing DC-SMSA residents. Separate samples of telephone numbers were selected for the District and the Virginia and Maryland suburbs. For the District, the frame of telephone numbers was sorted by exchanges. For the Virginia and Maryland suburbs, the frame was sorted by State, rate center city, and within rate center city by exchange code. This resulted in a frame ordered essentially by geographic area. To obtain sufficiently accurate estimates for the District, oversampling was needed since the District population is less than $\frac{1}{4}$ that of the entire SMSA. The sample design can be briefly described as a stratified random sample where exchange codes form the strata. A total of 5,542 D.C. area residents age 12 and older completed interviews.

B.   The CHEVS Sample Design

The target population of the CHEVS was the employees of the House of Representatives and the Senate and related Congressional offices, excluding the elected members of Congress themselves. The offices and organizations included were the Senate, the House of Representatives, the Library of Congress, the Architect of the Capitol, the Office of Technology Assess-

ment, and the Congressional Budget Office. The target population included all persons who were employed by these Congressional Offices at any time during 1982. This population is estimated to contain approximately 25,000 persons.

The sampling frame for CHEVS was constructed using lists provided by the Capitol Hill agencies for 1982 employees. RTI obtained machine readable files from the Library of Congress and the Office of Technology Assessment that contained the following information for each individual employed by the agency during 1982: (1) name, (2) Social Security Number, (3) work address, (4) work phone number, (5) home address, (6) home phone number, and (7) dates of employment. In addition, characteristics of the employee and his/her agency position were available for the Library of Congress; this information was used to improve the efficiency of the sample design. The Architect of the Capitol provided a machine readable file with name, Social Security Number, and home address for their 1982 employees. The Congressional Budget Office, the House of Representatives, and the Senate were unable to provide machine readable files for their 1982 employees. Instead, they provided printouts of employees as of late 1982 with their name and office. For these three agencies, the frame was subject to under-coverage of an unknown extent. In addition, since Social Security numbers were not available for all employees, the frame may contain multiple entries for the same employee. Employees who were listed on the frame more than once had more than one chance of selection. This event was accounted for by removing the duplicate listings whenever possible. The frame was sorted by agency and person characteristics when available and a stratified random sample selected where the agency groupings defined the strata. A total of 1,889 Capitol Hill employees completed interviews.

C. Construction of Sample Weights

To make inferences about the target population, sample weights were constructed that reflect the sample design. The weight of a sample unit can be viewed as the number of units in the survey population that the unit represents. Since sampling for each survey was without replacement, the initial sample weight was computed as the inverse of the selection probability of the sample unit. In some cases sample units had multiple opportunities for selection into the sample and the frame multiplicity of the sample unit were unknown. For instance, the fact that more than one telephone number can be associated with a sample residence introduces multiplicity for the DCHVS sample. As described in Appendix A, an unsuccessful attempt was made to identify these multiple-telephone-households. If Social Security numbers had been known for all Capitol Hill employees, the CHEVS sample frame could have been constructed so that employees who worked for more than one agency would have only one chance of selection. When Social Security Numbers were available for Capitol Hill employees, multiple listings were removed from the frame. Because of lack of complete data on frame multiplicity, both samples can be expected to contain a few selections that had multiple opportunities for selection. This frame multiplicity could not be removed by sample weighting since the units subject to multiple chances of selection could not be accurately identified.

These initial sampling weights were adjusted to account for nonresponse and undercoverage. Post-stratification adjustments were made to 1980 Census data for the DCHVS and to frame totals for the CHEVS. The final sample weights serve to differentially weight the sample data from individuals to reflect the level of disproportionality in the final sample relative to the population of interest. Both household level and person level weights were constructed for the DCHVS.

10

CHAPTER 3. DATA COLLECTION

Two computer assisted telephone interview (CATI) surveys were conducted as part of the District of Columbia Crime Victimization Study. For both surveys, the District of Columbia Household Victimization Survey (DCHVS) and the Capitol Hill Employees Victimization Survey (CHEVS), the same data collection instrument was used. This instrument was modeled after the experimental version of the National Crime Survey instrument tested in the Peoria Pilot Study. However, the contact and screening procedures employed for each survey varied because of the different sampling procedures used to identify the target populations. Random digit dialing was used to identify eligibles for the DCHVS; the sample for the CHEVS was selected from lists of Capitol Hill employees. Data collection began in mid-May and continued through the end of August. During that time, a team of approximately 27 interviewers working over three shifts conducted interviews for both surveys.

A.   The CATI System

1.   System Description. In computer assisted telephone interviewing (CATI), the survey instrument is stored within the computer, and questions or items are displayed for the interviewers in program-controlled sequences on cathode ray tube (CRT) terminals. The telephone interviewers read questions as they are relayed from the computer to the viewing screen; as the respondent answers, the interviewers record the answer and enter it as data into the computer by depressing keys on the connecting terminal keyboard.

The use of the computer in questionnaire administration offers the capability for collecting high quality data in an efficient manner. Because skip patterns are computer-controlled rather than interviewer-controlled,

11

the incidence of missing or inconsistent data is greatly reduced under CATI. Since interviewers are freed from time lapses caused by turning pages and monitoring skip patterns, the time required for questionnaire administration is reduced. Moreover, by entering responses directly into the computer as the questions are answered, the data entry step is eliminated.

2. CATI Programming. RTI's CATI system, installed onto a VAX 750 minicomputer, requires no special programming language to develop the CATI version of a questionnaire. Instead, once the user indicates the task to be performed (e.g., questionnaire development), the system provides a series of prompts to follow in completing the task. Survey specialists, experienced in both survey adminstration and CATI programming, were responsible for programming the instrument for the study. With the exception of the eligibility screening questions for the two surveys, all victimization screening and incident data were collected using the CATI system.

The programming of the questionnaire involved the development of a set of logically linked screens, which were displayed to the interviewer on a CRT during the course of the interview and which usually contained one or more questions. Each screen was constructed by completing the following activities:

- defining such screen attributes as the screen name, the number of distinct responses that would be entered on the data file, and the normal sequence of screen display,

- entering the text of the questions and any necessary interviewer prompts,

- identifying the variables that are to be used in questions (i.e., names, pronouns, etc.),

- defining the input variable attributes, including the type of data (i.e., alpha, numeric), the variable identification and a short descriptive name, the format of the input and the output, and the acceptable values of the input, and

- defining any special skip logic or consistency checks (this activity does not necessarily apply to the construction of every screen).

Once the CATI program was complete and had been accessed, the interviewer read the questions as they appeared and entered the respondent's answers. Editing procedures were included as a part of the CATI program so that the survey data were edited as they were entered. The computer immediately performed programmed checks for valid codes, consistency, and completeness, and the system required that invalid and inconsistent entries be corrected by the interviewer while the interview was still in progress. The program had control functions that allowed the interviewer to override the program logic and move forward or backward to selected screens in order to make necessary corrections.

3. The Data File and Data Collection Management. As the interview was conducted and the respondent's answers keyed, the CATI system entered the data directly onto a computer-readable data file which included numeric, alpha, and alpha-numeric data. Because CATI created this data file as an on-going operation, the file could be accessed and analyzed during the course of the survey. As part of the CATI program, a current status code was incorporated as an item of data to be entered for each sample case. This status code identified the action taken on each case and the result of that action. These codes identified completed interviews, refusals, no answers, busy signals, etc. Routine tabulations of these codes were made to allow project management to monitor data collection activities and to make necessary procedure or scheduling adjustments.

B. OMB Clearance

A clearance package was prepared and submitted to the Office of Management and Budget through the appropriate clearance process. Copies of the

survey instruments, a project justification statement, a work plan, time schedule, publication plans, an estimation of respondent reporting burden, and other materials necessary for clearance were submitted by BJS for review and approval by OMB. Approval was obtained on May 19 and extended through August 31.

C. CATI Interviewers

1. Interviewer Manual. It was essential that all data collection procedures be specified and adhered to in order to obtain consistent, high-quality data from respondents. Toward this end, project staff prepared a Telephone Interviewer's Manual to serve as both a training manual and an interview procedures guide. This manual included comprehensive coverage of such topics as:

- purpose, sponsorship, and importance of the survey,

- the interviewer's responsibilities,

- confidentiality of data collected,

- CATI operations,

- contacting sample members,

- explaining the study and overcoming respondent objections,

- procedures for conducting interviews and keying responses,

- question-by-question specifications for administering the survey instruments,

- scheduling work, and

- completing project forms and records.

2. Interviewer Training. A training session was conducted by project staff to teach the telephone interviewers and supervisors before interviewing began. During training, the Telephone Interviewer's Manual was thoroughly reviewed with particular emphasis placed on familiarizing the staff with the questionnaire and item-by-item specifications, as well as

with procedures and techniques to be used in contacting sample members. Telephone interviewers were given background information on the survey objectives and other possible concerns of the respondents and were trained in confidentiality and privacy requirements for the study. They learned answers to anticipated respondent questions, such as questions concerning the sponsorship of the survey, its purpose, sample member selection, and authenticity of the survey. Interviewers were also trained to handle problems such as refusals and postponements. The principal instructor, after covering the above topics in lecture fashion, demonstrated how an interview for the study should be conducted. Finally, a major component of the training session was simulated practice interviews in which the interviewers conducted computer assisted telephone interviews following a prepared script designed to give the interviewer experience in dealing with problems that were likely to arise during actual interviewing.

As a supplement to this interviewer training session, a half-day debriefing/retraining session was held approximately two weeks into the data collection period. The purpose of this session was to discuss in a group setting those problems that have been most common during the first week of data collection and to present standardized solutions to them. An additional half-day session was held approximately five weeks into the data collection period to review procedures to minimize survey nonresponse and to convert respondents who were reluctant to participate.

D.   The District of Columbia Household Victimization Survey

1.   Identifying Eligible Housing Units. For each random telephone number selected for the DCHVS, interviewers received a Random Telephone Number Screening Form that included the telephone number, a case identification number, screening questions, and a space for recording and coding

calls. This screening form identified residential units eligible for participation in the DCHVS by questions designed to elicit the following information:

- the telephone number reached and

- the type of place the number served (i.e. residence, business or institution, or pay phone).

If the number was dialed correctly and served a residence (or a business or institution that included resident quarters served by the same number), the interviewer asked to speak with someone 18 years of age or older and began the interview. If no one of this age was available, the interviewer scheduled a time to call back.

The majority of these screening calls resulted in ineligible telephone numbers. Codes were assigned to these numbers and established procedures followed regarding the number of call backs required before considering the number as definitely ineligible. Listed below are the definitions for telephone screening calls and the minimum call-back procedures required before coding them as final:

- nonworking number (recorded intercept)- after 2 calls, code final ineligible,

- temporarily nonworking number (recorded intercept) - after 5 calls, code final ineligible,

- wrong connection (another number reached) - after 2 calls, code final ineligible,

- no result from dial (no connection) - after 5 calls, code final ineligible,

- fast busy signal (accelerated busy signal) - after 5 calls, code final ineligible, and

- ring, no answer (normal ring with no response) - after 8 calls, code final indeterminate.

Numbers verified as serving a business or institution with no resident quarters or serving a public pay phone were also coded as final ineligibles.

In the event a single phone number was found to serve more than one residence, all residence(s) were included in the survey and each residence was assigned a unique housing unit identification code.

2. _Conducting the Interview_. Victimization data were collected for all 12-year and older members of sample residences in the DCHVS. The first interview was conducted with someone 18 years or older. At the conclusion of this interview, a roster was completed and the interviewer requested to speak with other eligibles. For 12- and 13-year-olds, the interview was conducted with a parent as proxy; all other interviews were conducted with the eligibles themselves.

In the event that all interviews for a residence could not be completed during the initial contact, the interviewer identified convenient times to call back. The interviewer was responsible for maintaining a record of such appointments and for making timely call-backs. In instances when an eligible was identified and four call-backs had been made without success, the telephone task supervisor reviewed the recorded information and discussed the case with the interviewer. The supervisor then decided to continue the case or to terminate action on the case. If the decision was to continue, the supervisor advised the interviewer as to the plan of action, which might have involved assigning the case to another interviewer, suggesting alternative times to call, or some other action. The decision to terminate action on a case was only made by the supervisor.

E. The Capitol Hill Employees Victimization Survey

1. _Lead Letters_. A lead letter announcing the study, explaining its importance, requesting participation, and alerting the individual to the RTI telephone contact was sent to each sample member about one week before the CHEVS data collection began. The letter, which was on Congressional stationery and signed by Congressional representatives, stressed that all

17

interview data would be treated in a confidential manner and that participating sample members would remain anonymous.

A return postal card, addressed to RTI, was included with each lead letter. Sample members were requested to complete the card, providing their home and office telephone numbers and indicating a time when they would prefer to be called. Information from returned cards aided in scheduling the telephone interviews efficiently and at the convenience of the sample members. Approximately 25 percent of the sampled employees returned the postal card with the requested information.

2. Conducting the Interview. In addition to sample members' names, the sample listing of Capitol Hill employees selected for participation in the CHEVS contained addresses and work and home phone numbers when such information was available. Sample members who returned postal cards were contacted at the time they indicated as preferable. Initial attempts to contact others were made at their work telephone number if that number was available. The purpose of these calls was to establish when and where the employee wished to be interviewed. Interviews were completed during the initial call if the sample member desired; otherwise, the interviewer called back to complete the interview at the time and place designated by the respondent. If the employee could not be contacted at work, an interviewer called the individual at home during night or weekend hours.

Since a list sample was used to identify the target population for the CHEVS, the telephone screening process was much simpler than that used for the DCHVS. The interviewer determined if the correct number had been reached and the Capitol Hill employee was still available at that number. Tracing was needed to locate sample employees who had moved.

18

F. Telephone Interviewing

Telephone survey efforts were scheduled to obtain optimal results at minimal cost based upon consideration of such factors as volume of work, appropriate contact times, at-home probabilities, shift differential costs, and staffing implications. Interview assignments were made by the telephone supervisors under the direction of the data collection task leader. Supervisors were present during all working hours to observe and monitor interviewing activities, and any problems were reported to the data collection task leader for resolution.

Interviewers were trained to meet objections to participation raised by sample household members. Respondents who continued to express doubts as to the authenticity of the study were provided with the telephone number of a government official who were prepared to provide information about the study and its goals. DCHVS respondents were given a telephone number in the Bureau of Justice Statistics. CHEVS respondents were given a telephone number in the Congressional Research Service. The Bureau of Justice Statistics received approximately 50 calls from D.C. area residents; the Congressional Research Service received approximately 20 calls from Capitol Hill employees.

Interviewers did not unduly pressure any individuals to respond. Each case where a designated respondent was reluctant to be interviewed was set aside by the interviewer and discussed with the supervisor. Depending upon the circumstances, the supervisor might attempt to contact the sample member in an effort to obtain cooperation, direct the interviewer to make another attempt using a different approach, assign the case to a different interviewer, or determine that no further action is reasonable and terminate work on the case.

Performance standards were established for telephone interviewers. Initial interviews completed by each interviewer were monitored and critiqued by a supervisor. Should any problems be identified in an individual's work, retraining was conducted and observations continued until the interviewer's work met the prescribed standards. When quality control measures indicated that standards were being met, the supervisors continued to check the performance of interviewers by monitoring ten percent of each interviewer's calls using "silent" telephone monitoring equipment.

CHAPTER 4. DATA PROCESSING

The Office of Technology Assessment (OTA), the Architect of the Capitol, and the Library of Congress provided RTI with a data tape containing a roster of persons who were employed by their agency in 1982. The roster included the following information (not all of which are available for the Architect of the Capitol): name, home and work addresses, home and office telephone numbers, Social Security number, and person characteristics. Documentation accompanied each tape providing the record layout, a description of each variable, a definition of the values used for each variable including missing values and consistency codes, and the tape specifications. The data files were compared with the documentation to insure that the data were complete and consistent and that the documentation was accurate. Any differences between the data files and the documentation or any discrepancies in the data were resolved as the differences were located.

The next step was to convert the data on each tape to a uniform format. Depending upon the data received, RTI recoded, reformatted, and collapsed variables. The reformatted data from each agency was then merged and this merged file was checked to determine whether there were duplicate names on the file, that is, persons employed by more than one of the agencies listed during 1982. Duplicate records were removed from the file. The resultant file was the sampling frame from which the automated portion of the CHEVS sample was selected.

The House, Senate, and CBO provided RTI with a hard copy listing of persons who were employed by their agency during certain time periods in 1982. Using these listings as a sampling frame, the balance of the CHEVS sample was selected as discussed previously. Using the data provided in

the listings and the record format already established for the machine readable files, a file was constructed for sampled employees of the House, Senate, and CBO. These data were merged with data for sampled employees of OTA, Architect of the Capitol, and Library of Congress and the merged file to constitute the CHEVS sample.

Because the data for the CHEVS sample were from two media, the merged file was scanned to determine whether there are duplicate listings on the file. When duplicate listings were found, one of the records was kept and one deleted. Next, a tape containing a sequence number, name, street address, city, state, zip code, home and office telephone numbers, person characteristics, and sampling information including the sample type (CHEVS) was prepared. The sequence number was the only required item on each record; it was necessary for CATI record access.

After the DCHVS sample telephone numbers were selected, a tape containing the sequence number, telephone number, and sampling information including the sample type (DCHVS) was prepared in accordance with specifications provided by RTI's CATI programming staff. The data recorded on the DCHVS sample tape had a format similar to the CHEVS sample tape; data items that were not available from the DCHVS sample were left blank.

Extensive edits were performed by the CATI computer program at the time of data collection. Therefore, machine edits that were performed after data collection was completed were cursory. After reformatting the data, type of crime recodes were developed. The specifications for the type of crime recode were modeled after that used by the National Crime Survey for coding crime type. RTI then developed software to assign a type of crime recode to each victimization. The victimizations that could not be categorized using the computer software were reviewed and coded manually.

Detailed specifications that were used in developing the type of crime recode and other data recodes are included in Appendix A.

A tape for delivery to BJS was prepared after the sampling weights were computed and added to each respondent's data record. Only data collected as a part of the DCHVS were delivered to the government. Frequencies were run on all discrete data items and means on all continuous data items. The file contains no information that will permit an individual or the agency at which he/she is employed to be identified.

The documentation includes the name of the data item, a description of the data item, frequencies of the possible values including consistency codes and missing values, a description of the values, the position of the data item in the record, and the format of the data item.

The tape specifications include information on the number of files, the record lengths, the block sizes, the recording density, and the number of records on each file. The tapes have IBM standard OS labels and the file names included in the tape specifications.

CHAPTER 5.   GENERAL ANALYSIS APPROACH

Since the inception of the National Crime Survey (NCS), questions have been raised as to the validity of data collected in victimization surveys. Many methodological studies have addressed measurement issues in the past or are presently in progress as a part of the NCS redesign effort. Initially, questions were raised concerning whether respondents would discuss their victimization experiences and how well they could recall victimization episodes. Record check studies using known victims revealed that victimization data could indeed be obtained in a household survey but that the design of the instrument, reference period length, and data collection procedures can seriously affect the quality of the resultant data. For this reason, D.C. crime survey procedures were modeled after those in current use or planned for use by the National Crime Survey. Financial constraints prevented the use of certain procedures such as clustered area sampling and personal interviewing. The questionnaire used by the D.C. crime study is a modified version of the standard NCS instrument; the questionnaire was developed by the Crime Survey Redesign Consortium as a prototype for the future NCS data collection approach. Because of the similarity between the two studies, the analysis plan for the D.C. Crime Victimization Study was modeled after that of the National Crime Survey. The remainder of this chapter outlines the general features of the analysis plan and comments on questions that had to be resolved in order to complete the analysis.

A.   Comparison of the NCS and the D.C. Crime Study

In order to compare the D.C. Crime Victimization Study to the NCS, the characteristics of the National Crime Survey need to be described. The NCS

sample design can be described as a stratified, multistage, cluster sample of approximately 73,000 housing units. The entire sample is divided into seven rotation groups of approximately 10,000 dwelling units each. Each rotation group is in the sample for 3 years with the rotation groups at any point in time differing in their length of stay in the survey. Every six months, a new rotation group is selected and the oldest rotation group leaves. Each rotation group is divided in six panels, each panel assigned to a particular interviewing month within the six month period. This division of the sample into rotation groups and panels yields a design in which each dwelling unit is visited seven times at six month intervals. Each respondent is asked to report victimizations that occurred in the six months previous to the month in which the interview occurs.

Retrospective reporting is subject to errors due to forward telescoping - the reporting of events as happening in a certain time period when they actually occurred during an earlier time period. In the NCS the effect of forward telescoping is minimized by bounding. In every interview after the first, the interviewer is supplied with a control card summary of the previous interview. If an event similar to one described on the control card is reported, the respondent is queried as to whether the event is actually the same one that was reported earlier. The first set of interviews conducted for an incoming rotation group is used strictly for bounding purposes and is not used for computing NCS study estimates.

It is important to note that the D.C. crime study, by necessity, collects unbounded data. Another difference between the D.C. crime study design and the NCS is in the length of the reference period. The reference period for the D.C. study is from January 1, 1982 to the date of interview

with an average length of 18 months. For analysis purposes, only the victimization data for May 1, 1982 to April 30, 1983 were used; the earlier and later data are for pseudo-bounding purposes. To the extent to which forward telescoping occurs, the interviews completed late in the data collection period will tend to under represent the victimization experience. This will result in yet another difference between the NCS and the DC crime study. In contrast, the NCS is based upon a six month recall period and the interviews are bounded after the first interview.

Even more important interview mode treatment differences exist between the D.C. study and the National Crime Survey. The NCS uses personal interviews for first contacts and a mixture of personal and telephone interviews thereafter. Since the D.C. study was all telephone, it may be subject to increased levels of undercoverage bias (due to loss of nontelephone households) and nonresponse bias (due to the higher refusal rates encountered in telephone surveys). Post-stratification adjustments were used in the D.C. study to reduce this bias but the extent to which differential levels of bias exist for the two studies is unknown. Because of these unknown factors, no direct comparisons should be made between D.C. Crime data and NCS data.

In spite of these differences, the two surveys have many similar features as well. Respondents are asked to report incidents of criminal victimization that happened to them and the information collected about the victimizations is very similar in the NCS and D.C. surveys. The presence of injury and weapons and other details about the victim-offender encounter including offender characteristics are gathered; information about property loss and the aftermath of victimization is also collected. Because there are fundamental similarities between the NCS and the D.C. surveys and

because the NCS has a well developed plan for describing NCS findings, the D.C. analysis was designed in a similar manner as the NCS. The classification of crimes was comparable and the choices of variables for analytic emphasis took direction from these same choices in NCS analyses.

NCS classifies crimes into two broad categories--crimes of violence (rape, robbery, and assault) and crimes of theft (personal and household larceny, burglary, and motor vehicle theft). For analyses of the D.C. and Capitol Hill survey data, similar type of crime (TOC) classifications were used as are used in NCS. Because of differences between the surveys, the offense categories could not be exactly the same, but TOC definitions were matched as closely as possible. The D.C. study also gathered data about crime types that are not included in NCS--most notably threats and vandalism. Findings for these victimization types were included in the analyses of the D.C. and Capitol Hill findings.

In summary, a basic goal of the D.C. analysis was to analyze the data and present findings in a way that conformed with the established NCS approach. Design and methodological differences prevented direct comparisons but fundamental similarities provided a basis for discussions of the findings from the two studies. Special features of the D.C. surveys, such as inclusion of a broader range of crimes, were exploited in the D.C. analyses.

B.   The Comparative Approach

A general feature of the analyses of the D.C. study data was the comparison of victimization rates and other victimization aspects for the different population groups. Most of the analyses categorized the data into two groups and compared the results for these groups; D.C. city residents versus D.C. suburban residents was one grouping and DC-SMSA employees

versus Capitol Hill employees was another grouping. Thus, victimization rates or the percentage of victimizations that involve injury to the victim were compared for D.C. residents versus D.C. suburban residents and for Capitol Hill employees versus DC-SMSA employees. The rationale for this approach was that the victimization experiences of individuals are most meaningful in comparison to others who are similarly situated.

It was decided during the design phase of the D.C. victimization study that it would be important to set the victimization experiences of individuals in and around D.C. in a national context. Since it was not possible to include a national sample in the D.C. study, the decision was made to use NCS data to make these national comparisons. Due to the design and methodological differences described above, direct comparisons of D.C. and national NCS data are not valid. For this reason it was decided to use NCS data only in making comparisons between the DC-SMSA, other urban areas, and the nation as a whole.

The details of this analysis are described in the next chapter of this report. To summarize, the approach was similar to that which was used in the analyses of the DC area/Capitol Hill survey data. Victimization rates and other aspects of victimization were compared for: (1) households and individuals in the DC-SMSA, (2) households and individuals in other urban areas of a similar population size, and (3) households and individuals in the nation as a whole. Just as the comparison of victimization findings for the three population groups in the D.C. area survey assist in understanding victimization within the DC-SMSA, the comparison of NCS findings for the DC-SMSA, other urban areas, and the nation made it possible to view the victimization experience of DC-SMSA residents in a larger context.

There are two additional analysis issues that need further discussion. These are the unduplication of reported victimizations and standardization of victimization rates for demographic subgroups. Strategies for unduplication and standardization were developed in conjunction with other data processing procedures. The issues are discussed in the remainder of this section.

C.   Unduplication of Incidents

In reporting the results from victimization surveys, a distinction needs to be made between incidents and victimizations. To illustrate the distinction, consider a hypothetical event where two persons on an evening out are accosted and robbed of their belongings. The event involves one criminal incident but two victimizations. The two victims may or may not be from the same household. Depending upon the analysis in question, this event may contain a potential for duplicate reporting. If victimizations are being described, a separate report from each victim of the incident is desirable. If incidents are being counted, the fact that more than one person can report the event needs to be accounted for, either as a part of the data collection effort or in after-the-fact data processing. Not all duplicate reporting can be identified during data collection. When the victims of an incident reside in different households, it is not feasible to resolve duplicate reports in the data collection stage. The methods that were available for use in this study to account for duplicate reporting will be discussed after noting the procedures used by previous victimization surveys.

The National Crime Survey approach to this problem is to use victimizations as the principal analysis unit rather than incidents in most analyses. The exception is for household crimes such as burglaries, household

29

larcenies, and motor vehicle thefts which are reported as incidents. The NCS approach in data collection is to ask household crime screening questions of a single respondent within the household. Should someone other than the household respondent mention a household crime, the interviewer determines if the event was already described by the household respondent; if not already described, an incident report is completed. This approach reduces the extent of duplicate reporting of household crimes. However, if the household respondent is not knowledgeable about all household crimes occurring during the reference period, some undercounting of household crimes may occur. For personal crimes, victimizations rather than incidents are usually analyzed. Common estimators are the victimization rate per 1,000 persons (e.g. the number of assaults per 1,000 persons) and the percent of the victimizations of a particular type that have a particular characteristic (e.g., percent of assaults where the offender was a stranger). Two types of incident-level victimizations are reported, however: the ratio of incidents to victimizations and the percent distribution of incidents. To convert victimization reports to incidents, the NCS uses questions that determine how many other persons were victimized in the incident that the respondent described.

The National Crime Survey collects data using hard copy methods even when the interview is completed by telephone. Besides the present D.C. study, the only other CATI survey of crime victimization was the Peoria Pilot Study conducted by the Survey Research Center of the University of Michigan as a part of the Crime Survey Redesign effort. The Peoria Pilot Study was a methodological investigation that contrasted the results from a police sample and a random digit dialed sample when the standard National Crime Survey instrument was used versus when a modified version was used

that incorporated a different approach to crime screening. For methodological reasons, no attempt was made to prevent duplicate reporting at the data collection stage, including within person duplicate reporting. The latter type of duplicate reporting of crimes was not common. More common was duplicate reporting among household members. Across person duplication of an incident report was identified by a computer match of the summary crime description, the date and location of the crime, and the type of crime recode. In developing incident estimates, the incidents were weighted based upon the number of reported mentions.

With respect to the treatment of duplicate reporting, the D.C. study used the most feasible of the two approaches outlined above. During interviews for the DCHVS, each resident of a household was asked to report both personal and household victimizations. In households where more than one person was interviewed, it was possible that more than one respondent reported the same crime, particularly burglaries and household larcenies.*
The interviewer was instructed to remove duplicate mentions of crimes by the survey respondent, but no attempt was made during the interview to determine whether duplicate reports were being made across household members. In analyzing the data, victimizations were focused on in describing rates of personal crimes. For household crimes, the crime reports of the first person responding were used. This approach was used since survey resources precluded the manual or computer matching of crime reports of household members.

D.   Adjustment and Standardization

Many of the analyses done for the D.C. study involved comparisons between the population groups of D.C. residents, suburban residents of the DC-SMSA, and Capitol Hill employees. There also are characteristics of

these population groups such as the age, race, and sex distribution that are highly related to the risk of victimization; these differences had to be considered in comparing the population groups with respect to crime victimization. As an example, the Capitol Hill employee population is distributed quite differently with respect to age than the general DC-SMSA population. Crime victimization also differs by age with the young being victimized more often and the old less often than the population as a whole. Inferences made by a simple comparison of Capitol Hill victimization rates to DC-SMSA rates could be misleading because of the differential age distribution between the two groups. In analyzing the effect of observed differences between the victimization rates for population subgroups, the effect of population characteristics that are not directly involved in the comparison must be accounted for or removed to avoid confounding the comparison.

These population characteristics that are extraneous to the comparison of interest but can confound the comparison may be referred to as "extraneous variables." The first step in adjusting for extraneous variables was to identify population characteristics that affect victimization risk. Historical data from the National Crime Survey were used in identifying these characteristics. For the variables that were identified, the next step was to determine if there were differences in the distribution of the extraneous variables between the population groups being compared. Variables that relate to the risk of victimization and are differentially distributed across the population subgroups need to be accounted for in order to avoid confounding these characteristics with risk factors of interest, such as Capitol Hill employment, for instance.

---

*Duplicate reporting was not a problem for the Capitol Hill survey since only the employee is interviewed and not other household members.

One approach to remove the effect of an extraneous variable on survey comparisons is to compute victimization rates within levels of the confounding variable. Thus, the victimization rates for Capitol Hill employees might be compared to the victimization rates for DC-SMSA employees, within age categories related to differential victimization risk. When there are several extraneous variables associated with a comparison, this approach may not be feasible since the sample may be partitioned into a large number of cells with a small sample for many of the cells. A large number of category-specific victimization rates may also result from the multi-way cross of all the confounding variables, making overall comparisons difficult.

In this situation, a reasonably simple standardization approach is available to control for the effect of extraneous variables. This approach uses a post-stratification adjustment in which the distributions within the population subgroups are forced to a "standard" distribution with respect to the extraneous variables. A major advantage of this approach is the relative ease of computation. Standardizing post-stratification adjustments can be applied to the sample weights. Then standardized estimates can be computed directly using these adjusted weights.

This later method was used when the victimization experience of different population groups was compared for significant differences. In these situations, it was important to know whether observed differences could be explained by the characteristics of those in the subgroups. For comparison of D.C. city residents to D.C. suburban residents, each of the two sets of household respondents were standardized to the DC-SMSA age, race, and sex distribution. For employee level comparisons, the DC-SMSA employees were standardized to the CHEVS distribution by age, race, and sex.

E.   Overview of the Analysis Strategy

When a standardization approach is used, the resultant estimates of differences between the population groups are not descriptive of the populations being studied. In many cases, the purpose of an analysis is to describe the victimization characteristics of the subgroup, as they actually exist. In this situation, a standardization approach may be misleading and inappropriate. In many cases, this was true for the analyses planned for the D.C. crime study. The approach that was used in analyzing the data was to perform a thorough descriptive analysis of the data. As described in the next chapter, this descriptive analysis presented estimates for each subpopulations of interest. Then comparative analyses employing standardization methods were implemented. The subjects that were investigated revolve around comparisons of the victimization experience for DC City residents versus DC suburban residents and DC-SMSA employees versus Capitol Hill employees. These results of these analyses were described in the Report to Congress and the District of Columbia Government.

CHAPTER 6. NATIONAL COMPARISONS

The analyses described in the previous chapter involve the description
of the victimization experience of D.C. residents and Capitol Hill employees
and internal comparisons within the DC-SMSA. To put this D.C. victimization
experience into perspective, comparisons were needed of victimization for
the DC-SMSA and the nation. These comparisons were made using recent data
from the National Crime Survey (NCS). Data from the D.C. crime study could
not be used in making these comparisons due to the many methodological
differences between the two studies. Instead, NCS-based estimates for the
DC-SMSA were compared directly with NCS-based estimates for major metropoli-
tan areas and the nation as a whole.

National comparisons using NCS data were possible since the DC-SMSA
contributes several primary sampling units (PSUs) to the NCS. From the
entire DC-SMSA, approximately 1,100 respondents are interviewed every six
months. For annual statistics, this sample size is relatively small,
particularly when data from the incoming rotation group cannot be used. As
a rule of thumb, the Census Bureau requires ten incident reports in a cell
in order to report a statistic for that cell. In 1979, 62 burglary reports
and 69 violent crime reports were obtained for the DC-SMSA; of the violent
crimes, 5 were rapes, 14 were robberies, and 50 were assaults. However, by
aggregating victimization data over the five year period from 1977 to 1981,
sufficient victimizations were obtained to allow comparisons of the DC-SMSA
to the nation and to metropolitan areas. The unbounded first interview
data was not used in making these comparisons.

The Bureau of the Census (BOC) provided tables that served as the
basis for comparing the victimization of DC-SMSA residents with that of

residents of major metropolitan areas and the nation. To ensure timely production, these tables were formulated assuming standard NCS definitions and procedures would be used. Comparisons of victimization rates were based upon the major analysis variables of victim gender, age, and race and for selected victimization event characteristics such as victim injury, use of weapons, offender relationship to victim, and amount of economic loss. In addition to these tabulations, BOC also provided formulas that allowed us to determine sampling errors for these tables.

The national comparison data were discussed in a separate section of the Report to Congress and the District of Columbia Government. It emphasized that the findings had not been derived from the D.C. victimization surveys. Differences in the data collection instrument and interviewing mode that preclude valid comparison of the D.C. Crime Study and NCS results were discussed.

The NCS based comparisons provide a useful basis for making comparative statements about how the quality of life on an important dimension (victimization) compares for D.C. and other parts of the nation. Political leadership prefers and political constituencies expect to consider issues like the risk of victimization in a comparative framework. Because the DCHVS and CHEVS could not be used directly in national comparisons, the use of NCS data to compare the DC-SMSA to the nation served an important public information function.

APPENDIX A

SPECIFICATIONS USED IN IMPLEMENTING SAMPLING, DATA PROCESSING,
AND ANALYSIS TASKS

MEMORANDUM

August 24, 1983
Revised 11/15/83

TO: Brenda Cox

FROM: Jane Bergsten

SUBJECT: Description of the DCHVS and CHEVS Sample Designs

I. The DCHVS Sample

The DCHVS sample is a random digit dialing (RDD) sample of tele-
phone numbers serving the District of Columbia Standard Metropolitan
Statistical Area (DC-SMSA). A sampling frame was constructed using the
April 1983 AT&T computer tape containing all working telephone exchanges
in the nation, as well as the rate-center city and vertical and horizon-
tal coordinates associated with each exchange. Those telephone ex-
changes serving the DC-SMSA were extracted from the tape, using the
rate-center city and the coordinate information to determine the loca-
tion, and thus the survey eligibility, of the exchange. Those telephone
exchanges known to be entirely nonresidential (usually governmental)
were eliminated from the frame. Checking by telephone with the tele-
phone companies involved revealed that no new exchanges had been added
since the tape had been prepared.

Taking into consideration the desired oversampling of DC City
residents, as specified in the DC Crime Victimization Study Design
report, the sampling rate for DC City residents was set at 2 1/3 times
the rate for Virginia or Maryland suburbs. These rates, after allowing
for the fact that a smaller proportion of DC City telephone members are
working residential numbers, yield a DCHVS sample with an expected
distribution of 40 percent DC City cases and 60 percent DC suburb cases,
as specified in the design report.

Table 1 shows the structure of the DCHVS sample design. A simple
random sample sufficient for 5 waves was selected from each exchange,
resulting in the selection of 105 telephone numbers per exchange in DC
City and 45 telephone numbers per exchange in the suburbs. The selec-
tions within each exchange were then randomly partitioned into 5 equal
size subsamples, one for each of 5 waves of interviewing. Data collec-
tion costs would determine the number of waves that would be used.

Waves 1 and 2 were processed in their entirety and cost projections
indicated that Wave 3 could also be implemented in its entirety. Midway

38

Table 1. Structure of the Sample Design for the Random Digit
Dialing Telephone Survey for DCHVS

Location	No. of Exchanges (Each Exchange is a Stratum)	No. of Random Telephone Selections Per Wave Per Exchange	No. of Selected Telephone Numbers Per Wave
DC City	160	21	3,360
DC SMSA - MD Suburbs	162	9	1,458
DC SMSA- ·VA Suburbs	141	9	1,269
		Total	6,087

into Wave 3, however, unexpected costs made it desirable to cut the sample size. This was done by randomly subsampling one fifth of the DCHVS cases for which no final classification of the telephone number had been made. This subsampling involved 272 of the 6,087 Wave 3 cases, of which one fifth or 55 were retained in the sample and 217 were eliminated. This method of subsampling resulted in a valid probability sample but one for which the overall probability of selection is unknown. In order to obtain a sample for which the probability of selection was known, completed Wave 3 interviews would have had to be thrown out. Because of the inherent waste involved, (most of the sample had already been at least partially worked), we chose this approach instead. A later memoranda describes the approach used to construct sample weights. Although an unbiased weighting procedure was possible, an alternative weighting approach was chosen that has a smaller mean square error.

II. The CHEVS Sample

The CHEVS sample was selected from computer files and hard copy lists of Capitol HIll employees.

The target populations for the survey consist of all employees who worked on Capitol Hill or its immediate vacinity at some time during 1982 for any of the following governmental organizations:

Congressional Budget Office (CBO)
House of Representatives (H) ⎫ excluding elected officials
Senate (S) ⎭
Architect of the Capitol (AC)
Library of Congress (LC)
Office of Technology Assessment (OTA)

Some employees of the above organizations did not work on Capitol Hill and were consequently eliminated from the sampling frames where possible (LC), were eliminated after selection but before screening (H), or were eliminated during the telephone screening (principally H and S). The eliminations consisted primarily of people working in the home district office of a Senator or Representative or were Library of Congress employees based at any of the following locations:

Navy Yard Annex
Landover Center Annex
Taylor Street Annex
Pickett Street Annex.

Table 2 shows the structure of the CHEVS sample. Additional information on the sample selection procedures follows.

The basic sampling procedure involved 1) the formation of strata, 2) the selection of a simple random sample of one-fifth of the persons within each stratum, 3) random partitioning of selections within each stratum into five equal subsamples, one for each of the five potential waves of interviewing.

40

Table 2. Structure of the Sample Design for the Telephone Survey for CHEVS

Organization	Sampling Frame	Number on Frame	Number of Strata	Number of Selections Per Stratum Per Wave*	Total Number of Selections Per Wave	
					Selected*	To Be Screened
Congressional Budget Office	Hard copy listing sent March 3, 1983 from CBO	207	1	8	8	8
House of Representatives	Clerk of the House July 1, 1982 - September 30, 1982 Directory as frame; U.S. House of Representatives Spring 1982 Telephone Directory for telephone numbers	13,397	43 1 1	12 10 ~10(9)	~536(535)	~417
Senate	February 16, 1983 computer printout as frame	6,963	33 1	8 ~15(14)	~279(278)	~279
Architect of the Capitol	Computer file	2,498	11 1	8 12	100	100
Library of Congress	Computer file	5,822	28 1	8 ~9(8)	~233(232)	~233
Office of Technology Assessment	Computer file	297	1	~12(11)	~12(11)	~12
Total number of selections					~1168	
Total number of selections for screening (after eliminating non-Capitol-Hill employees)						~1049

*Numbers in parentheses indicate sample size for one or two of the five waves.

41

For the Congressional Budget Office, House of Representatives and Senate, hard copy lists were used as sampling frames. For the House of Representatives, strata were formed using an alphabetized listing of employees. Selections were checked against a House telephone directory listing, and employees located outside of Washington D.C. were eliminated prior to telephone screening. For the Senate, strata were formed using a listing ordered by office. For CBO and Senate employees, no elimination-before-screening was carried out.

Samples for the Architect of the Capitol, Library of Congress, and Office of Technology Assessment were selected from computer files. The computer files used as sampling frames were first cleaned of 1) persons hired in 1983 2) duplicate listings where a name and Social Security Number match was found, and 3) Library of Congress employees based away from Capitol Hill. Within each of the three organizations, Architect of the Capitol, Library of Congress, and Office of Technology Assessment, the records were alphabetized before forming strata. For the Library of Congress, records were first sorted by sex (judged from title, Mr., Mrs., Ms. or Miss) and then were alphabetized within sex groups, prior to forming strata. No elimination-before-screening was carried out.

Waves 1 and 2 were processed in their entirety. After data collection for Wave 3 had started, a random elimination of 90 percent of the Wave 3 cases that had not yet been contacted also had to be made. This was carried out by separating the unworked case screening forms into piles by organization, combining piles, and systemmatically assigning a digit 0 through 9 to the forms. A random number, 6, was picked and all forms bearing this digit were activated. All other forms, bearing digits 0-5 or 7-9, were eliminated from further screening. This resulted in similar problems with respect to defining the probability of selection as that described for the DCHVS.

September 22, 1983
Revised 11/8/83

MEMORANDUM

TO: Wendell Refior

.. FROM: Jane Bergsten
           Brenda Cox

SUBJECT: Computing Sample Weights for the DCHVS and the CHEVS

The assignment of sample weights for DCHVS will be of two sorts:

1.   Individual weights for the DCHVS sample
2.   Household weights for the DCHVS sample

The CHEVS will only have an individual-level weight. This memorandum outlines the weighting procedure for both samples and describes the formation of a stratum identifier for use in analysis.

Household and Individual Weights for the DCHVS Sample

1.   The procedure for calculating weights will include:

   a.   Computation of an initial sample weight for working residential telephone numbers.

   b.   Households within telephone numbers and persons within household selection probabilities are 1.

   c.   No nonresponse adjustments will be used.

   d.   Post-stratification adjustments will be made using 1980 DC-SMSA Census population counts.

2.   The information needed in order to compute the sample weights is, for each interview:

   a.   The CATI ID number - on CATI file

   b.   The CAC ID number - on CAC file and CATI file

   c.   The SRDC ID number - on CAC file and SRDC file

   d.   The household ID number - on CATI file

43

e. Location of household. Recode to classify as PLACE recode

   (1) MD suburb: code 1,2 or 3 for MDLOC
   (2) DC city; code 1 for STATE
   (3) VA suburb: code 1,2,3,4,5,6,7, or 8 for VALOC
   (4) MD outside DC-SMSA: code 4 for MDLOC
   (5) VA outside DC-SMSA: code 9 for VALOC
   (6) Not in DC, MD, or VA: code 4 for STATE.

f. Sex: Get from answer to SEX variable.

g. Race: Get from answers to RACE variable to calculate RACER as:

   (1) Nonblack: code 1, 3, 4, 5 or 6 for RACE
   (2) Black: code 2 for RACE

h. Race of householder. The householder will be defined as the oldest (AGE) person in the household (HUID). Recode as 1 = nonblack and 2 = black.

i. Age: use AGE variable. Recode as:

Age	Recode #1	Recode #2
12-14 15-19	11	11
20-24	21	21
25-29		
30-34	31	31
35-39		
40-44 45-49	41	41
50-54 55-59	51	52
60-64		
65+	61	

Recode #2 will be used only if collapsing is needed.

    j.   1980 Census population counts from <u>General Population Characteristics</u>: key from table 25, "Age by <u>Race, Spanish Origin, and Sex,</u> <u>for</u> Areas and Places: 1980" Washington D.C. - MD.-VA. SMSA.

       Also key from Table 27, "Household Relationship of Persons by Race and Spanish Origin for Areas and Places: 1980" the required information.

3.   Calculation steps for household weights

    a.   Calculate the initial sample weight for working residential telephone numbers as follows:

       (1) <u>Separately</u> for DC City and the DC suburbs, estimate the population total working residential numbers as

$$\hat{N}_{WR} = N \; \hat{P}_{WR}$$

where

        N is the total number of possible residential telephone numbers[*] for the area, and

        $\hat{P}_{WR}$ is the estimated proportion of telephone numbers in the area that are working residential numbers.

       (2) The proportion of working residential numbers within an area will be estimated as

$$\hat{P}_{WR} = [n_{WR}(1) + n_{WR}(2)]/[n_{SC}(1) + n_{SC}(2)]$$

where

        $n_{WR}(i)$ is the total sample numbers in the i-th wave that were identified in screening to be working residential numbers, and

        $n_{SC}(i)$ is the total sample numbers in the i-th wave for which screening was completed.

The sample counts are provided in the memorandum to the record entitled, "Actual Versus Projected Response and Eligibility Rates for the District of Columbia Crime Victimization Study." Screening is defined to be complete when the telephone number can be classified as eligible or ineligible. By definition an eligible telephone number is classified as working residen-

---

[*]Some exchanges known to be entirely business were eliminated from the frame. "Possible residential telephone numbers" are the remaining telephone exchange numbers with all possible four digits added.

tial. An ineligible number can be nonworking, temporarily nonworking, double wrong connection, business or institution, no result from dial, fast busy, or public pay phone.

(3) Using the estimates derived for the area (i.e., DC City or DC suburbs), each identified working residential number from an area will be assigned as its initial sample weight:

$$\hat{N}_{WR} / [\sum_{i=1}^{3} n_{WR}(i)]$$

where $n_{WR}(i)$ is the sample count of screened working residential numbers in Wave i.

b. Sort by PLACE recode: from 2e above into six groups.

c. For PLACE 1, MD suburbs,
   PLACE 2, DC city and
   PLACE 3, VA suburbs,

separately, compute post-stratification ratio adjustment factors as follows:

(1) Sort by race of householder.

(2) If any cell has fewer than 20 interviewed households, combine race groups only as necessary to make each cell at least 20 cases. We will need to look at them at this stage.

(3) We will fix the race post-strata for each of the three places.

(4) For the fixed post-strata, aggregate the 1980 census figures from 2j above, separately for each place. Note that "nonblack" figures are obtained by:
   Total - black = nonblack

(5) For each post-stratum in each of the three places, calculate the ratio of the census number in (4) above to the sum of the sample weights for each interviewed household in the post-stratum. This is the post-stratification adjustment.

(6) Record the post-stratification adjustment factor on your file and print out, for each post-stratum:

   (a) the description of the post-stratum, that is, place and race of householder,
   (b) the post-stratification adjustment factor,
   (c) the Census total population for that post-stratum,

(d) the sum of the sample weights for that post-stratum, and

(e) the number of records (interviewed households) for that post-stratum.

(7) We will review the post-stratification adjustment factors to see if any smoothing is necessary. Factors of 2 and perhaps those between 2 and 3 will be acceptable. Larger factors, in certain circumstances, may also be accepted.

(8) We will carry out any necessary smoothing operations, documenting all decisions made and procedures used.

(9) The final post-stratification adjustment factor will then be added to each record, for places 1,2 and 3. In addition, it should be added to all records in places 4 and 5, as follows:

(a) Link places 1 and 4 as MD suburbs and 3 and 5 as VA suburbs.

(b) For each place 4 record, determine which place 1 post-stratum it fits into and assign that final post-stratification adjustment factor to it.

(c) For each place 5 record, determine which place 3 post-stratum it fits into, and assign that final post-stratification adjustment factor to it.

(d) Every record having a place recode of 1,2,3,4, or 5 should now have both a sample weight and a final post-stratification adjustment factor. All other records will be assigned a post-stratification factor of one.

(10) Compute the _final household weight_ for each record as the product of the sample weight and the final post-stratification adjustment factor. Record this on each record.

(11) Sum the final household weights for each post-stratum for each place, and print this sum together with the Census total and the ratio of the latter to the former for each post-stratum in each place. Theoretically, the sum of weights and the Census totals should be the same and the ratios should be about 1.

4. Calculation steps for person weights:

a. Begin with the post-stratified adjusted household weight. Attach to each person.

b. Sort by PLACE recode: from 2e above into six groups.

c.   For PLACE 1, MD suburbs,
        PLACE 2, DC city and
        PLACE 3, VA suburbs,

separately, compute post-stratification ratio adjustment
factors as follows:

(1) Sort by sex, race recode, and age recode #1.

(2) If any cell has fewer than 20 interviewed cases, combine age
     groups only as necessary to make each cell at least 20 cases
     using age recode #2.

(3) We will fix the age by sex by race post-strata for each of
     the·three places.

(4) For the fixed post-strata, aggregate the 1980 Census figures
     from 2j above, separately for each place. Note that "non-
     black" figures are obtained by:
                    Total - black = nonblack.

(5) For each post-stratum in each of the three places, calculate
     the ratio of the Census count in (4) above to the sum of the
     sample weights for each interviewed person in the post-
     stratum. (Use the post-stratified household weight for each
     sample person responding.) This ratio is the post-stratifi-
     cation adjustment.

(6) Record the post-stratification adjustment factor on your
     file and print out, for each post-stratum:

     (a) the description of the post-stratum, that is, place,
          age, sex and race recodes,
     (b) the post-stratification adjustment factor,
     (c) the Census total population for that post-stratum,
     (d) the sum of the sample weights for that post-stratum
          (Use the post-stratified household weight for each
          sample person responding.)
     (e) the number of records (interviewed persons) for that
          post-stratum.

(7) We will review the post-stratification adjustment factors to
     see if any smoothing is necessary. Factors of 2 and perhaps
     those between 2 and 3 will be acceptable. Larger factors
     may also be accepted.

(8) We will carry out any necessary smoothing operations, docu-
     menting all decisions made and procedures used.

(9) The final person post-stratification adjustment factor will
     then be added to each record, for places 1,2 and 3. In
     addition, it should be added to all records in places 4 and
     5, as follows:

(a) Link places 1 and 4 as MD suburbs and 3 and 5 as VA suburbs.

(b) For each place 4 record, determine which place 1 post-stratum it fits into and assign that final post-stratification adjustment factor to it.

(c) For each place 5 record, determine which place 3 post-stratum it fits into, and assign that final post-stratification adjustment factor to it.

(d) Every record having a place recode of 1,2,3,4, or 5 should now have both a sample weight and a final post-stratification adjustment factor. All other records will be assigned a post-stratification factor of one (i.e., those with PLACE = 6).

(10) Compute the <u>final person weight</u> for each record as the product of the sample weight, the household post-stratification adjustment factor, and the person post-stratification adjustment factor.

(11) Sum the final person weights for each post-stratum for each place, and print this sum together with the Census total, and the ratio of the latter to the former for each post-stratum in each place. Theoretically, the sum of weights and the Census totals should again be the same and the ratios should be about 1.

Employee Weights for the CHEVS Sample

For the CHEVS, an employee level weight is needed. Follow this procedure to calculate the weight. All computations are <u>within</u> agency. (You probably will have to collapse the CBO and OTA together because of their size.) Each eligible responding employee within an agency will be assigned a weight of

$$\hat{N}_E / [n_{ER}(+)]$$

where

$\hat{N}_E$ is the estimated population count of eligible employees in the agency and

$n_{ER}(+)$ is the total number of eligible responding agency employees summed over all three waves of the sample.

The population total eligible employees is estimated as

$$\hat{N}_E = N \, \hat{p}_E$$

where

49

N is the total number of persons on the agency frame, and

$\hat{p}_E$ is the estimated proportion of the frame listings for the agency that are eligible for the study.

For the House of Representatives and Senate, N will be an estimate obtained as the count of the number of selected employees times the selection interval. This will be after we removed obvious non-DC employees. For the House, we selected, eliminated obvious ineligibles, and then phoned to screen. The proportion eligible employees is estimated from Wave 1 and Wave 2 data as

$$\hat{P}_E = \sum_{i=1}^{2} [n_{ER}(i) + n_{EN}(i)] / \sum_{i=1}^{2} [n_{ER}(i) + n_{EN}(i) + n_I(i)]$$

where

$n_{ER}(i)$ is the total number of agency employees in the Wave i sample who are eligible and respond

$n_{EN}(i)$ is the total number of agency employees in the Wave i sample who are eligible and nonresponding (i.e., complete the screening interview so that their eligibility can be established but not the core questionnaire).

$n_I(i)$ is the total number of agency employees in the Wave i sample who are identified as ineligible by screening.

For checking purposes, print out all components of the weights. Also print out a cross tab of agency by response status indicator.

Stratum Identifiers

Both the DCHVS and the CHEVS were selected as stratified random samples. The DCHVS was deeply stratified based upon exchange code. Because of the large number of strata (exchange codes) and the small sample within many of these (several have only one observation), the strata need to be collapsed. Order the exchange codes within each area code and collapse downward when needed so that each stratum has at least ten respondents. The CHEVS strata had somewhat larger sample sizes and therefore shoul not need collapsing although you will to construct a stratum identifier.

/pp

RESEARCH TRIANGLE INSTITUTE
POST OFFICE BOX 12194
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27709

September 28, 1983

MEMORANDUM

TO: Brenda Cox

FROM: Jane Bergsten

SUBJECT: Weight adjustments for multiple telephone numbers at the sample
          dwelling: DC Crime, Project No. 2634.

A dwelling with more than one residential telephone number has a
larger probability of selection in a RDD survey. One typically applies to
the sample weight a weight adjustment factor equal to the inverse of the
number of different telephone numbers linked to the sample dwelling. We
will not make such an adjustment in the DC Crime Survey sample weights, for
reasons detailed below.

For the 1,020 cases for which a control form was completed on Wave I
of DCHVS, the answers to Q2 "Is there a telephone with a different number
in your home/residence on which you could also be reached?" were distri-
buted as follows.

	Frequency	Percent
Yes	151	15
No	836	82
Refused	12	1
Not answered	21	2
Total	1,020	100%

The 15 percent of households with more than one telephone number is many
times the 1 to 2 percent we had expected. The answers to Q3. "How many
different telephone numbers are there for your home/residence?" were distri-
buted as follows

Number of Phone Numbers	Site: DC	MD	VA	DK	TOTAL
1	5	2	2	-	9
2	51	48	14	-	113
3	3	2	-	-	5
4	-	1	1	-	2
5	1	-	-	-	1
Refused	-	-	-	2	2
Not answered	-	-	-	19	19
Total	60	53	17	21	151

51

The results from these hand tallies made from the Wave I control forms suggested that the questions had possibly been answered about extension telephones rather than different telephone numbers.

A check of about 1,500 residential telephone listings was made for each of DC, Maryland suburbs and Virginia suburbs using May 1982, October 1982 and January 1983 directories, respectively.

Muliple phone numbers discovered were

Frequency	Percent	Site
2	$\frac{2}{1500} = 0.1\%$	DC
17	$\frac{17}{1500} = 1\%$	Maryland
11	$\frac{11}{1500} = 1\%$	Virginia

The results of our checking convinced us that the response to Q2 and Q3 on the control from were undoubtedly referring to telephone instruments rather than mulitiple telephone numbers. Any adjustment using these data would, therefore, introduce much more bias than would result from making no adjustment at all. The latter course of action is, therefore, being taken.

/pp

October 25, 1983
Revised 11/14/83

MEMORANDUM

TO: Wendell Refior

FROM: Jane Bergsten
           Brenda Cox

SUBJECT: Standardization for the DC Crime Victimization Study

A. Standardizing DC City and DC Suburbs to DC-SMSA Characteristics for
    the Resident-Level Analyses:

    1. 1980 Census population estimates are available for the DC-SMSA by
        location (DC City, DC Suburbs) by age by sex by race (black,
        nonblack). This will be the basis for determining standardizing
        weights. We will develop two standardized weights, one for DC
        City and one for the DC Suburbs. Fringe areas will be included
        and linked to city versus suburb location by state of residence
        and area code. This is the same approach that we followed in
        developing the unstandardized weight.

    2. Create for each of the two locations separately, age by sex by
        race (black, nonblack) groups. Collapse age groups, if neces-
        sary, to assure at least 20 interviews in a cell. (See the
        September 22 memo for forming and collapsing age groups.)

    3. For each of the two locations separately, compute a (LOCATION)
        resident standardizing adjustment factor for each cell as

        (adjustment factor for cell i) = $[C(i)/C(+)] \div [WS(i)/WS(+)]$

            where C(i) = 1980 Census population count for cell i of the
                         DC-SMSA,

                  C(+) = 1980 Census population count for the total DC-SMSA,

                  WS(i) = sum of the final person weights for all persons
                          in cell i for (LOCATION), and

                  WS(+) = sum of the final person weights over all cells for
                          (LOCATION).

4. Record the (LOCATION) resident standardizing adjustment factor on each record falling into (LOCATION).

5. Compute the resident standardizing weight for (LOCATION) as the product of the final person weight and the (LOCATION) resident standardizing adjustment factor.

6. Give CHEVS records a resident standardizing weight of zero and a resident standardizing adjustment factor of zero.

7. Check: the sum of the resident standardizing weight for each of the two locations should equal the sum of the final person weights for the same location.

8. Check: for each location, the percentage falling into each age x sex x race cell using the resident standardized weights should be identical to the percentage falling into the same cell for the 1980 DC-SMSA Census population counts.

B. Standardizing DC-SMSA employees to characteristics of CHEVS employees for the Employee Level Analyses.

1. All CHEVS interviews will be considered employees. Use the final person weights. Age, sex, and race groups will be defined as in the September 22 memorandum. Collapse across age groups where necessary to insure a minimum of 20 interviews per cell. Form age by sex by race cells for CHEVS employees keeping track of the number of interviews and the sum of the final person weights for each cell.

2. DCHVS interviews will be classified as employees if they were employed at least one month during the survey reference period. (P8a = 1 or code 1,2,3,...,11, or 12 for P8b). Using final person weights, form age by sex by race groups, keeping track of the number of interviews and the sum of the final person weights for each cell. Collapse to keep minimum of 20 interviews in a cell.

3. Collapse CHEVS employee cells or DC-SMSA employee cells further, if necessary, so that the partitioning for each group is based upon identical divisions.

4. Note that we are including DC-SMSA interviews that were fringe cases on location classification.

5. Form an employee standardizing adjustment factor for each cell i as

(adjustment factor for cell i) $= [CH(i)/CH(+)] \div [WS(i)/WS(+)]$

where

$CH(i)$ = sum of the final person weights for cell i of the CHEVS sample,

$CH(+)$ = sum of the final person weights over all cells of the CHEVS sample,

$WS(i)$ = sum of the final person weights for cell i of the DCHVS sample, and

$WS(+)$ = sum of the final person weights over all cells of the DCHVS sample.

6. Put this employee standardized adjustment factor on each DCHVS employee record in the cell.

7. Compute for each DCHVS employee: Employee standardizing weight = (final person weight) * (employee standardizing adjustment factor).

8. Record the employee standardizing weight on each DCHVS employee record.

9. CHEVS employees receive an employee standardizing adjustment factor of one and an employee standardizing weight equal to their final person weight.

10. DCHVS non-employees get an employee standardizing adjustment factor of zero and a employee standardizing weight of zero.

11. Check: for DC-SMSA employees the sum of the final person weights over all DCHVS employees in cell i is equal to the sum of the employee standardizing weight over all DCHVS employees in cell i.

12. Check: the percentage falling into each age by sex by race cell using the employee standardized weight for DCHVS employees should be identical to the percentage falling into these same cells using the final person weight for CHEVS employees.

13. We need to look at distributions of final standardizing weights so we will need a PROC FREQ or PROC MEANS run. We may need to do some smoothing, but this is doubtful.

14. In doing the standardizing:

    a) DCHVS persons living outside of VA, MD or DC city will be included.

To construct these variables, sort the data file by sample type (DCHVS versus CHEVS), by telephone number, and then by household (HUID). A simple hot deck procedure will be used to replace missing values. In order to implement this process you will need "seed" values for the hot deck variables. The seed values will be defined based upon the values expected for the first record in the sorted data file for each sample type. Two imputation classes will be used to separate the two samples and imputation will be independently implemented within the classes.

As an example, the age variable is created for each record as follows. If P7 is between 12 and 90, then AGE = P7 and AGEII = 0 and the value for P7 is used to update the hot deck register for P7, that is HDAGE = P7. If P7 is missing (P7 = 98 or 99), then the value in the hot deck register is imputed for the age or AGE = HDAGE and AGEII = 1. Similar processes are used for race and sex.

For the residence variables, STATE is imputed first in a manner similar to AGE with the associated imputation indicator defined. If STATE = 1 after imputation, then VALOC = 10 and VALOCII = STATEII, MDLOC = 5 and MDLOCII = STATEII. If STATE = 2 after imputation, then SECTOR = 5 and SECTORII = STATEII, CHLOC = 3 and CHLOCII = STATEII, and VALOC = 10 and VALOCII = STATEII. If STATE = 3 after imputation, then SECTOR = 5, CHLOC = 3, and MDLOC = 5, further SECTORII, CHLOCII and MDLOCII are all set equal to STATEII. If STATE = 4 after imputation, then SECTOR = 5, CHLOC = 3, VALOC = 10, MDLOC = 5, and the associated imputation indicators are set equal to STATEII.

If STATE = 1, then SECTOR and CHLOC need to be defined. If P2b = 1,2,3, or 4, then SECTOR = P2b and SECTORII = 0 and the hot deck is updated, e.g. HOTSECT = P2b. If P2b ≠ 1,2,3, or 4, then SECTOR = HOTSECT and SECTORII = 1. The variable CHLOC is defined in a similar manner. Note that HOTSECT can only take on values 1-4 just as HOTCHLOC will only take on values 1 or 2.

If STATE = 2, then MDLOC needs to be defined. If P2d = 1,2,3, or 4, then MDLOC = P2d, MDLOCII = 0, and the hot deck is updated HOTMDLOC = P2d. If P2d ≠ 1,2,3, or 4, then MDLOC = HOTMDLOC and MDLOCII = 1.

If STATE = 3, then VALOC needs to be defined. The procedure is similar to that for Maryland.

bkp

October 4, 1983
Revised 11/4/83

MEMORANDUM

TO: Danny Allen

FROM: Brenda Cox

SUBJECT: Additional Recoding and Editing Needed for the Analysis Files

An examination of the sample data for the District of Columbia Crime Victimization Study indicates that additional editing and recoding is needed to construct the analysis data files. This memorandum outlines the additional work that needs to be done.

Based upon discussions of the number of persons for whom more than six long forms were needed, it has become apparent that we will need to impute for missing long forms. In order to do this, we will need to have two recodes defined. Both recode varibles will be defined for all crimes in the short form only file and the short form/long form file.

The first variable is crime category or CRM_CAT and is defined as follows:

    1 - Robbery or Attempt
    2 - Injury or Attempt
    3 - Threat to Injure
    4 - Burglary or Attempt
    5 - Personal Larceny or Attempt
    6 - Household Larceny or Attempt
    7 - Intentional Damage
    8 - Not a Crime of Interest

CRM_CAT will be a hierarchal variable with code 1 having the most priority and code 8 the least. The levels are defined as follows:

    a.   CRM_CAT = 1. Robbery or Attempt. If $D2n = 1$ and either $D2i = 1$ or $\overline{D2j} = 1$.

    b.   CRM_CAT = 2. Injury or Attempt. If $D2o = 1$ or $D2p = 1$.

    c.   CRM_CAT = 3. Threat to Injure. If $D2n = 1$ and $D2o \neq 1$ and $D2p \neq 1$.

d.   CRM_CAT = 4.   Burglary or Attempt.   If D2e = 1 or D2f = 1 or D2g = 1 or D2h = 1.

e.   CRM_CAT = 5.   Personal Larceny.   If D2i = 1.

f.   CRM_CAT = 6.   Household Larceny.   If D2j = 1.

g.   CRM_CAT = 7.   Intentional Damage.   If D2m = 1.

h.   CRM_CAT = 8.   Not a Crime of Interest.   If D2e ≠ 1, D2f ≠ 1, D2g ≠ 1, D2h ≠ 1, D2i ≠ 1, D2j ≠ 1, D2m ≠ 1, D2n ≠ 1, D2o ≠ 1, and D2p ≠ 1.

Print out all records that are unclassified under the rules. Also print out 15 records for each category of CRM_CAT. Note that no record in the short/long form file should be classified as CRM_CAT = 8, by definition. Print out any records that you encounter of this sort.

The other variable is an Analysis Time Period Indicator or ANALIND that will tell whether or not a crime occurred within the analysis time period. ANALIND will be defined as

1 - Crime Within Analysis Period
2 - Crime Outside Analysis Period
3 - Not a Crime of Interest

The variable levels are defined as follows:

ANALIND = 1   if CRM_CAT ≠ 8 and the crime falls within the analysis time period

ANALIND = 2   if CRM_CAT ≠ 8 and the crime does not fall within the analysis time period

ANALIND = 3   if CRM_CAT = 8.

A crime is defined to fall within the analysis time period if it occurs between May 1, 1982 and April 30, 1983. If any of the following is true, then the event falls within the analysis time period:

a)   D9 = 2 and D10a = 5-12

b)   D9 = 3 and D10a = 1-4

c)   (D9 = 2 or D13a = 2) and D13b = 1 and D13b1 = 5-12

d)   (D9 = 3 or D13a = 3) and D13b = 1 and D13b1 = 1-4

e)   (D9 = 2 or D13a = 2) and D13b = 2 and (D13b1 and D13b2 are not legitimate skip, DK, RE, or other missing codes) and (D13b1 < D13b2) and D13b2 > 4

f) (D9 = 3 or D13a = 3) and D13b = 2 and (D13b1 and D13b2 are not
    legitimate skip, DK, RE, or other missing codes) and
    (D13b1 < D13b2) and D13b1 < 5.

Otherwise, the event falls outside the analysis time period.

Note that the following should be true. All records within the short/
long form file should have ANALIND = 1. Print out all records that don't.
Also print out 50 records from the short form only file and 50 from the
short/long form file for the purpose of verification.

Please let me know of any difficulties that you encounter in imple-
menting these specifications.

bkr

October 7, 1983
Revised 11/4/83

MEMORANDUM

TO: Danny Allen

FROM: Brenda Cox

SUBJECT: Completing Missing Long Forms for Eligible Crimes

The instrument for the District of Columbia Crime Victimization Study included space for 20 victimizations to be listed and classified and dated via the short incident form (Section D of the Core Questionnaire). To avoid burdening the respondent, provisions were made for long incident forms (Sections E-O of the Core Questionnaire) to be completed for no more than six victimizations that fell within the analysis time period. Therefore, there will be some short forms for which a long form should have been filled out but wasn't. The long form data are required in order to include the victimization in the analysis. These victimizations must be included in order to avoid an undercount of the rate of crime victimization. Creating a crime-level weight was considered but rejected since we cannot simultaneously control for type of crime and for all the analysis variables of interest. Instead a hot deck imputation will be implemented to replace the missing long form data. This memorandum provides specifications for that hot deck imputation.

A victimization was eligible to have a long form completed for it when the short form indicated that it was a crime of interest and that it occurred within the analysis time period of May 1, 1982 to April 30, 1983. In terms of my memorandum entitled, "Additional Recoding and Editing Needed for the Analysis Files," a short form is eligible for a long form when CRM_CAT = 1-7 and ANALIND = 1. If CRM_CAT $\neq$ 1-7 or ANALIND $\neq$ 1, then no long form is needed.

Extract from the short form only file all records with CRM_CAT = 1-7 and ANALIND = 1. Add these records to the short/long form file. Separate out all short/long form combinations that have CRM_CAT$\neq$1-7 or ANALIND$\neq$1. Do not include these records in the remaining operations. Class the remaining records by CRM_CAT and sort them by sample type, then by sex, then by race, and then by age. The sample type is CHEVS, D.C. proper, and D.C. suburbs.

Within each class defined by CRM_CAT, a sequential hot deck imputation procedure will be used to replace the missing long form data. A long form imputation indicator (LFORMII) will be created that is "0" for real data and "1" for imputed data. The imputation will be implemented independently within each imputation class defined by CRM_CAT. Initial long form values are determined for each class in the hot deck based upon the data for the first record encountered with a long form completed. As new records are processed, the imputation class to which each record belongs is determined. If the record being processed has long form data, then that individual's long form data replace the responses stored in the relevant class of the hot deck. Thus, new long form responses are supplied for each cell of the hot deck as they appear in the data file. When a record is encountered with missing long form data, the long form data in the same class of the hot deck is imputed for the missing long form data.

When the imputation is completed, the type of crime variable (TOC) will need to be defined for the imputation-revised records.

bkp

October 7, 1983
Revised 11/04/83

MEMORANDUM

TO: The Record

FROM: Brenda Cox

SUBJECT: Type of Crime (TOC) Specifications

Specifications for a type of crime classification were developed and sent to the government in August. The memorandum provides detailed computer specifications for the type of crime variable (TOC) that was created as a result of those specifications. TOC is a hierarchal variable with level 1 having the most priority and level 36 the least priority. As an example, if a crime could be classified as level 1 or level 4 then the lower number had priority; that is, the crime would be classified as TOC = 1. The TOC variable was only created for completed interviews and only for records with an associated long form.

TOC = 1. Rape with Serious Injury. If injury occurred (D2o = 1) and rape indicated (J6b = 1 or J13 = 5) and either an obviously serious injury indicated (J13 =1, 2, 3, 4, or 6) or an injury with hospitalization for more than one night indicated (J16c = 3 or 4).

TOC = 2. Rape with Minor Injury. If injury occurred (D2o = 1) and rape indicated (J6b = 1 or J13 = 5) and a minor injury occurred (J13 = 7 or 8 and J16c $\neq$ 3 or 4).

TOC = 3. Rape with No Other Injury. If injury or attempt (D2o = 1 or D2p = 1) and rape indicated (J6b = 1 or J13 = 5) but no other injury indicated (J13 $\neq$ 1, 2, 3, 4, 6, 7, or 8) and hospitalization for more than one night not indicated (J16c $\neq$ 3 or 4).

TOC = 4. Robbery with Serious Injury. If personal or household belongings taken or an attempt made to take them (D2i = 1 or D2j = 1) and injury occurred (D2o = 1) and either an obviously serious non-rape injury indicated (J13 = 1, 2, 3, 4, or 6) or an injury with hospitalization for more than one night indicated (J16c = 3 or 4).

TOC = 5. Robbery with Minor Injury. If personal or household belongings taken or an attempt made to take them (D2i = 1 or D2j = 1) and injury occurred (D2o = 1 and J4a $\neq$ 3) but the injury was not obviously serious and did not require hospitalization for more than one night [(J13 $\neq$ 1, 2, 3, 4, 5, or 6) and (J16c $\neq$ 3 or 4)].

TOC = 6.   Robbery with No Injury.   If personal or household belongings taken or an attempt to take them (D2i = 1 or D2j = 1) and injury is threatened or attempted but no injury occurs (D2n = 1 and D2o ≠ 1 and J4a ≠ 3).

TOC = 7.   Assault with Serious Injury.   If injury occurred (D2o = 1) and was an obviously serious non-rape injury (J13 = 1, 2, 3, 4, or 6) or required hospitalization for more than one night (J16c = 3 or 4).

TOC = 8.   Assault with a Weapon.   If weapons are involved (J4b = 1, 2, or 4 or J7a = 1 or J7c = 1) and injury or an attempt to injure occurred [(D2o = 1 or D2p = 1) and (J4a ≠ 3)] with no obviously serious injury and no hospitalization for more than one night [(J13 ≠ 1, 2, 3, 4, 5, or 6) and (J16c ≠ 3 or 4)].

TOC = 9.   Sexual Assault (Excluding Rape).   If injury or attempt (D2o = 1 or D2p = 1) and sexual assault occurred (J6a = 1) but rape did not occur (J6b ≠ 1 and J13 ≠ 5).

TOC = 10.   Simple Assault with Injury.   If injury occurred (D2o = 1 and J4a ≠ 3) that was not obviously serious and did not require hospitalization for more than one night [(J13 ≠ 1, 2, 3, 4, 5, or 6) and (J16c ≠ 3 or 4)].

TOC = 11.   Attempted Assault with No Weapon.   If an attempt to injure occurred but no injury (D2o ≠ 1 and D2p = 1 and J4a ≠ 3) and no weapon was involved (J4b ≠ 1, 2, or 4 and J7a ≠ 1 and J7c ≠ 1).

TOC = 12.   Threats to Injure: Face to Face Contact.   If a threat was made to injure but no injury or attempt occurred (D2n = 1 and D2o ≠ 1 and D2p ≠ 1) and the threat was made in person (J1 = 1).

TOC = 13.   Threats to Injure: Other Contact.   If a threat was made to injure but no injury or attempt occurred (D2n = 1 and D2o ≠ 1 and D2p ≠ 1) and the threat was not made in person (J1 ≠ 1).

TOC = 14.   Forcible Entry.   If burglary or attempt (D2e = 1 or D2f = 1 or D2g = 1 or D2h = 1) and the burglar broke in (F1 = 1 and F3 = 1).

TOC = 15.   Unlawful Entry Without Force.   If burglary or attempt (D2e = 1 or D2f = 1 or D2g = 1 or D2h = 1) and the burglar did not break in but did enter (F1 = 1 and F3 ≠ 1).

TOC = 16.   Attempted Forcible Entry.   If burglary or attempt (D2e = 1 or D2f = 1 or D2g = 1 or D2h = 1) and the burglar tried but failed to get in (F1 ≠ 1 or 3).

TOC = 17.   Completed Motor Vehicle Theft.   If theft or attempted theft of household or personal belongings (D2i = 1 or D2j = 1) and a motor vehicle stolen (G2c = 1).
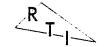
TOC = 18. <u>Attempted Motor Vehicle Theft</u>. If theft or attempted theft of household or personal belongings (D2i = 1 or D2j = 1) and a motor vehicle was not stolen but an attempt was made (G5b = 1 and G2c $\neq$ 1).

TOC = 19. <u>Completed Purse Snatching or Pocket Picking</u>. If theft or attempted theft of personal belongings (D2i = 1) and the victim saw the offender or was in the same place at the same time as the offender (D1a = 1 or D1b = 1) and a purse or wallet stolen (G2c = 4).

TOC = 20. <u>Attempted Purse Snatching or Pocket Picking</u>. If theft or attempted theft of personal belongings (D2i = 1) and the victim saw the offender or was in the same place at the same time as the offender (D1a = 1 or D1b = 1) and an attempt made to steal a purse or wallet (G2c $\neq$ 4 and G5b = 4).

TOC = 21. <u>Other Personal Larcenies With Contact: $50 or more</u>. If personal belongings taken or an attempt to take (D2i = 1) and the victim saw the offender or was in the same place at the same time as the offender (D1a = 1 or D1b = 1) and a purse or wallet was not stolen nor was an attempt made to steal a purse or wallet (G2c $\neq$ 4 and G5b $\neq$ 4) and the total value of the property taken was $50 or more (G3 = 3,4,5,6, or 7).

TOC = 22. <u>Other Personal Larcenies With Contact: Less Than $50</u>. If personal belongings taken or an attempt to take (D2i = 1) and the victim saw the offender or was in the same place at the same time as the offender (D1a = 1 or D1b = 1) and a purse or wallet was not stolen nor was an attempt made to steal a purse or wallet (G2c $\neq$ 4 and G5b $\neq$ 4) and the total value of the property taken was less than $50 (G3 = 1 or 2).

TOC = 23. <u>Other Personal Larcenies With Contact: Amount Not Available</u>. If personal belongings taken or an attempt to take (D2i = 1) and the victim saw the offender or was in the same place at the same time as the offender (D1a = 1 or D1b = 1) and a purse or wallet was not stolen nor an attempt made to steal a purse or wallet (G2c $\neq$ 4 and G5b $\neq$ 4) and the total value of the property taken is not known (G3 $\neq$ 1,2,3,4,5,6, or 7).

TOC = 24. <u>Household Larceny: $50 or More</u>. If household belongings taken or an attempt to take (D2j = 1) and the total value of property taken was $50 or more (G3 = 3,4,5,6, or 7).

TOC = 25. <u>Household Larceny: Less Than $50</u>. If household belongings taken or an attempt to take (D2j = 1) and the total value of property taken was less than $50 (G3 = 1 or 2).

TOC = 26. <u>Household Larceny: Amount Not Available</u>. If household belongings taken or an attempt to take (D2j = 1) and the value of the stolen property is not known (G3 $\neq$ 1, 2, 3, 4, 5, 6, or 7).

TOC = 27. <u>Personal Larceny Without Contact: $50 or more</u>. If personal belongings taken or an attempt to take (D2i = 1) and the victim was not in the same vicinity as the offender (D1a $\neq$ 1 and D1b $\neq$ 1) and the total value of the property taken was $50 or more (G3 = 3,4,5,6, or 7).

TOC = 28. Personal Larceny Without Contact: Less than $50. If personal belongings taken or an attempt to take (D2i = 1) and the victim was not in the same vicinity as the offender (D1a ≠ 1 and D1b ≠ 1) and the total value of the property taken was less than $50 (G3 = 1 or 2).

TOC = 29. Personal Larceny Without Contact: Amount Not Available. If personal belongings taken or an attempt to take (D2i = 1) and the victim was not in the same vicinity as the offender (D1a ≠ 1 and D1b ≠ 1) and the total value of the property taken was not known (G3 ≠ 1,2,3,4,5,6, or 7).

TOC = 30. Vandalism: $50 or More. If intentional damage done (D2m = 1 and H1 ≠ 8) and the damage was $50 or more (H3 = 3,4,5,6, or 7).

TOC = 31. Vandalism: Less Than $50. If intentional damage done (D2m = 1 and H1 ≠ 8) and the damage was less than $50 (H3 = 1 or 2).

TOC = 32. Vandalism: Amount Not Available. If intentional damage done (D2m = 1 and H1 ≠ 8) and the amount of the damage is not known (H3 ≠ 1, 2, 3, 4, 5, 6, or 7).

TOC = 33. Injury or Attempted Injury: Later Unconfirmed. If injury or attempt mentioned (D2o = 1 or D2p = 1) and later denied (J4a = 3).

TOC = 34. Burglary: Later Unconfirmed. If burglary or attempt mentioned (D2e = 1 or D2f = 1 or D2g = 1 or D2h = 1) and later denied (F1 = 3).

TOC = 35. Vandalism: Later Unconfirmed. If intentional damage mentioned (D2m = 1) and later denied (H1 = 8).

TOC = 36. Not A Crime of Interest. If no crime mentioned (D2e ≠ 1, D2f ≠ 1, D2g ≠ 1, D2h ≠ 1, D2i ≠ 1, D2j ≠ 1, D2m ≠ 1, D2n ≠ 1, D2o ≠ 1, and D2p ≠ 1).

After the TOC variable was defined, we checked to verify that a value had been defined for each crime record. Fifteen records from each type were printed out and examined to verify the correctness of the TOC definition.

bkp

November 14, 1983

MEMORANDUM

TO: Wendell Refior

FROM: Brenda Cox

SUBJECT: Type of Crime Recode Needed for Analyzing Crime Data

For use in all analyses of the D.C. Crime Victimization Study data, the following crime recode needs to be created.

RTOC=1.   Robbery.   If TOC=4,5, or 6.

RTOC=2.   Assault.   If TOC=1,2,3,7,8,9,10, or 11.

RTOC=3.   Threat to Injure.   If TOC=12 or 13.

RTOC=4.   Personal Larceny With Contact.   If TOC=19,20,21,22, or 23 or [D2i=1 and (D1a=1 or D1b=1) and (TOC=17 or 18)].

RTOC=5.   Personal Larceny Without Contact.   If TOC=27,28, or 29 or [D2i=1 and D2j≠1 and D1a≠1 and D1b≠1 and (TOC=17 or 18)].

RTOC=6.   Personal Vandalism.   If TOC=30,31, or 32 and D2k=1 and D2ℓ≠1.

RTOC=7.   Burglary.   If TOC=14,15, or 16.

RTOC=8.   Household Larceny.   If TOC=24,25, or 26 or [D2j=1 and (TOC=17 or 18)].

RTOC=9.   Household Vandalism.   If TOC=30, 31, or 32 and D2ℓ=1.

It is important to note that RTOC=4 takes precedent over RTOC=8.

Note the following definitions for use in table generation.

Personal Crimes: RTOC=1-6
Crimes of Violence: RTOC=1-3
Crimes of Theft and Damage: RTOC=4-6
Household Crimes: RTOC=7-9

bkp

RESEARCH TRIANGLE INSTITUTE

POST OFFICE BOX 12194

RESEARCH TRIANGLE PARK, NORTH CAROLINA 27709

(919) 541-6000

COMPUTER APPLICATIONS CENTER

November 11, 1983

TO: Brenda Cox

FROM: Danny Allen

SUBJECT: D. C. Crime - Person 1 Data and Income Coding

The CATI program was designed to request certain information only from the first respondent in the HUID. Questions included were "1a-2f" and "16a-16f" in Section "P." Situations were encountered whereby:

1. more than one respondent was indicated as a first person interview,

2. there were no respondents indicated as first person interviews; however, there were subsequent interviews witin the same HUID,

3. first person interviews were not completed and data was not collected for the given questions; however, subsequent interviews within the same HUID were made, and,

4. first person interviews were not completed but data was collected for the given questions.

Computer listings for all interviews within HUID's that do not have "FIRSTPER=1" are available. Interviewer error for HUID's could have contributed to discrepancies.

Assignment of 1st person data to subsequent persons within the HUID and income coding was implemented based on the following:

1. This applied to the random sample only. The random sample can be determined by "V2" = "2.".

2. The housing unit identifier ("V4") is unique for each household.

3. "V8" is a first person identifier whereby "1" indicates "yes" and "2" indicates "no."

4. Processing was restricted to completed interviews (i.e., result code=80).

5. Applicable data for the first person was inserted into subsequent person records for a given HUID.

6.   If there was more than 1 first person indicated for a HUID, the
     lowest CATI ID with result code '80' was used as the determining
     factor for establishing a first person.

7.   If there were no first persons indicated, the lowest CATI ID with
     result code '80' was used as the determining factor for assigning
     a first person. This usually resulted in missing data for questions
     that were copied and inserted. In this case, missing data was coded
     with missing data codes.

8.   Income recoding and assignment to all records within a given HUID
     was based on the attached flow chart.

9.   The income variable and all copied fields were appended to person
     records as new variables.

10. Recoding was complicated as a result of lost data.

DA/ah

Income Recoding

P16a
≥ 25,000

N=2

Y=1

P16b
≥ 30,000

N=2

Income=5
(25,000 –
29,999)

Y=1

P16c
≥ 50,000

N=2

Income=6
(30,000 –
49,999)

Y=1

A

DK=8
RE=9
Missing

Income=7
(≥ 50,000)

P16d
≥ 5,000

N=2

A

Y=1

Income=1
(0-4,999)

P16e
≥ 10,000

N=2

A

Y=1

Income=2
(5,000 –
9,999)

P16f
≥ 15,000

N=2

A

Y=1

Income=4
(15,000 –
24,999)

Income=3
(10,000 –
14,999)

72

RESEARCH TRIANGLE INSTITUTE

POST OFFICE BOX 12194

RESEARCH TRIANGLE PARK, NORTH CAROLINA 27709

(919) 541-6000

COMPUTER APPLICATIONS CENTER

November 17, 1983

TO: Brenda Cox

FROM: Danny Allen

SUBJECT: DC Crime – Multiple Response Questions

CATI structuring for multiple response questions was defined for a fixed number of entry fields that often did not correspond to the number of possible codes. Codes were keyed and recorded in any order as specific values corresponding to question segments. Unused positions were coded as zeros or blanks depending upon CATI programming and/or interviewer techniques. "Refusal" and "Don't Know" codes were keyed in the first entry position only. Skipped questions (i.e., legitimate skips) were defined with all blank entries.

Software for restructuring was developed based on the criteria defined above. In some cases this involved expanding the number of fields. "Don't Know" or "Refusal" responses were recoded throughout the entire question. The entire question was recoded to blank when the first response was blank. Otherwise the entire question was initialized to zeros and valid responses were assigned specific output positions. Positive responses were then assigned the code of "1."

Various checks were implemented in order to check the validity of recoding. Verification of the procedure included a separate computer comparison and manual review of input data versus the recoded output. The verification process revealed (1) duplicate responses for the same quesion and (2) a limited number of responses that were not recorded as defined in the criteria for recoding.

The recoding process resulted in dropping duplicate responses. An edit/ update process was implemented to correct other responses.

Specific questions affected by the multiple response edit/recode process include the following:

Section	Questions
E	4, 22
F	2
G	2c, 5b
H	1, 2

73

Section	Questions
J	2b, 3, 4b, 9, 11, 13, 14, 16b
K	4b, 5b
O	5, 6b
P	8c, 8f

DA/ah

APPENDIX B

INSTRUCTIONS DEVELOPED TO ANSWER PROCEDURAL QUESTIONS OF INTERVIEWERS

May 27, 1983

MEMORANDUM

TO: D.C. Crime Study TSU Staff

FROM: Dale DeWitt

SUBJECT: Additional Interviewing Instructions

During the early days of interviewing, a number of procedural questions have arisen. Please review the following information and follow the instructions given when applicable.

1. Explanation of source of sample member's name for CHEVS. If asked how we got a sample member's name, state:

    "Your name and work affiliation were obtained from public documents." If appropriate, you may also say: "We did not have access to confidential information."

2. Why we need information about crime events that did not occur on Capitol Hill or in the DC-SMSA:

    "For purposes of analysis, we need to obtain crime event data for the full-time period from January 1, 1982 until today regardless of where the events occurred."

3. Use of "Section C - Examples and Reminders":

    Interviewers are to make all reasonable efforts to read the complete list of examples and reminders. If a respondent raises objections, explain that --

    "There are particular events of interest to the study and I'm reading these examples to help you remember events that may have occurred."

    If a respondent becomes agitated or refuses to continue the interview if the examples are continued, stop reading them and proceed with the interview. Indicate in the notes section of the screening form the approximate point where you stopped reading the list.

4. DCHVS contacts with embassies or other facilities serving foreign governments:

    Citizens of foreign countries who live in an embassy structure or compound and are served by a sample number are ineligible for the survey. The number should be given a final screening Code 14 (Business/Institution).

    If American citizens working for the facility live there and their residential unit is served by the sample number, they are eligible and the number should be treated as a Code 21 (Working residential).

5.   DCHVS <u>number</u> <u>serving</u> <u>a</u> <u>teenager</u> <u>in</u> <u>a</u> <u>household</u> <u>also</u> <u>served</u> <u>by</u> <u>a</u> <u>household</u> <u>telephone</u> <u>number</u>:

The entire household is eligible for the survey and all members should be interviewed. The teenager's phone is to be counted in the number of telephones serving the household in Question 2 of the DCCF.

6.   <u>Roomers</u> <u>served</u> <u>by</u> <u>their</u> <u>own</u> <u>telephones</u>:

When a sample number is a private number for persons living in a room or living unit of a rooming house or dormitory, only the persons served by the sample number are considered members of the residential unit to be interviewed. (If, however, the sample number is a general number serving a number of residents in different rooms or units, they are all to be interviewed, or treated as a group quarters if more than ten are served.)

7.   DCHVS <u>numbers</u> <u>serving</u> <u>government</u> <u>offices</u> <u>or</u> <u>other</u> <u>businesses/institutions</u>:

When an assignment batch is received with all or many sample numbers in the same exchange, the first number called is identified as a government agency office or office within a business or institution, and subsequent numbers appear to be associated (e.g., 252-8000, 252-8001, 252-8002, etc.), time may be saved by obtaining the number for the agency or other organization's central switchboard operator. The remaining numbers may then be considered complete if the operator verifies that they serve business/institutional offices only.

8.   <u>Questions</u> <u>about</u> <u>length</u> <u>of</u> <u>interview</u>:

If a respondent questions you about the time it will take to complete the interview, advise that:

"The average time is about 30 minutes but it does vary from interview to interview."

9.   <u>Referrals</u> <u>to</u> <u>Ms.</u> <u>Taylor</u> <u>or</u> <u>Dr.</u> <u>Langan</u>:

Page II-1 of the project interviewer manual provides instructions for referring questions about the authenticity of the survey to government contacts. <u>These</u> <u>referrals</u> <u>should</u> <u>be</u> <u>made</u> <u>only</u> <u>when</u> <u>your</u> <u>best</u> <u>efforts</u> <u>to</u> <u>explain</u> <u>the</u> <u>survey</u> <u>have</u> <u>been</u> <u>unsuccessful</u>. They are not to be made routinely.

10.   CHEVS <u>postal</u> <u>card</u> <u>name</u> <u>changes</u>:

Some CHEVS postal cards have been returned with the sample member's name crossed through and another person's name written on the card. <u>The</u> <u>originally</u> <u>named</u> <u>person</u> <u>is</u> <u>the</u> <u>sample</u> <u>member</u> <u>who</u> <u>is</u> <u>to</u> <u>be</u> <u>interviewed</u>. We are not to interview substitutes or replacements.

11. CHEVS <u>sample members who did not receive the lead</u> letter:

The CHEVS lead letter may not have been forwarded to sample members who have moved. If, when introducing the study, it appears that the person may not have received advance notice, ask:

"Did you receive the letter from Senator Baker and Representative O'Neill explaining the survey and its importance?"

If the letter was not received, explain that such a letter was sent but apparently was not forwarded to them. Relate the information about the study contained in the letter as necessary to answer the sample member's concerns (Summarize points as needed; <u>DO NOT READ THE ENTIRE LETTER</u>.)

If the sample member's questions cannot be satisfied, advise that we will remail the letter if he/she will give you a current mailing address. Note the information on the screening form, call your supervisor's attention to the need for a mailing, and schedule a call-back ten days later.

DD/sf

June 6, 1983

MEMORANDUM

TO: D.C. Crime Victimization Study TSU Staff

FROM: Brenda Cox

SUBJECT: Additional Interviewing Instructions: Number 2

During the first retraining discussion with TSU staff, several procedural questions were raised. Please review the following information and follow the instructions given when they are applicable.

1.   How to read Section C Examples and Reminders. The Section C examples and reminders should not be read as fast as possible. Timing and tone of voice should be used that create the impression that a check list is being read rather than questions that have to be answered "yes" or "no." The respondent needs to think about each reminder so you should not read them too fast. If you read them too slowly, the respondent may become impatient, however. I suggest that you read the examples at a somewhat faster pace than you read the questions in the later sections of the questionnaire. If you sense that the respondent may need more time to think about an example, use the probe: "Am I going too fast?"

2.   The examples and reminders are too long. This is our problem more than it is the respondent's. As interviewers, you will get to read the list many times. The respondent hears it only once. Be aware of the fact that this section is not as interesting to you, the interviewer, because the respondent usually does not give you verbal feedback (answers) as you read the individual reminders. If you convey the impression to the respondent that the list is boring, the respondent is likely to react in a negative manner. Therefore, I suggest that you train yourself to think positively about the list and your positive reaction will be conveyed to the respondent.

3.   The examples related to "things done by people you know" are confusing the respondents. Several interviewers reported that the respondents were confused by this question and thought we were interested in things done to people they know. To avoid this problem, I suggest that you read the statement clearly and distinctly and accent the word "by."
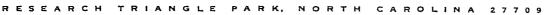
79

4. Explain Question 1 in Section D. The question first asks "Did you see
    the offender?" If R saw the offender, R may be able to provide infor-
    mation to the police about characteristics of the offender. The
    second question asks, "Were you and an offender both at the same place
    at the same time?" If R and the offender were in the same place at
    the same time, then R was potentially in danger. This does not repli-
    cate the information provided in the first question. R could have
    seen the offender stealing his car from a distance and not been in the
    same place or in any danger. Similarly, R could have returned home
    and heard an intruder in the house who fled when the intruder heard R
    arrive. In this case, R did not see the intruder but R was in the
    same place at the same time and was in potential danger. The last
    question asks, "Was there any communication between an offender and
    you?" R may never have seen the offender but he may have received
    threatening phone calls from him. Written communication is not in-
    cluded since we are interested in two-way communication between R and
    the offender.

5. Distinguish between "burglary, illegal entry, and attempted break-in".
    For this study, a "burglary" will be defined to be the act of il-
    legally entering the dwelling place of another to commit a felony or
    theft. An "illegal entry" is entering the dwelling place of another
    without their permission. An "attempted break-in" is the act of
    attempting to illegally enter the dwelling place of another. An event
    involving a stranger entering R's residence or trying to enter without
    his permission would be considered a burglary, illegal entry, or an
    attempted break-in. An event involving a friend of R's child who
    stole something while visiting the child at home would not be con-
    sidered to be a burglary, illegal entry, or attempted break-in since
    the friend was not in the residence illegally. The event does count
    as a theft when answering the questions, "During this event, did
    anyone take or try to take anything belonging to you personally?" and
    "Did they take or try to take property that belonged to your entire
    household, such as furniture or appliances?"

6. Should break-ins involving cars, boats, or offices be included when
    responding to "Was there burglary, illegal entry, or attempted break-
    in?" If only a car or an office is involved, the answer is "no". If
    a boat is involved, the answer is "yes" only if people live on the
    boat (weekend use is included). If the respondent answers "yes" and
    you feel that they are referring to an event that involves a car,
    boat, or an office only, you may probe: "Did this event involve
    illegal entry or attempted entry into a residence where people live or
    have lived in the past?"

7. Question P6 about race is causing a problem. Question P6 may be read
    in this manner, "What is your race? White? Black? American Indian,
    Aleut, or Eskimo? Asian or Pacific Islander?" It is better not to
    read the "Hispanic" or "other" response. The first four categories
    include all races. The "Hispanic" and "other" categories are to be
    used for responses that do not fit into the four categories.

8. Reluctance to answer the Income Items (P16). If R appears reluctant to answer the income items or seems suspicious, you may state: "These questions are to determine the range into which your family income falls and not the specific amount of your income."

9. Questions that don't make sense or seem to have words missing. Since we allowed for 20 sets of short forms and six long forms, many screens had to be copied. In copying them an error could have been made. If you think a question is not phrased correctly, note the screen number and discuss it with your supervisor. If you don't understand a question or why it is being asked, make a point of discussing the question with your supervisor. Questions should be read as written even if they seem repetitive or illogical. Interviewers are not to make judgments about skipping questions or rephrasing questions. The probes may be modified if required but not the question. Bring all questions to your supervisor's attention.

10. Visitors to home have items taken. An interviewer noted that one respondent reported that guests who were visiting him had items stolen. Unless belongings of the respondent or his household are taken, this event is not to be listed. If needed, you may use the probe: "Were belongings of yours or your household stolen or damaged in this event?" If the answer is "no," do not list this event.

11. Treatment of deaf or otherwise mentally or physically incapable respondents. For the CHEVS, complete the screening interview by proxy if possible. Then complete the control card giving "30" for "Screening Completed" as the Screening Result Code and "61" for "Physically/mentally incapable" as the Interview Result Code. For the DCHVS, you may complete the entire interview by proxy under this stipulation: the proxy must have already completed the interview or the proxy is inelgible for interview.

bkp

RESEARCH TRIANGLE INSTITUTE

POST OFFICE BOX 12194

RESEARCH TRIANGLE PARK, NORTH CAROLINA 27709

R
T-I

MEMORANDUM   June 8, 1983

TO: D.C. Crime Victimization Study TSU Staff

·· FROM: Brenda Cox
           Dale DeWitt   ·

SUBJECT: Additional Interviewing Instructions: Number 3   .

     Some additional questions need to be discussed that arose out of the
first retraining discussions. Please review the following information and
follow the instructions when they are applicable.

1.   Some respondents are becoming irritated when we ask Question P8b.
     "For how many months from May 1, 1982 to April 30, 1983 did you have a
     job?" is asked after we determine that the individual was mainly
     looking for work, keeping house, in school, unable to work, or re-
     tired. The individual may have worked at some time during this period
     so we cannot skip the question. To get around this problem, a probe
     may be asked when needed. The probe will be: "Were you employed at
     any time during this period?" If the answer is no, then "0" should be
     entered. If the answer is "yes", the original question should be
     repeated.

2.   Should business crimes be listed. Crimes that involve a business only
     are not included in the survey. However, if personal or household
     property of the respondent is taken or if the respondent is injured or
     attempts or threats are made to injure the respondent, then the crime
     is included. You usually will not know that a crime is business only
     at the listing stage and whether theft or physical danger was in-
     volved. For this reason, the crime should be listed.

3.   How are business crimes handled in answering Section D questions.   In
     answering Question D2a, "Was there burglary, illegal entry, or
     attempted break-in?", a break-in to a store or business is not con-
     sidered to be a burglary or break-in so the answer is "no". This
     question applies to structures for residential use and associated
     property such as garages, yards, or sheds. If the respondent answers
     "yes" and you feel that they are referring to a business break-in
     only, use the probe: "Did this event involve illegal entry or
     attempted entry into residential property?" Question D2i to D2m will
     determine if personal or household property of the respondent was
     taken or damaged in the incident. In answering these questions about
     theft and damage, business property is not included. If R owes a

(919) 541-6000   FROM   RALEIGH,   DURHAM   AND   CHAPEL   HILL

store that is broken into, the property that is taken should be considered to be all business property. The only exception that should be made is when R has a residence attached to his business and this residence was also involved in the crime. The next set of questions determines if R was injured or if attempts or threats were made to injure him in the incident. After these questions are asked, the CATI determines if the crime is of interest to us. If the crime involves a business only, it will not be classified as a burglary or attempt (Q. D2a = No). Also it will not be classified as theft or attempt or intentional damage unless personal or household property of R was involved. Usually business crimes will not be classified as a burglary since there will not be an attached residence, they will not involve theft or damage, and R will not be injured or have attempts or threats made to injure him. Under this circumstance, the crime is not eligible for·the study, and the CATI program will go to the next listed crime.

4. <u>Should the interviewer probe if they feel that household crimes such as burglary are not being reported by all respondents within the household</u>. No probe should be used. However, we do want the respondents to report all crimes that come to mind. If R mentions a crime and then says, "But my wife already told you about that," you are to respond, "Different people can give us a different description of an event. We would like to get a description from you as well." Unless R clearly indicates that he will not provide a description, the event should be listed.

5. <u>Distinguish between household and personal property</u>. This needs to be put in context. In answering Q. D2i and D2k, "During this event, did anyone take or try to take anything that belonged to you personally" or "Was there damage to anything that belonged to you personally?", personal property is that property that can be considered to belong to the respondent as an individual rather than the common property of the household. The household property referred to in Q. D2j and D2ℓ is that property that can be considered to belong to the household as a whole rather than to individuals (e.g., the refrigerator, stove, living room sofa). Roommates living together do not constitute a household for these questions. If one of several roommates has his television stolen, the roommate it belongs to is the only one who should report. For the other roommates, it is not considered their personal property or property that belongs to the household as a whole.

In completing the Stolen Goods Table, two entires are "Other Personal Valuables" and "Household Furnishings." In this case, "Other Personal Valuables," are items that are typically carried on the person. The "personal stereo" referred to in the listing is the Walk-Man variety. The "Household Furnishings" are items that are generally used in the home.

6. <u>Call-backs to follow-up on broken appointments</u>. When an eligible fails to keep an appointment for interview, but has not refused up to

five (5) additional attempts to reach and interview the individual are
required before terminating work on the case. The Final Code to be
assigned if no interview is obtained is 71, since this is, in effect,
an implied refusal.

7. Hard-to-contact CHEVS cases. Unless definitive information is ob-
tained indicating that a CHEVS sample member will be unavailable
during the survey period, efforts to contact individuals who are not
in their office, in meetings, etc. should be continued at reasonable
intervals throughout each data collection wave. Interviewers should,
of course, attempt to learn the best times to call, obtain the sample
member's home phone number for evening/weekend calls, etc. All such
cases in active status at the end of a wave will be reviewed and
decisions made about additional action or assignment of a final code.

8. Answering machines for businesses. If eight (8) calls made at appro-
priate intervals all result in contact with an answering machine that
clearly identifies a business, Final Code 14 is to be assigned.

BC:mc

June 22, 1983

MEMORANDUM

TO: D.C. Crime Study TSU Staff

FROM: Brenda Cox

SUBJECT: Additional Interviewing Instructions: Number 4

We are encountering problems in the Capitol Hill survey with respect to offices that we call frequently. This memorandum discusses this problem and procedural details associated with both surveys. Please review the following information and implement the instructions when they are applicable.

1. Calls to the Doorkeeper's Office. The Doorkeeper's Office of the House of Representatives has been upset by our frequent calls to their office. The Doorkeeper's Office had a number of temporary staff who are now gone (pages) and nonoffice staff (elevator operators) who do not work within the office. We have discussed the situation and worked out the following compromise. Wave 1 individuals will not be traced by calling the Doorkeeper's Office. Those that we have not contacted to date will be traced using Metropolitan Directory Assistance. If no number can be found for them, they are to have the final status code of "Unable to Locate" assigned and the case closed out. For Wave 2 and thereafter, we are to follow these procedures. First, check the latest directory for the House of Representatives. (I have sent one over to the TSU Unit marked "Latest Directory.") If the individual is listed in the latest directory, you may call the indicated number even if it is the Doorkeeper's Office. If you are told that the individual no longer works in the office or otherwise cannot be reached at the number, do not ask for an alternate number at which they may be reached. Instead, thank the individual you are speaking to and close the conversation. Except under the above mentioned circumstances, you are not to call the Doorkeeper's Office. Instead, the Metropolitan Directory Assistance will be used for tracing. The Doorkeeper's Office has agreed to provide location information for up to 10 of the difficult to locate cases. I will request this information for the cases we cannot locate.

2. Calls to the Clerk of the House. I received a call from the Assistant to the Clerk of the House about the disruption caused by our letters and calls to staff of the Clerk's Office. Apparently when they re-

85

ceived the letters and/or got a call, the Clerk's staff verified the
authenticity with him, etc. It was not our calls per se but their
verification calls to him that was the problem as they took a lot of
his time. Together, the Assistant to the Clerk and I figured out a
solution to his problem whereby he would notify them that they would
receive a call and tell them what to expect. The Assistant does not
have any objections to our calling the Clerk's Ofice so we may con-
tinue to do so.

3.   Calls to the Architect of the Capitol. The personnel officer of the
     Architect of the Capitol indicated to me that the bulk of his staff
     were janitors and hence cannot be reached at the Architect's number.
     The Senate Superintendent Office from the Architect's Office has now
     requested that we no longer call his office for this reason. To
     prevent burden on the Architect's Office, we will try to locate these
     employees using the Metropolitan Directory Assistance first. The
     Architect's Office has indicated that they will help us with those
     that we are unable to locate. To prevent burdening them, I will send
     lists for future waves to them after we have made our best attempt to
     locate the employees.

4.   OTA and Library of Congress Employee Tracing. If we have difficulty
     contacting an OTA or Library of Congress employee, let me know. I
     have sources within the agency who have agreed to provide location
     information for those that we are unable to locate.

5.   Frequent Calls to an Office. We are wearing out our welcome with some
     of the Congressional agencies. We will try to reduce this problem in
     Wave 2 by grouping the telephone numbers. However, if you call an
     office and encounter resistance or outright refusal from the recep-
     tionist who answers the telephone, advise your supervisor of the
     problem. The supervisors in turn will discuss the matter with either
     Dale DeWitt or me.

6.   Respondent's Reluctantance To Listen to the Examples and Reminders.
     Betsy Martin, one of the staff who developed the Core Questionnaire,
     provided this example of how the interviewer may explain the reasons
     for going through the list of examples and reminders:

         Survey statistics show that 60% more crimes are remembered
         when examples like these are used. People we interview are
         often surprised at the things that don't come to mind until
         specific reminders are given.

         These examples will also let you know better the kinds of
         events this survey covers.

         Please bear with me while I go through the list.

7.   Overall Comments. Thus far we have been satisfied with the survey
     results with the exception of the response rate for the DCHVS which is
     somewhat low. We are now investigating the problem. You should
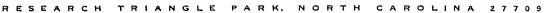
    expect to be advised of steps that you can take to minimize the extent
    of refusals. In the meanwhile, be aware of this problem and carefully
    describe the circumstances that led to refusal and the characteristics
    of the nonrespondent, e.g. the age, sex, and race if discernable.

bkp

APPENDIX C

SUMMARIES OF TASK ACTIVITIES AND PROBLEMS IN ACHIEVING TASK OBJECTIVES

October 19, 1983

MEMORANDUM

TO: The Record

FROM: Brenda Cox

SUBJECT: The District of Columbia Crime Victimization Study: Project
         Summary and Evaluation

The District of Columbia Crime Victimization Study is only the second
application of computer assisted telephone interviewing (CATI) to obtain
crime victimization data.* It is the largest application to date with
approximately 7,500 completed interviews. The study used an experimental
version of the National Crime Survey (NCS) instrument which had been deve-
loped as a prototype for future use in the NCS. This instrument was de-
signed as an improvement on the NCS instrument and the instrument tested in
the Peoria Pilot Study and differed substantially from both instruments.

In the process of implementing the study, we have encountered un-
expected problems, particularly with CATI and the new instrument. As
problems have been encountered that resulted in increased costs, correc-
tions have been made in study plans to avoid cost overruns. However,
several tasks have recently encountered problems that cannot be totally
resolved within the budget. This memorandum reviews all of the unantici-
pated problems and the measures that were instituted to solve these prob-
lems. Since this memorandum reflects my observations as project director,
if focuses on time and money considerations.

The District of Columbia Crime Victimization Study has two phases.
Phase I involved the design of sampling, data collection, data processing,
and data analysis procedures for the study. These procedures were to be
implemented in Phase II of the study. After Phase I was essentially com-
plete, two activities had to be added to the contract specifications for
Phase I in order to satisfactorily complete Phase I of the contract.

---

*The first application was the Peoria Pilot Study conducted by the Crime
Redesign Consortium, which interviewed approximately 2,000 Peoria resi-
dents, approximately 1/3 of which were identified via randomly selected
telephone numbers and 2/3 from police records.

The major added activity was revising the design described in the Phase I Draft Final Report to allow for a redefinition of the objectives of the survey. This change was needed as the result of a decision made by the Congressional advisory panel that the study must compare the victimization experience of District of Columbia residents to that of the nation as a whole and other comparable metropolitan areas.

The second change in the scope of work was relatively minor and associated with instrument development for the study. Originally, a modified version of the present NCS instrument was to be used in the study. This instrument had been used in Peoria Pilot Study CATI application. In December, the decision was made to use the "uniform" instrument being developed for future NCS use since this instrument was expected to be more productive in the sense of stimulating victimization recall. Since the uniform instrument had not been programmed for CATI, RTI had to provide advice to BSSR, which was developing the instrument under another OJARS contract, as to (1) the suitability of the questionnaire for CATI implementation, (2) the factors that would adversely affect interview response time, and (3) the sampling, data processing, and analysis implications of the instrumentation approach.

These changes in the Phase I scope of work added to the costs for Phase I reducing the funds available for Phase II implementation. In addition, these additional activities delayed the start of Phase II. Since victimization data were to be collected for the time period from January 1, 1982 to the interview date, this implied that vicimitzation data would be collected for 17½ months rather than 15½ months, which would increase the costs per completed interview in a proportional manner.

The cost implications of the additional work and the time required to complete the work was recognized in revising the draft report to produce the Phase I Final Report. New projections of the cost per completed interview were prepared for the two surveys and the sample sizes for the surveys reduced so that Phase II projcted costs would be within the targeted amount.

In actually implementing Phase II, unanticipated problems were encountered, most of which were due to the fact that there was little prior information as to situations that could be expected to arise from the use of CATI methods or the use of the "uniform" questionnaire. To the extent possible, modifications were made in project activities to adjust for these problems and the increased costs that resulted.

As a part of Phase I, the "uniform" NCS questionnaire was reviewed and revisions proposed in the instrument. Since extensive changes had to be made in the draft instrument as a result of this review, a second full scale review of the revised instrument was required to verify its accuracy and completeness. The questionnaire was examined by instrument specialists for format, accuracy, and ease of administration by CATI after it was received in early April. The revised instrument was sent to BJS, CRS, and BSSR on April 18th for comments. Comments from BJS and CRS were received by April 22 and BSSR comments on April 29. As these comments were received, the instrument was revised. In making these changes, we again had to review the entire instrument for accuracy, with particular attention paid to the accuracy of skip patterns and the logical flow of the ques-

tions. The final version of the instrument was not completed until the week of May 16. Errors were still being detected and resolved up to the time data collection began on May 22.

In terms of calendar time and person time, the review and revision of the instrument took four times longer than anticipated. This increased time as the result of the newness of the "uniform" approach to data collection; this approach promised to be more productive with respect to victim recall but had not been field tested.

Data collection had been scheduled to begin on May 4. Because of the delay in finalizing the instrument, data collection did not actually begin until May 22. In order to begin on this date, we had to start programming the CATI version of the instrument before the instrument had been finalized. When the BSSR instrument was received in early April, instrument specialists reviewed and revised the instrument a section at a time. As the sections were revised, they were given to the CATI programmers to begin programming. After all sections had been revised, the instrument was reviewed as a unit. This review identified modifications that had to be made in the sections already given to the CATI programmers. A revised version of the entire instrument was given to the CATI programmers in the third week of April. Since the CATI programmers were well into programming the instrument, these changes resulted in additional programming effort. Later changes requested by the government requied additional changes in the CATI program.

It should be noted that we had no choice but to begin CATI programming prior to finalizing the instrument. If we had waited till the instrument had been finalized, data collection would have been delayed by almost two months. This time delay would have made it impossible to deliver the Report to Congress on schedule.

However, the successive changes to the CATI program built in a potential for programming errors. Since CATI data collection is all by computer with no hard copy records, programming errors can result in serious data losses. To prevent such errors, the CATI program was subjected to an extensive review and correction process extending over a two week time period. The debugging process was complicated by the large number of computer screens involved (1,136 screens in all) and the large number of variables in the CATI data base (2,895 variables in the data record).

Hence, the extensive revisions of the instrument had implications beyond the increased personnel time required for instrument specialists to make the corrections. Because the time schedule for report delivery was fixed, CATI programming could not wait till the questionnaire was approved. The changes made to the instrument in turn resulted in additional time required for revising and debugging the CATI program.

Frame development and sample selection began in April and was completed in early May. Unlike the instrument revision and CATI programming task, there was an adequate amount of time in which to draw the sample, print labels, and otherwise have the sample ready for data collection on May 22. However, the CHEVS sample selection was more complicated and time consuming than we had projected. The difficulty centered around sampling

from the hard copy lists that the Senate and the House of Representatives had provided. Instead of employee records, the Senate and House had provided a list of disbursements. Additional time was required to construct the sample since multiple documents had to be searched to obtain address and telephone numbers for each sample listing from the disbursements. This information then had to be transcribed onto coding sheets, keyed, and verified in order to produce a data file for use in generating mailing labels and in setting up the CATI data files. These efforts required increased clerical time above that needed for the simple procedures assumed in costing Phase II.

As data collection progressed in June, problems developed that were the result of frame inaccuracies. From each agency, we had requested the most current home and office addresses and telephone numbers. Only the Library of Congress and the Office of Technology Assessment were able to supply this information. The Architect of the Capitol could only provide home address and no telephone numbers at all. The House and the Senate provided the address and telephone number of the office to which the employee was assigned at the time that the payment records were compiled, which meant the information was about a year out of date.

To obtain telephone numbers and encourage response, a lead letter was sent out to each sample employee prior to interviewing with a post card attached for the employee to complete with the telepone number and time where he/she could be reached. In most cases, only the work address was available for sample employees so the letter was sent there. Only ten percent of the sample employees returned the postcards. In costing Phase II, we had assumed that 50 percent of the employees would return the postcards and provide telephone numbers.

Because of this inaccurate and unavailable information, tracing and locating were needed for about three times more employees than we had projected. This additional effort substantially increased the interviewer time spent to complete each sample case and the associated telephone charges. Data collection costs per completed CHEVS interview were 28 percent higher than we had projected.

In late June, we became aware that we were encountering unusual levels of nonresponse for the DCHVS. For the Wave 1 sample at that time, 28 percent of the working residential numbers had been finalized as nonrespondents and a potential existed for as much as 40 percent nonresponse depending upon how the pending cases were resolved. The reasons for the unusual level of resistance to the survey were unclear. We hypothesized that the residents of D.C. were a more difficult population to interview to begin with and that there might be instrument or interview design problems that were exacerbating the situation.*

---

*The results of the Peoria Pilot Study indicate that the instrument can have an important influence on response. In the random digit dialed component of that study, a household-level response rate of 85 percent was obtained for the National Crime Survey instrument as compared to 80 percent for the experimental instrument. At the person-level, a response rate of 80 percent was achieved for the NCS instrument as compared to 70 percent for the experimental instrument.

To deal with the problem of nonresponse, the decision was made to focus the second interview retraining on nonresponse conversion. (The first retraining had centered on instrumentation problems and the use of CATI.) Training in nonresponse conversion occurred in early July. Wave 1 nonrespondents were then recontacted and many of these were converted.

The training in how to deal with nonresponse paid off in substantially increased response rates to the survey. At the conclusion of the survey, completed interviews had been obtained with 82 percent of the identified working residential numbers with 83 percent of the identified eligible persons within these responding households completing an interview. However, much more interviewer effort had to be spent in obtaining cooperation than we had projected. This additional effort increased the cost of a completed interview. It also made it unlikely that we could finish data collection on schedule. To insure that data collection was completed on schedule, additional interviewers had to be hired and trained. This resulted in additional costs for project staff to train them as well as the additional interviewer training costs.

At the time that the Phase II costs were prepared, it was recognized that CATI interviewing was new enough, particularly with the use of "uniform" instrument, that completely accurate predictions of data collection costs were not possible. For this reason, data collection was set up in waves so that the early results could be used to project survey costs. In mid-July, we assessed the status of survey costs and projected that we would be able to include 18,261 telephone numbers in the DCHVS and 3,147 sample employees in the CHEVS. At that time, charges were only complete through the end of May. These sample cases were released and telephone surveying began.

In early August, complete data collection charges through the end of June were available. In reexamining the data collection costs, it was estimated that unless the sample was cut, data collection costs (Tasks 4-6) would overrun by a substantial amount. In consultation with BJS, the decision was made to subsample unworked Wave III cases at a 20 percent rate for the DCHVS and at a 10 percent rate for the CHEVS. Only unworked cases that were subsampled had data collected for them.

Even with this reduction, the data collection tasks were projected to exceed the amount budgeted for these tasks by approximately $5,000. In addition to the factors described earlier, there was one additional problem that led to increased data collection costs. For both surveys, the yield of completed interviews per sample case was much lower than we had projected. Based upon previous RTI surveys in the D.C. area, we estimated that 28 percent of the telephone numbers would be working residential numbers Instead, we found that only 21 percent were working residential numbers. (This lower yield apparently resulted from the fact that we oversampled D.C. city numbers in order to insure separate estimation capability for the city.) In order to obtain the required number of households, we had to dial many more telephone numbers than anticipated. Even after the Wave III cut back, 13 percent more sample numbers were surveyed than we had projected in the Phase I report. A related event occurred for the CHEVS as well. The hard copy lists used in sample selection were not accurate, including both non-Capitol Hill employees as well as location

information that was out of date. Hence the yield of locatable, eligible employees per sample listing was much lower than we anticipated.

Processing the CATI data began in July by using RTI general purpose software to develop a machine readable codebook and supporting documentation directly from the CATI program. Actual processing of the CATI data began in late August with test programs ran on the Wave I data set. As a result of these operations, we discovered that the data file produced by CATI was not as clean as we had assumed in costing Phase II.

An assumption made in costing the study was that CATI would produce a file that was essentially ready for production applications. This was not the case. Situations contributing to this included CATI software restrictions, variation in programming techniques between programmers, and the instrument changes described previously. In addition, the interviewers induced errors into the data set when they failed to follow program instructions. As an example, identification numbers were erased from a few records when the interviewer backed up over them contrary to instructions. Thus, various post-CATI processing steps have had to be implemented in order to create a data file that could be used for analysis.

Additional data processing was also needed to replace missing data. When we prepared the Phase I design, we assumed that only in a very few cases would an individual have been victimized more than six times during the analysis time period. Hence the CATI program, for space saving reasons, only allowed six long forms to be completed (Section E-O of the Core Questionnaire). The assumption was made that so few victimizations would be missed with this restriction that the lost reports could be ignored. (BSSR had allowed for only four long forms in designing the instrument.) This was not the case. For this reason, we have had to develop an imputation procedure to replace the missing long form data. In addition, we have also had to develop procedures to replace missing age, race, sex, and residence data so that these variables can be used in sample weighting.

All of the above activities went far beyond the limited personnel and computer time that had been allocated to produce analysis files from what we thought would be a clean CATI data base. Some of the problems that we encountered might have been avoided if more time had been available to develop the CATI program and to pilot test it. Other problems are typical of conventional data entry situations and suggest that CATI data, although cleaner than other forms of survey data, still require editing in order to produce a data set of the quality that is needed for analysis.

bkp

94

August 22, 1983

MEMORANDUM

TO: The Record

FROM: Dale DeWitt

SUBJECT: D.C. Crime Study Data Collection Observations

1. Instrument Development Activities

A factor that had major impact on the preparations for data collection and early data collection activities was the amount of unanticipated developmental work required to prepare the instrument for use. This work impacted on the data collection budget and infringed upon a preparations schedule that was already too limited. The time required to prepare the instrument created difficulties for CATI programming and preparations for interviewer training. Also, some problems remaining in the instrument at start-up required additional CATI programmer time and caused problems for interviewers in the early stages of data collection.

2. Complexity of the Instrument

The instrument, as designed, was an extremely complex interview schedule for CATI programming. It required considerably more programming time than had been anticipated and also required more computer capacity than was originally expected. The programming time requirement had a major impact upon the data collection budget, which eventually (combined with some other cost factors), required reduction of sample size. The computer space requirements also had significant effects. To minimize the load on the computers and to prevent jeopardizing other activities to which the computers were committed during the D.C. Crime Study data collection period, certain activities (e.g., telephone number screening, CHEVS sample member screening, DCHVS household rostering, etc.) were done manually rather than on CATI. This resulted in additional work for the Telephone Survey Unit staff, difficulties in maintenance of progress reports, etc.

3. Constraints on Data Collection Preparation Activities

The schedule provided minimal time for the activities required to prepare for data collection. Given the schedule constraints and the effects of the problems already discussed, there was insufficient time to develop data collection procedures and to refine the Interviewer Manual, training plan, etc.

While we believe an adequate job was done under the circumstances, additional time would have allowed for refinements that would have enhanced the efficiency of the data collection operations, provided for improved management control, and reduced the nonresponse problems encountered.

4.   The DCHVS Screening Form

The screening form used for the DCHVS could have been improved in ways that might have enhanced the response rate. The initial activities required to screen the telephone number did not require the amount of explanation and reference to the U.S. Code, for example. The explanation of the study should have been placed after identification of an eligible and should have been worded in such a manner that the respondent could readily sense the potential importance of the outcome to his/her safety and lifestyle.

5.   Length of Interview

For respondents who had crime events to report, the interview was quite lengthy. While the interviewing staff was able to minimize breakoffs, they did occur. A relatively large number of complaints about the length of the interview were reported, and some nonresponse in multi-eligible households resulted because other members were aware of the time it had taken for the initial respondent to complete the interview.

6.   Examples and Reminders

The long list of examples and reminders caused some difficulty. Particularly in the early stages of interviewing, the interviewers were uncomfortable with this section because they perceived that it could be annoying to respondents and feared that they might breakoff. With experience, the interviewers generally overcame this problem, but some respondent complaints about this section were reported throughout the data collection period.

7.   Response Problems

For CHEVS, the major response problems resulted from certain agencies that were either reluctant to have their staff participate or who could not provide the time to aid in locating and contacting sample members for whom telephone numbers and addresses were not made available to RTI. Another factor that contributed to nonresponse was the inclusion of interns and other temporary employees in the sampling frame. These people required more tracing and locating than anticipated and a number of them could not be located. It should also be noted that the decision was made with government project staff that refusal conversion activities would not be undertaken with CHEVS sample members.

The DCHVS presented all of the response problems inherent in random-digit-dial telephone surveys as well as some that were related to the nature of the study (e.g., length of interview, need to interview all residents

served by the sample number who were 12 years of age or older, reluctance of some respondents to answer questions about crime, etc.). To counter such problems, selected interviewers were specially trained (not at project expense) to deal with DCHVS refusals, and the other interviewers were also given additional instruction. While multi-eligible households generally appear to have been less of a response problem than anticipated, difficulties were encountered when an adult (parent or guardian) refused for younger members of a household. Also, individuals who refused to complete the initial telephone screening usually continued to refuse when recalled. Another nonresponse category of concern included those who were away for the summer, which appears to have occurred most often with younger members of multi-eligible households.

8. Telephone Strike

In the final weeks of the study, the nationwide telephone strike caused concern and inefficiency. For example, one entire day was lost because of sabotage of a major carrier line. Sporadic interruptions of service, up to two hours in length, occurred throughout the strike period.

DDeW/lsm

September 30, 1983

MEMORANDUM

TO: The Record

FROM: Brenda Cox

SUBJECT: Nonresponse Types and Conversion Approaches for the District of
          Columbia Crime Victimization Study

   To train project staff in nonresponse conversion, the Telephone Survey
Unit brought in Ms. Dorothy Grossman, the RTI field supervsor in St. Louis.
Ms. Grossman spent several days here monitoring our progress, converting
nonrespondents, and training staff in methods that she had found most
successful in the past. After she completed her stay here, I discussed
with her the aspects of instrument, survey design, and interview design
that she felt affected response for the D.C. study and the procedures that
she recommended for nonresponse conversion. This memorandum summarizes her
observations and comments.

   The first type of nonresponse that we encountered was nonresponse at
screening. The screener determined whether or not a telephone number was a
working residential number and hence eligible for inclusion in the study.

   To prevent this type of nonresponse, Ms. Grossman recommended that the
introduction be read in a slow, deliberate, sincere manner. The person
answering the phone naturally anticipates that the call will be from some-
one with whom he/she is familiar. That person needs time to assimilate who
is calling and why they are calling. If the introduction is rushed, then
the person may become suspicious or may attach little importance to coop-
erating.

   The wording of the introduction may have lead to screening refusals,
too. Ms. Grossman suggested a slightly longer introduction (a short para-
graph) that would provide a nontechnical description of the survey and
hence establish our credibility and allay suspicions. Also, she noted that
the first screening questions could be rephrased to make them less sensi-
tive. Finally, there may have been a tendency for the interviewers to be
over polite and too willing to accept a putoff. For instance, some inter-
viewers were adding the phrase, "Would you have time to help us out?" to
their prepared script. Interviewers need to be assertive in their efforts
to get an interview once they find someone at home.

With respect to converting screening nonrespondents, Ms. Grossman suggested that these cases can be the easiest to convert. By calling at different times, you may get another household member who will respond. In other cases, the original respondent may have been in a hurry or have not understood the introduction. Some people are seldom at home. When you get them, complete the interview. Ms. Grossman related a case where a number was dialed a large number of times with no result. When she reached him, he was just getting ready to leave. She explained how many times we had tried to reach him and said, "Now that I've finally got up with you, won't you finish the interview. I may not reach you again." The man laughed and explained that he had two jobs and didn't stay home when he was not working. He completed the interview.

The next type of nonresponse was individuals who refused to complete an interview after they or someone else within their family had provided screening information. Ms. Grossman indicated that after the screening was completed, the interviewer had a difficult time period to bridge in which they had to key in a number of data items before they could bring up the CATI program. (The screening was done from hard copy.) Many of the interviewers adlibed to fill this time with remarks such as "I am going to ask you a series of questions. If there are any that you would rather not answer, please let me know and I'll go on to the next question." Ms. Grossman suggested that only as a last resort should interviewers or converters tell respondents that they can refuse to answer any questions they would rather not answer. This approach causes the respondent to immediately become suspicous and to be apprehensive about the nature of the interview. This introduces unnecessary probems and can result in the loss of an interview or at least the loss of valuable information. The pause before the CATI program was ready could better be filled by factual statements such as, "We are conducting the interview using a computer terminal so that it will take less of your time. Let me set it up. This will take just a few seconds. I am now entering some data and then we will be ready to go." For future studies, the time delay should be eliminated altogether, in Ms. Grossman's opinion, because of its deleterious effect on response and the difficulties that it presented for the interviewer.

The other reasons for interview nonresponse after screening completion were unrelated to CATI use and instead reflected the respondent's characteristics and attitude to being interviewed.

Some respondents tend to be suspicious, particularly of strangers calling them on the telephone. Once they hear the questions, they will understand the survey is for real. For these cases, the interviewer should say briskly and with confidence "Let's do the interview now" or "Let me start and you can see what the questions are like." or "Let's just start."

Other individuals are simply busy with little time to spare. For these busy people, the interviewer should say, "Depending upon your responses, this interview may not last longer than 15 minutes" and then start the interview.

Another nonresponse type is those who feel the survey is not relevant to them, e.g., those who say no crimes occurred to them. Ms. Grossman's

suggestion was to say, "I'm so glad." and then "Here's the first question."
In other words, get them started and they will generally finish the interview.

An even more common form of nonresponse is the "put offs" who say
"Call me back next week" or "I haven't got time to talk now." Avoid rescheduling the interview since appointments can easily be broken. Again
once the interview is started, it will usually be finished. If rescheduling is unavoidable, the interviewer should set the time and let the respondent know that it is a firm appointment. Phrases may be used such as,
"I will set up an appointment for you at 7:00 or 8:00. Which would you
prefer? Good, I have put you down on my calendar for 8:00."

As long as they are handled right, argumentative types will almost
always give an interview. These individuals are usually young men who
actually want to be interviewed but also want to give the interviewer a
hard time first. These people like to argue and make remarks such as, "I
read enough about this in the paper." or "You should visit the police
stations if you want to know about crime." Ms. Grossman's suggestion was
to bear with them. Don't argue or try to set up an alternate appointment.
After they give you their opinions, then they will answer the questions.
As long as they keep talking, the interviewer can get an interview.

Individuals who have been vicitimized will want to participate in the
study once they understand what the study is about and the subjects that we
are interested in. This implies that the interviewer must give the respondent a chance to learn about the survey and to want to particpate. The
interviewer should stress the importance of the survey by words and manner.

Some nonrespondents just cannot be interviewed by telephone. These
include those with language barriers, hearing problems, the elderly, and
the physically/mentally incapable. Unless we allow proxy interviews, the
individuals are automatically respondents. It would have helped if the
D.C. study had had a Spanish speaking interviewer, however.

Finally, Ms. Grossman hypothesized that some of the people that we
were calling may be drug addicts or criminals themselves and may not believe that it is victimization that we are interested in. These will be
almost impossible to convert.

The next form of nonresponse that was discussed was breakoff interviews. These people are usually busy people. The best approach is to
avoid the breakoff interview in the first place if possible. Some people
will not have the time to finish and will have to break off. Breakoffs are
easy to convert. Remarks can be used such as, "Hello, I'm _____. I
called you last Saturday. We didn't quite get finished then. Let me ask
these last remaining questions." Above all, don't acknowledge if the
person refused. Use remarks instead such as, "You got busy the other day."
or "You had to leave the house." Knowing the circumstances leading to the
breakoff is important and should be documented thoroughly since these
provide the lead in to follow-up conversations.

Ms. Grossman suggested that changes in the instrument design might
reduce nonresponse. For instance, almost the first question that we ask is

the number of persons in the household. This is a sensitive question for individuals living alone and may cause them to become suspicious. Ms. Grossman suggested that it would be better to ask nonpersonal questions about crime first, particularly opinion questions. This would get the respondent interested in the survey and convince them that it is for real and not a crank call or someone selling something. The respondent wants to tell you what he thinks about the subject so give him an opportunity. Then, the credibility of the study will be established and personal questions can be asked.

The final form of nonresponse that we encountered was roster nonresponse. In the D.C. study, the roster was obtained after the first interview. Ms. Grossman felt that this was the most difficult form of nonresponse that we were faced with. In some cases, she felt that the person lived alone but did not trust us enough to admit the fact. In other cases, she suggested that after participating in a long interview the respondent is reluctant to give informaton about other family members so that we can bother them too. The conversion approach that worked best was to get another family member to complete the roster and to do that first. "Someone earlier talked to us. Now we need to complete the information for other family members." In some cases, the original person provided the roster when called back at a later time.

At this point, I discussed with Ms. Grossman the characteristics of hard core nonrespondents - those people who refused and could never be converted. Ms. Grossman indicated that as long as a person will talk to the interviewer, then the interviewer has a good chance to get the required information. Hard core nonrespondents are those who will not talk to an interviewer. These people make remarks such as, "Don't call this house again!" or "I'm not interested." and then hang up immediately. Some of the hard core nonrespondents are anti-goverment people; a very persuasive converter can sometimes get these to respond. In some instances, Ms. Grossman suggested that interviews could be obtained for hard core nonrespondents from other family members if proxy interviews were allowed.

In concluding our conversation, Ms. Grossman gave some tips for interviewers to use in converting nonrespondents and for supervisors to improve response. The conversion tips for the interviewers were:

- When nonresponse occurs, document it as fully as possible with characteristics of person (sex, age, race) and circumstances leading to nonresponse. These provide lead ins when calling back to convert.

- Don't speak too quickly during the introduction - the respondent may feel you are rushing and not attach importance to your call.

- If the original interviewer was able to get the respondent's name, use it when you call.

- Attempt to speak to the respondent rather than someone else in the household. If one spouse refused for another, don't speak to that spouse. If your respondent is not available, thank the person and hang up.

- Examine comments on call record sheet for clues as to best time to reach respondent. If husband refused for wife - make your call during day in hopes he won't be there and vice versa.

- Be positive in your approach - explain what you want and suggest starting the interview now. Appointments are easily broken. Work quickly when the respondent finally agrees. Remember you are dealing with reluctant respondents!

- Work on easiest refusals first to increase production, then if time permits work on others.

- Be ready to counter every objection and above all don't ask them any questions to which the respondent can answer no - and keep talking.

- If necessary to call back - you suggest the appointment time.

- Don't let refusals on the screening forms intimidate you. Actually these are fairly easy refusals to convert. In many instances another household member will answer the phone and in other cases perhaps the original respondent was in a hurry, didn't understand the introduction.

Her suggestions for the supervisory staff were as follows:

- Train the interviewers in how to handle nonresponse, both initially and as the study progresses. Cite examples from your experience.

- Show concern over refusals. Discuss specific refusals with individual interviewers and offer suggestions on how to handle the problem next time.

- Be positive and supportive when interviewers are converting nonrespondents.

- Indoctrinate the interviewers on the importance of a high response rate and good persuasive interviewing techniques. Explain the biasing impact on the study of low response.

- Post completion rates and production figures prominently on a weekly basis. Have a 15 minute meeting each week to present them and to boost morale.

- Make some time available to personally conduct interviews and convert refusals so that you are aware of the problems the interviewers face and so that you can demonstrate that they can be solved.

- Monitor some portion of each interviewer's work each week so that you are aware of the quality of the work that they are doing and how they can improve their performance.

- Evaluate the performance of each interviewer. Those who get excessive nonresponse should be terminated. Alternatively, above average performers should receive recognition and a merit raise. Interviewers who show the talent and willingness to convert nonrespondents should be paid more.

- In training interviewers to convert nonrespondents, demonstrate first, train second, assist as interviewers convert third, and reinforce good work. Say, "I couldn't have got him to respond either." when true or if the interviewer used a poor conversion method say, "Nice try. Next time you might want to try this approach..." Praise the interviewer who completes the conversion.

- Give the interviewer goals to work for and recognize their good work. Compliments are cheap but they raise everyone's morale.

- If you have difficulty with monitoring and participating in the interviewing and conversion process because of paper work, get a clerk or administrative assistant to help with the paper work or try to get the volume of paper work reduced.

Finally, Ms. Grossman noted that these comments were based upon her experience with personal interview surveys and list frame telephone surveys but that they have value for random digit dialed surveys as well. She expressed doubt that a random digit dialed survey could ever achieve response rates as high as those obtained by personal interview or list frame telephone surveys but improvements are possible. Random digit dialed surveys will always require more commitment and effort to obtain satisfactory response rates.

bkp

October 24, 1983

MEMORANDUM

TO: The Record

FROM: Brenda Cox

SUBJECT: Actual Versus Projected Response and Eligibility Rates for the
          District of Columbia Crime Victimization Study

With data collection completed in August, response and eligibility
rates can now be computed for the District of Columbia Crime Victimization
Study. When the sample size determinations were made, we used the avail-
able data from past RTI telephone surveys as well as crime victimization
studies by the Census Bureau and the University of Michigan to project the
rates that we would encounter. This memorandum summarizes that process and
compares these assumptions with our actual survey experience. By this
documentation, it is hoped that insight can be gained into the data collec-
tion process for the D.C. study as well as for future surveys. It should
be noted that minor discrepancies may exist in the rates cited in this
report since some are based upon field counts and others on data base
counts.

To begin with the most complicated survey first, the District of
Columbia Household Victimization Survey (DCHVS) was a telephone survey of
residents of the DC-SMSA. The sample was selected as a stratified random
sample from an ordered list of all telephone numbers assigned to the DC-
SMSA with 40 percent of the sample allocated to D.C. proper and the re-
maining 60 percent to the Virginia and Maryland suburbs. (Approximately 25
percent of the DC-SMSA population lives in D.C. proper.) When a telephone
number was associated with a residence, all individuals 14 and up were
interviewed beginning with adult members of the household. Responses for
12 and 13 year olds were obtained from their parents.

To estimate the distribution of telephone numbers, the experience of a
recent RTI study was used. That study included a telephone survey of
DC-SMSA residents with the sample randomly selected from all telephone
numbers associated with the DC-SMSA. Based upon that study's results, we
estimated that 46 percent of the telephone numbers would be nonworking, 20
percent would be business numbers, and 6 percent would be indeterminable
(mostly ring no answers), leaving 28 percent of the numbers working resi-
dential numbers. An examination of the control cards for that study re-
vealed that nonworking numbers could be identfied in the majority of cases
by a recorded message.

Table 1 summarizes the actual data collection experience and contrasts it with the originally projected experience. Approximately 50 percent of the selected numbers were identified as nonworking. These include 7,500 nonworking numbers, 457 temporarily nonworking numbers, 417 double wrong connections, 466 no result from dial numbers, and 115 fast busy numbers. An additional 22 percent were noneligible working numbers with 51 of these public pay phones, 3,899 businesses and institutions, and 58 other ineligibles (foreign embassies, etc.). A total of 3,728 working residential numbers were identified or 21 percent of the total numbers dialed. Screening interviews were not completed for 7 percent of the sample numbers with 1,071 of these ring no answers, 84 regular busy, 24 language barriers, and 174 refusals. Since the screening interview determines eligibility, the numbers for which screening was not completed were classified as indeterminable. The "regular busy" designation may be a misnomer. Many of these were not a normal busy signal nor were they a fast busy signal. These may not all be working numbers.

Note that we selected 2,098 more numbers than we originally anticipated selecting but that we still identified less working residential numbers than we had projected. This resulted from the fact that only 21 percent of the numbers were working residential numbers instead of 28 percent as we originally projected. To determine if our oversampling of District phone numbers was the cause of this problem, we tabulated the results for DC proper versus the suburbs.

For D.C. proper, 50 percent of the numbers were again identified as nonworking with 4,006 recorded-message nonworking numbers, 361 temporarily nonworking numbers, 269 double wrong connections, 348 no result from dial, and 43 fast busy's. A larger percentage were noneligible working numbers, however. Of the total D.C. proper telephone numbers selected, 27 percent were ineligible working numbers of which 2,721 were businesses or institutions, 27 were public pay phones, and 19 were other ineligibles. A total of 1,419 working residential numbers were identified or only 14 percent of all numbers dialed. Finally, screening interviews were not completed for 9 percent of the sample numbers with 738 of these ring no answer's, 54 regular busy's, 6 language barriers and 64 refusals.

For the D.C. suburbs, 49 percent of the numbers were identified as nonworking with 3,494 recorded-message nonworking numbers, 96 temporarily nonworking numbers, 148 double wrong connections, 118 no result from dial's, and 72 fast busy's. Of the 7,953 D.C. suburban numbers dialed, 15 percent were ineligible working numbers of which 1,178 were businesses or institutions, 24 were public pay phones, and 23 were other ineligibles. A total of 2,309 working residential numbers were identified or 29 percent of all D.C. suburban numbers dialed. Finally, screening interviews could not be completed for 6 percent of the sample numbers with 333 of these ring no answer's, 30 regular busy's, 18 language barriers, and 110 refusals.

These tabulations do indicate that a substantially lower percentage of the assigned telephone numbers for D.C. are working residential numbers than for the suburbs. The patterns described above were also consistent across all three waves of the survey. For those readers desiring more details of the screening results, Tables 2-4 give the results by wave for the DC-SMSA, D.C. proper, and the D.C. suburbs.

The next step in the projections was to specify the response rate that we would achieve in the study. The results of past RTI studies were examined to make the projection as well as the response rates that the University of Michigan had achieved in a similar study. Based upon these past studies, we projected that at least one completed interview would be obtained from 80 percent of the residences that completed the screening interview. To determine the total number of completed interviews, we had to project the average number of persons 12 or older that would be found in these homes. Census data for 1980 was used in projecting that 1.91 eligible persons would be found on the average. Within responding households, we then estimated how many persons would respond given that at least one person had responded. University of Michigan results were again examined. Their results suggested that persons after the first responded at a lower rate than did the first person. For this reason, we projected that subsequent persons would respond at a 75 percent rate. Thus with an average of 1.91 persons within responding households, we could expect to obtain responses from 1.68 persons [1 + .91 (.75)].

In actuality, we obtained at least one completed interview from 3,026 of the 3,728 identified working residential numbers resulting in a response rate of 81 percent. However, not all of the responding residential numbers provided a roster of household members 12 and up. Roster questions were asked after the first completed interview. Of the 3,026 responding households, 2,922 or 97 percent provided rosters. Without rosters, we cannot determine how many additional persons remain to be interviewed, if any. In computing the person within responding household rate, only responding households that provide a roster can be included. From these 2,922 households, we identified 6,637 eligible persons or 2.27 per household. Of the 6,637 persons, we obtained interviews from 5,477 persons or 1.87 per household. This implies that our response rate from subsequent persons within houeholds where at least one person responded and provided a roster was 69 percent. The total number of completed interviews from all responding households (whether or not a roster was completed) was 5,581 or 1.84 per responding household. Thus, we obtained a household response rate that was better than anticipated but a person within-responding-household response rate that was lower than anticipated. We also identified more eligibles per responding household than we had predicted based upon Census data.

At this point, it may be of interest to contrast the experience for D.C. proper versus that for the suburbs.

Within D.C., we obtained at least one interview with 1,142 of the 1,419 identified working residential numbers for a household response rate of 80 percent. Rosters were obtained from 1,102 of these responding households for a roster response rate of 96 percent. Within responding households completing a roster, 2,301 eligible persons were identified or 2.09 per household. We completed interviews with 1,864 of these eligible persons, implying a response rate for subsequent persons within responding households of 64 percent.

For the suburban areas, we obtained at least one interview with 1,884 of the 2,309 identified working residential numbers for a household response rate of 82 percent. Rosters were obtained from 1,820 of these responding households for a roster completion rate of 97 percent. Within

responding households completing a roster, 4,336 eligible persons were identified or 2.38 per household. We completed interviews with 3,613 of these eligible persons, implying a subsequent persons within responding households response rate of 71 percent.

The response rate for the suburban areas of D.C. was higher than for the city itself but not by any appreciable amount except for the subsequent persons within responding households response rate. In all cases, the subsequent persons rate was lower by at least ten percent from the first person rate. This does not necessarily imply that if an eligible person had been randomly selected from each household that the overall response rate would be higher, however. The first interview is obtained from an easier group than subsequent interviews. For instance, the initial interview is conducted with the household's telephone answerer (or with the first household member who is cooperative) who will tend to be more verbally inclined and to not have a physically/mentally incapability or a language barrier. Secondly, if a randomly selected respondent were interviewed instead of every eligible household member, then the rostering would have to be done at the beginning rather than the end of the interview. We rostered after the interview since we felt that asking the sensitive rostering questions first would result in more nonresponse. It might be better to ask selected survey questions first if a respondent were to be randomly selected. For instance if it were O.K. to obtain the household crimes from any responsible person answering the telephone, then the household crime questions could be asked and then the roster obtained and a random respondent selected to provide data on personal crimes.

To provide a better understanding of the person-level response rate, Tables 5-7 summarizes the results for the 6,741 eligible persons identified in the DCHVS. This includes the 104 first persons who completed an interview but did not provide a roster. Interviews were completed for 83 percent of the group with refusal the primary source of nonresponse (8 percent). Another 5 percent of the sample could not be interviewed at all due to physical/mental incapability, language barriers, or nonavailability (out of town during survey period), etc. The response rate was lower for D.C. proper at 81 percent response. Refusals accounted for 8.3 percent of the 19.1 percent nonresponse with another 5.4 percent incapable of being interviewed. A higher response rate of 83.5 was obtained for the D.C. suburbs. The refusal rate was 7.8 percent and incapable of interview was 4.8 percent.

The final item that we had to project was the number of short incident forms and long incident forms that we would have to complete per person. Each person was asked to report the crimes that had occurred since January 1, 1982. The analysis, however, will focus on crimes occurring in the period from May 1, 1982 to April 30, 1983. The short form (Section D of the Core Questionnaire) determined if the event was a crime of interest and if it fell within the analysis time period. If both were true, a long form was completed for the crime (Sections E-O of the Core Questionnaire). To make these projections, National Crime Survey (NCS) data for major metropolitan areas was used. These data were adjusted to account for underreporting anticipated due to the longer DCHVS reference period and for the greater productivity that was projected for the instrument. In costing the study, the assumption was made that the non-NCS reportable crimes of threats and vandalism would not have a long form completed for them. We

projected that 1.607 events would be reported per person within the reference period. Of these events, 1.123 would fall within the analysis time period and 0.337 of these would be NCS crimes. Hence we projected that 1.607 short forms would be completed per sample person and 0.337 long forms.*

In actuality, we obtained 0.828 events per person in the DCHVS. Of these 0.352 were crimes falling within the analysis time period and 0.282 were NCS crimes. A decision was made prior to data collection to complete long forms for non-NCS as well as NCS crimes. Therefore, short forms were completed for 0.828 crimes per person and long forms for 0.352 per person.

Differences were also observed between the central city and the suburbs. D.C. city residents reported 0.820 events per person, of which 0.335 were eligible crimes falling within the analysis time period and 0.292 of these were NCS crimes. D.C. suburban residents reported 0.832 events per person, of which 0.360 were crimes falling within the analysis time period and 0.277 were NCS crimes.

Based upon the assumptions described above, we projected that the cost per completed DCHVS interview would be $21.46. In actuality, we spent $18.88 per completed interview. It should be noted, however, that if the actually occurring rates were used with our estimated cost components, the cost per completed interview would be estimated as $21.07.

The other survey that was done as a part of the study was the Capitol Hill Employees Victimization Survey (CHEVS). The CHEVS was a telephone survey of employees of the Senate, House of Representatives, Library of Congress, Congressional Budget Office, Architect of the Capitol, and the Office of Technology Assessment who had worked on Capitol Hill at some time in 1982. The sample was selected as a stratified random sample from lists provided by the six agencies.

Table 8 presents the assumptions that were made in costing the study. We projected that 2,994 employees would be selected, of which ten percent would need to be traced. Out of these 2,994 employees, we projected that we would complete interviews for 85 percent, that 10 percent would refuse and that 5 percent would not be located. From the 2,545 responding employees, we projected that we would get 4,090 crimes requiring that a short form be completed and 858 that required a long form in addition. For lack of information to the contrary, we used the projected crime rates estimated for the DCHVS. That is, we were presuming 1.607 crimes reported per person, of which 1.123 would fall within the analysis time period with 0.337 of these NCS crimes. Since we were again projecting that long incidence forms would only be completed for NCS crimes, this implies that a total of 1.607 short forms would be completed per person and 0.337 long forms.

---

*In projecting NCS crimes I used 1980 NCS data for cities with a central city of 1,000,000 or more. It would have been more appropriate to use cities with a central city of 500,000 to 1,000,000 since this is the way Census classifies the DC-SMSA.

projected that 1.607 events would be reported per person within the reference period. Of these events, 1.123 would fall within the analysis time period and 0.337 of these would be NCS crimes. Hence we projected that 1.607 short forms would be completed per sample person and 0.337 long forms.*

In actuality, we obtained 0.828 events per person in the DCHVS. Of these 0.352 were crimes falling within the anaysis time period and 0.282 were NCS crimes. A decision was made prior to data collection to complete long forms for non-NCS as well as NCS crimes. Therefore, short forms were completed for 0.828 crimes per person and long forms for 0.352 per person.

Differences were also observed between the central city and the suburbs. D.C. city residents reported 0.820 events per person, of which 0.335 were eligible crimes falling within the analysis time period and 0.292 of these were NCS crimes. D.C. suburban residents reported 0.832 events per person, of which 0.360 were crimes falling within the analysis time period and 0.277 were NCS crimes.

Based upon the assumptions described above, we projected that the cost per completed DCHVS interview would be $21.46. In actuality, we spent $18.88 per completed interview. It should be noted, however, that if the actually occurring rates were used with our estimated cost components, the cost per completed interview would be estimated as $21.07.

The other survey that was done as a part of the study was the Capitol Hill Employees Victimization Survey (CHEVS). The CHEVS was a telephone survey of employees of the Senate, House of Representatives, Library of Congress, Congressional Budget Office, Architect of the Capitol, and the Office of Technology Assessment who had worked on Capitol Hill at some time in 1982. The sample was selected as a stratified random sample from lists provided by the six agencies.

Table 8 presents the assumptions that were made in costing the study. We projected that 2,994 employees would be selected, of which ten percent would need to be traced. Out of these 2,994 employees, we projected that we would complete interviews for 85 percent, that 10 percent would refuse and that 5 percent would not be located. From the 2,545 responding employees, we projected that we would get 4,090 crimes requiring that a short form be completed and 858 that required a long form in addition. For lack of information to the contrary, we used the projected crime rates estimated for the DCHVS. That is, we were presuming 1.607 crimes reported per person, of which 1.123 would fall within the analysis time period with 0.337 of these NCS crimes. Since we were again projecting that long incidence forms would only be completed for NCS crimes, this implies that a total of 1.607 short forms would be completed per person and 0.337 long forms.

---

*In projecting NCS crimes I used 1980 NCS data for cities with a central city of 1,000,000 or more. It would have been more appropriate to use cities with a central city of 500,000 to 1,000,000 since this is the way Census classifies the DC-SMSA.

The lists provided to us for sampling employees of the Senate, House of Representatives, and the Congressional Budget Office (CBO) were not extracted from 1982 personnel records as we had hoped. Instead the CBO sent a telephone directory and the House and Senate sent payment records. Because of this fact the frame was incomplete to an unknown extent and also inaccurate. Not all employees on the lists were 1982 employees and many others did not work on Captiol Hill. This event required that we include an eligibility screening interview prior to the actual interview and induced a new stage at which sample individuals could fail to respond. Finally, the work addresses and telephone numbers were not current, resulting in a substantially increased tracing and locating effort and a greater loss of unable to be located employees. In addition to these frame inadequacy problems, the population as a whole was a somewhat sensitive group to interview. For this reason, we were instructed by the client to forego extensive nonrespondent conversion.

A total of 2,504 employees were selected for the sample of which 1,979 were screened eligibles and 219 were screened ineligibles for a screening completion rate of 87.8 percent. A total of 157 employees or 6.3 percent of the sample were not screened because we were unable to contact them. An additional 23 employees or .01 percent of the sample were not available during the survey period, or were physically/mentally incapable of interview or deceased. Of the remaining nonrespondents, 219 employees or 8.7 percent of the sample refused screening.

Of the 1,979 employees screened and identified as eligible, 1,890 completed and interview for an interview response rate of 95.5 percent. The nonresponding employees included 3 breakoff interviews (0.2%), 59 refusals (3.0%), 9 employees not available during the interview period (0.5%), 6 employees who were deceased or otherwise physically/mentally incapable of being interviewed (0.3%) and 12 other nonrespondents (0.6%).

The 1,890 responding employees reported 0.968 events per person in the CHEVS. Of these 0.447 were crimes falling within the analysis time period and of these 0.355 were NCS crimes. Thus, short forms were completed for 0.968 crimes per person and long forms for 0.447 crimes per person, rather than the 1.607 short forms and 0.337 long forms that we had projected.

For the interested reader, we have attached Tables 9 and 10 providing the screening and interview results by wave.

Based upon the assumptions described earlier, we projected that the cost per completed CHEVS interview would be $19.68. In actuality, we spent $25.20 per completed interview. However, we cannot project the costs using the actually occurring rates since exact counts are not available for the number of employees requiring tracing.

bkp

Table 1. Projected Versus Actual Sample Sizes for the District of
          Columbia Household Victimization Survey (DCHVS)

Projected Sample Size	Actual Sample Size	Sample Component
15,946	18,044	Telephone Numbers Selected
7,335	8,955	Nonworking Numbers
3,189	4,008	Government/Business Numbers
957	1,353	Indeterminable Numbers
4,465	3,728	Working Residential Numbers
3,572	3,026	Responding Residential Numbers
6,823	6,741	Eligible Persons Identified
6,000	5,572	Responding Persons
9,642	4,599	Victimizations Reported
6,738	1,953	Victimization Reported for Analysis Time Period
2,022	1,567	NCS Crimes Reported for Analysis Time Period

Table 2. DCHVS Telephone Screening Results: DC-SMSA

Result	Wave I		Wave II		Wave III		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Nonworking	2,630	43.2	2,456	40.3	2,414	41.1	7,500	41.6
Temporarily Nonworking	134	2.2	180	3.0	143 ·	2.4	457	2.5
Double Wrong Connection	141	2.3	130	2.1	146	2.5	417	2.3
Business or Institution	1,239	20.4	1,348	22.2	1,312	22.4	3,899	21.6
No Result from Dial	163	2.7	170	2.8	133	2.3	466	2.6
Fast Busy	57	0.9	28	0.5	30	0.5	115	0.6
Ring No Answer	372	6.1	361	5.9	338	5.8	1,071	5.9
Public Pay Phone	18	0.3	15	0.2	18	0.3	51	0.3
Working Residential	1,250	20.5	1,279	21.0	1,199	20.4	3,728	20.7
Refusal	49	0.8	54	0.9	71	1.2	174	1.0
Regular Busy	12	0.2	37	0.6	35	0.6	84	0.5
Other	17	0.3	16	0.3	25	0.4	58	0.3
Language Barrier	5	0.1	13	0.2	6	0.1	24	0.1
Total	6,087	100.0	6,087	100.0	5,870	100.0	18,044	100.0

Table 3. DCHVS Telephone Screening Results: DC City.

Result	Wave I		Wave II		Wave III		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Nonworking	1,419	42.2	1,278	38.1	1,309	39.0	4,006	39.8
Temporarily Nonworking	104	3.1	135	4.0	122 ·	3.6	361	3.6
Double Wrong Connection	96	2.9	86	2.6	87	2.6	269	2.7
Business or Institution	828	24.6	954	28.4	939	28.0	2,721	27.0
No Result from Dial	120	3.6	125	3.7	103	3.1	348	3.5
Fast Busy	28	0.8	6	0.2	9	0.3	43	0.4
Ring No Answer	249	7.4	233	6.9	256	7.6	738	7.3
Public Pay Phone	12	0.4	8	0.2	7	0.2	27	0.3
Working Residential	471	14.0	478	14.2	470	14.0	1,419	14.1
Refusal	22	0.7	22	0.7	20	0.6	64	0.6
Regular Busy	3	0.1	25	0.7	26	0.8	54	0.5
Other	7	0.2	2	0.1	10	0.3	19	0.2
Language Barrier	1	0.0	4	0.1	1	0.0	6	0.1
Total	3,360	100.0	3,356	100.0	3,359	100.0	10,075	100.0

113

Table 4. DCHVS Telephone Screening Results: DC Suburbs

Result	Wave I		Wave II		Wave III		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Nonworking	1,211	44.4	1,178	43.2	1,105	44.2	3,494	43.9
Temporarily Nonworking	30	1.1	45	1.7	21 ·	0.8	96	1.2
Double Wrong Connection	45	1.7	44	1.6	59	2.4	148	1.9
Business or Institution	411	15.1	394	14.5	373	14.9	1,178	14.8
No Result from Dial	43	1.6	45	1.7	30	1.2	118	1.5
Fast Busy	29	1.1	22	0.8	21	0.8	72	0.9
Ring No Answer	123	4.5	128	4.7	82	3.3	333	4.2
Public Pay Phone	6	0.2	7	0.3	11	0.4	24	0.3
Working Residential	779	28.6	801	29.4	729	29.1	2,309	29.0
Refusal	27	1.0	32	1.2	51	2.0	110	1.4
Regular Busy	9	0.3	12	0.4	9	0.4	30	0.4
Other	10	0.4	8	0.3	5	0.2	23	0.3
Language Barrier	4	0.1	9	0.3	5	0.2	18	0.2
Total	2,727	100.0	2,725	100.0	2,501	100.0	7,953	100.0

Table 5.   DCHVS Person Interview Results: DC-SMSA

Result	Wave I		Wave II		Wave III		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Interview Completed	1,885	83.8	1,888	82.1	1,799	82.1	5,572	82.7
Breakoff Interview	41	1.8	14	0.6	16 ·	0.7	71	1.1
Refusal	158	7.0	222	9.7	158	7.2	538	8.0
Not Available During Survey	53	2.4	65	2.8	77	3.5	195	2.9
Language Barrier	7	0.3	13	0.6	10	0.5	30	0.4
Physically/Mentally Incapable	36	1.6	42	1.8	30	1.4	108	1.6
Deceased	1	0.0	3	0.1	0	0.0	4	0.1
Other Nonresponse	69	3.1	52	2.3	102	4.7	223	3.3
Total	2,250	100.0	2,299	100.0	2,192	100.0	6,741	100.0

115

Table 6. DCHVS Person Interview Results: DC City

Result	Wave I		Wave II		Wave III		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Interview Completed	638	81.3	607	79.6	654	82.5	1,899	81.1
Breakoff Interview	22	2.8	6	0.8	8 ·	1.0	36	1.5
Refusal	66	8.4	79	10.4	49	6.2	194	8.3
Not Available During Survey	14	1.8	27	3.5	25	3.2	·66	2.8
Language Barrier	1	0.1	6	0.8	7	0.9	14	0.6
Physically/Mentally Incapable	16	2.0	15	2.0	13	1.6	44	1.9
Deceased	1	0.1	1	0.1	0	0.0	2	0.1
Other Nonresponse	27	3.4	22	2.9	37	4.7	86	3.7
Total	785	100.0	763	100.0	793	100.0	2,341	100.0

Table 7.   DCHVS Person Interview Results: DC Suburbs

Result	Wave I		Wave II		Wave III		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Interview Completed	1,247	85.1	1,281	83.4	1,145	81.8	3,673	83.5
Breakoff Interview	19	1.3	8	0.5	8 ·	0.6	35	0.8
Refusal	92	6.3	143	9.3	109	7.8	344	7.8
Not Available During Survey	39	2.7	38	2.5	52	3.7	129	2.9
Language Barrier	6	0.4	7	0.5	3	0.2	16	0.4
Physically/Mentally Incapable	20	1.4	27	1.8	17	1.2	64	1.5
Deceased	0	0.0	2	0.1	0	0.0	2	0.0
Other Nonresponse	42	2.9	30	2.0	65	4.6	137	3.1
Total	1,465	100.0	1,536	100.0	1,399	100.0	4,400	100.0

117

Table 8. Projected Versus Actual Rates for the Capitol Hill
Employees Victimization Survey

Projected Count	Actual Count	Sample Component
2,994	2,504	Employee Listings Selected
2,844	1,979	Eligible Employees
0	219	Ineligible Employees
0	109	Screening Refusals
150	157	Unable to Locate Cases
0	40	Other Screening Nonresponse
2,545	1,890	Identified Eligibles Responding
4,090	1,829	Victimizations Reported
2,858	845	Victimizations Reported for Analysis Time Period
858	671	NCS Crimes Reported for Analysis Time Period

Table 9. Screening Results for the CHEVS

Result	Wave I		Wave II		Wave III		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Screened and Eligible	864	82.4	800	76.3	315	77.4	1,979	79.0
Screened and Ineligible	81	7.7	107	10.2	31	7.6	219	8.7
Breakoff/Partial Data	0	0.0	0	0.0	0	0.0	0	0.0
Refusal	40	3.8	56	5.4	13	3.2	109	4.4
Not Available During Survey	3	0.3	5	0.5	10	2.5	18	0.7
Unable to Contact	56	5.3	65	6.2	36	8.9	157	6.3
Deceased	2	0.2	0	0.0	0	0.0	2	0.1
Physically/Mentally Incapable	1	0.1	0	0.0	2	0.5	3	0.1
Other Nonresponse	2	0.2	15	1.4	0	0.0	17	0.7
Total	1,049	100.0	1,048	100.0	407	100.0	2,504	100.0

119

Table 10. Interview Results for the CHEVS

Result	Wave I		Wave II		Wave III		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Interview Completed	813	94.1	780	97.5	297	94.3	1,890	95.5
Breakoff Interview	1	0.1	2	0.3	0	0.0	3	0.2
Refusal	36	4.2	14	2.1	9	2.9	59	3.0
Not Available During Survey	3	0.3	3	0.4	3	1.0	9	0.5
Deceased	0	0.0	0	0.0	0	0.0	0	0.0
Physically/Mentally Incapable	4	0.5	0	0.0	2	0.6	6	0.3
Other Nonresponse	7	0.8	1	0.1	4	1.3	12	0.6
Total	864	100.0	800	100.0	315	100.0	1,979	100.0

November 18, 1983

TO: The Record

FROM: Danny Allen

SUBJECT: D. C. Crime Data Processing Activity Summary

The D. C. Crime Victimization Study has been RTI's first experience with implementation of a large and complex CATI application. To a great extent this application has to be considered a learning experience. The following project summary is centered around post CATI project activities. In addition, suggestions for future CATI applications are identified.

CATI Record Structure

Each CATI observation requires a fixed length record corresponding to all potential data to be collected for the given interview. For the D. C. Crime Victimization Study, the data record had 5,616 characters, but 6,143 were used in order to end on a buffer boundary. The record contained CATI interview control information, person data, and crime data. In order to allow for multiple crimes, there were 20 identical sections for short form crime data (Section D) and 6 identical sections for the long form crime data (Section E-O). Variables were established in CATI for purposes of linking long form data with appropriate short forms.

CATI programmers had to be extremely careful when programming repeat sections. Extensive code had to be generated (i.e., code for each repeat). Sets of code for repeat sections had to be identical by order of variables and widths of fields. Also, the repeating sections required new identifiers, output positions, etc.

This is in contrast to direct data entry procedures where a single definition of code is used for a given repeat. Direct data entry code is usually recorded only once and allows considerable flexibility in the number of repeats needed for a given instrument. The direct data entry structure also provides a means of considerably reducing space since a fixed number of repeats does not have to be defined.

Codebook Generation

The first step in data processing was to develop software to read and generate a codebook directly from the CATI screen file code. (As a point of emphasis a listing of the CATI screen file code was in excess of four inches of computer printout.) The computerized codebook consisted of variable definitions and controls used directly by CATI; however, it did

not include logic statements. The codebook was used by essentially all postprocessing steps.

A brief codebook review revealed that CATI output positions were scattered. This was the result of changes made in the questionnaire after CATI programming began. The codebook was sorted by CATI output position; Software was developed to read the sorted codebook and check for duplication or gaps in output positions.

Output revealed several duplicate CATI output positions and one case of duplicate output beginning position but a different number of characters for output. After reviewing reasons for duplication of output positions with CATI programmers, it was determined that the duplication was intentional and valid although the difference in the number of output positions was in error. The reasons for the duplication of output positions were based upon questionnaire flow and programming techniques.

The next step in codebook development involved eliminating duplicate output positions. Determination of variable definitions to keep was based upon maintaining those that appeared to have the most logical position within the initial codebook. Codebook IDs of records to be flagged as duplicates were keyed into a control file. The codebook and corresponding control file were sorted by ID. This provided input to software that was developed for purposes of flagging records as duplicates. The resultant output produced a new codebook file with a single definition for all output positions. This was verified by rerunning the software previously developed to check for duplicates in output positions or record gaps.

Review of the revised codebook revealed considerable scattering of output positions when compared to logical questionnaire flow. Reasons for this had to do with (1) CATI restrictions, (2) multiple CATI programmers, and (3) instrument changes made after CATI programming began.

The next step to codebook development resulted in redefining the order of codebook variables. The intent was to provide a mechanism for re-ordering variables in data records so the data record structure would correspond to the logical flow of the questionnaire. Considering the size of the codebook and the extent of variable scattering, the approach used was used to create a control file of variable IDs to be moved and corre-sponding relocation position within the codebook. Software was written to generate a new codebook with revised sequencing. Sorting the codebook on the new sequence number provided a codebook of single definitions for each variable and codebook variables were ordered in the desired logical ques-tionnaire flow.

Upon completion of WAVE 1 keying, the codebook and WAVE 1 data were copied to tape from the VAX (where CATI interviewing occurred) and trans-ferred to TUCC for data processing and analysis. A backup file was created of WAVE 1 data and processing of the data began.

Multiple Response Questions

The first postprocessing step of CATI data involved reformatting the data to conform to the record structure defined in the codebook. This

involved directly copying single response questions and recoding multiple response questions which at times required field expansion. The approach taken for CATI software development and handling of these questions varied by programmer to some extent. However, the general approach was as follows:

1. If the number of possible responses was less than nine, the number of entry fields allocated corresponded to the number of possible responses.

2. If greater than eight possible responses, eight maximum entry fields were provided in the sections 'E-O'. For section 'P' the number of entry fields corresponded to the number of possible responses.

3. Multiple response fields permitted entry of any value in any order. Thus, there were no designated fields for given responses. (This permitted duplication of responses.)

4. A "DON'T KNOW" or "REFUSAL" code in the first field was to be the (This was not always the case.)

5. Blanks in the first response field were to be the determining factor for a legitimately skipped question. (This was not always true.)

Recode Program

Software was developed for purposes of recoding and restructuring multiple response questions. The approach taken was to assign specific fields for each possible response. The stacked responses recorded during the CATI operation were reassigned to designated fields in the data record. In some cases this required expansion of the number of fields to allow for all possible codes.

Example: A question with 12 possible codes

-   CATI allowed eight fields and the values recorded were 4, 5 and 9 in the first three entry fields

-   Restructuring of the record provided twelve fields with each response having its designated position. The result of recoding generated response codes of '1' in fields 4, 5 and 9 and remaining fields were designated as nonresponse.

Further explanation of this procedure are defined in the memo "D.C. Crime Multiple Response Questions" in Appendix A.

Check Program

Software was developed to perform checking of the results of the recode program. Original data was compared against the output of the recode program. CATI data situations were discovered whereby (1) the first entry position(s) were blank and data followed, (2) values were not right

justified in entry fields, (3) and criteria for determining "DON'T KNOW," "REFUSAL" and nonresponse were not always reliable. Hence, corrections were ncesssary.

Split Program

Data for all result codes was restructured into three files per wave. File type 1 is considered the "PERSON" file. It contains all person level data including Sections A, B, C and P. Record identifiers and CATI control variables are also maintained within the file. File type 2 is a file of short forms (Section D) that do not have corresponding long forms. Record headers are available for purposes of linking to file types 1 and 3. File type 3 contains all short forms that have a corresponding long form (Sections E-O).

This process makes a much more efficient use of storage space by eliminating all blank repeat sections. It also provides a more efficient record structure for further processing of data. Observation has revealed that most of the CATI allocated record space was never used; however, there were occasions when the space was not sufficient to record all needed long form repeats.

The procedure for restructuring was as follows:

1. Person data was extracted for each record and written to the person file.

2. The input record was scanned in sequential order for occurrences of completed long forms.

3. Completed long forms were linked to appropriate short forms and then written to the short/long file. If proper linkage did not exist, error messages were printed.

4. Corresponding short form data in the input record was flagged as "used".

5. Steps 2-4 were repeated for all possible occurrences of long forms.

6. Next, all short form sections that were not flagged as "used" or blank were written to the short form file.

Type of Crime (TOC) Coding

Type of crime coding was initially implemented based on specifications that resulted in multiple classifications of some crime reports. Results were reviewed by analysts and decisions were made to revise TOC coding procedures to incorporate a hierarchial ordering to prevent this problem. The new procedures were implemented. Final corrections and review has now been completed. The TOC variable will be appended to records in file type 3 (i.e., long records with corresponding short forms).

Weight File Extraction

Data was extracted to create a file to be used for the computation of weights. Results revealed a need for additional data processing. CATI software was defined to collect certain household data based upon responses obtained by the first person interviewed within the housing unit. Subsequent persons within the housing unit were not asked the questions with the assumption that data for the first person would be directly linked to all others interviewed.

Copying First Person Data and Generation of an Income Variable

Software was developed to copy data collected for the "first person only" to subsequent person records within the HUID. Also, an income variable was created to define the level of income within a housing unit (HUID). The income variable was added to each person level record for completed interviews. Else, the code was identified as missing. The variable was assigned the following values based upon responses to questions "P16a - f:"

Code	Income ($)
1	0 - 4,999
2	5,000 - 9,999
3	10,000 - 14,999
4	15,000 - 24,999
5	25,000 - 29,999
6	30,000 - 49,999
7	50,000 and Above
8	Don't Know
9	Refusal

Problems with Person Level Records

Unfortunately the above process of copying person level data revealed the following data problems. (It should be remembered that the data was being processed without post CATI edit.)

1. missing HUIDs, person identifiers and/or phone numbers,

2. miskeyed HUIDs,

3. multiple first person identifiers for a HUID,

4. no first person identified within some HUIDs,

5. more than one person interviewed within a housing unit; however, the first person interviewed was a breakoff and thus household data was not collected for the housing unit,

6. phone numbers of all 9s or possibly blanks,

7. blank result codes.

Efforts were implemented to identify and correct these problems. (See memo "D.C. Crime - Person 1 Data and Income Coding," Appendix A). One thing is evident. Keying error and transpositions of numbers have contributed to HUID problems.

Post CATI Edit Needs

Postprocessing of data was implemented based upon the assumption that CATI would produce a file that was essentially ready for production applications. This was far from the truth. Situations attributing to this include CATI restrictions, variations in programming techniques, keyer error, program changes etc. Thus, various post CATI processing steps had to be implemented in order to create a desirable file for analysis and file delivery. It is evident there is a need for established quality control procedures for all CATI applications.

Suggested Areas of Improvement and Consideration for Future CATI Applications

Post CATI programming activities have definitely demonstrated that one cannot assume that CATI produces a clean data file ready for analysis. However, this has been a first time effort for an application as complex as the D. C. Crime Victimization Study. Many problems can certainly be avoided for future applications. Based upon experience to date, needed areas of improvement and consideration for CATI applications include:

1. a thorough understanding of the CATI application by project task leaders,

2. a single source of documentation other than the CATI source that identifies where program specifications deviate from the questionnaire,

3. retention of all variables that may have to be recreated,

4. record structuring within the confines of CATI that would simplify postprocessing,

5. generalized CATI techniques and procedures where feasible,

6. consistent programming techniques within a given application,

7. restrictions on program changes after implementation, especially inserts,

8. documentation and distribution of all changes to CATI software and a mechanism for identifying all records affected by changes,

9. generalized techniques for handling multiple response questions ment to designated positions,

10. informing others of potential "pitfalls" such as data problems that can be created by various keyer actions,

11. as much control as possible to eliminate keyer generated problems and communication with interviewers to explain proper use of CATI,

12. strict control over keying record identifiers, linking variables, etc. A check digit routine would prove useful. A double keying technique might be used until a check digit routine is available,

13. realizing that applications with repeat sections are more prone to error especially when the system requires duplication of code,

14. avoiding too tight a restriction on field widths that can create problems and not allow for sufficient codes,

15. right justification and preferably left zero fill of all categorical variables,

16. avoiding the combination of blanks and zeros to mean legitimate skip,

17. utilization of previously developed codebook generation software prior to CATI implementation for debugging purposes,

18. establishment of consistent codes for nonresponse, don't know, refusals, etc.,

19. improving programming efficiency and record structure requirements for applications with repeat sections. As an example, the D. C. Crime Victimization Study required 20 repeats of one section and six repeats of another. A specified number of repeat sections was mandated based on CATI record structure requirements. Sets of code corresponding to each repeat was required. This structuring (a) usually resulted in significant space that was required but not used, (b) did not permit recording of data that exceeded repeat restrictions, (c) provided the likelihood for interjecting programmer error and (d) had impacts on system requirements,

20. a definite need for established quality control procedures and post CATI editing procedures.

Suggested Review for CATI Applications

A significant level of effort could be devoted to ascertaining reasons for all data problems encountered. Reasons likely include specification errors, programming techniques, keyer error, and functions not yet realized in terms of how CATI does and does not function. It is also evident that some problems result from limitations imposed by CATI and techniques used to "make things work." An indepth study of various situations is not being done on this project. Types of problems encountered and materials including data files, are available for a thorough analysis of situations