

Robert M. Groves, James M. Lepkowski, Peter V. Miller, The University of Michigan

1. Introduction.

As part of a redesign effort of the National Crime Survey (NCS) involving researchers and survey organizations throughout the country, the Survey Research Center of The University of Michigan is investigating two areas of the NCS design. The first of these is the alteration of the NCS questionnaire to reduce the response errors to which the survey is subject. This work was stimulated both by early experimental tests of alternative instruments, which found that failure to report victimizations was the dominant response error, and by the effects of the addition of supplements, which generally were found to increase reporting of victimizations. The research activity involves the construction and testing of alternative instruments through the use of intensive interviews and reverse record check studies (i.e., designs which draw samples of known victims, often from police record systems, validates interview data on the reported crimes, and measures the extent of underreporting of these incidents). The second area of research involves a reexamination of several sample design features of the NCS in order to improve the precision of statistics given high priority by the sponsoring agency, the Bureau of Justice Statistics.

We are conducting the questionnaire design and sampling research in collaboration with other NCS redesign consortium members; many of the developments described here are the result of the sharing of ideas which the consortium has made possible and facilitated. This paper reports on our activities to date and describes future plans for the redesign effort.

2.0 Response Error Issues in the Measurement of Victimization

The current NCS instrument is divided into three parts: 1) a "control card," which contains information useful to the interviewer in reinterviews of the same household (e.g., age, sex, and identification of all household members); 2) a "screener" questionnaire, which asks whether the respondent has been a victim of specific crimes in the last six months; and 3) an incident report, which obtains detailed information about each distinct victimization mentioned in the screener questionnaire. One person, the "household respondent," provides the information on the control card and answers a set of "household victimization" questions; all persons 14 years or older more are asked to answer the other screener items and the incident form. Proxy respondents are permitted for household members who are 12 or 13 years of age. Sample housing units are enrolled in the sample for seven interviews taken at six month intervals. The first interview is used as a "bounding" interview only. Victimizations that are reported both on the first and the second interview are eliminated from tabulations from

the second interview. The data obtained in the bounding interview are not used for estimation in the survey. If a household moves out of a sample housing unit, the household that replaces the movers is interviewed for the remainder of the scheduled seven interviews (without another bounding interview).

There are a variety of response error issues which the instrument redesign work is addressing: 1) conceptualization of victimizations, 2) the desirable scope of victimization measurement (i.e., what crime types should be included in the instrument), 3) methods to reduce the underreporting of victimizations, and 4) other methodological issues, such as the optimal length of the reference period (i.e., the length of the time period for which victimizations are to be reported), the most desirable method of bounding responses, the extent of panel bias due to panel mortality and response error associated with reinterviews, and any telephone or personal interview differences that affect NCS data.

2.1 Conceptual Issues in the Measurement of Victimization

The National Crime Survey is subject to some measurement errors that arise not from the weaknesses of particular questions or interviewer procedures, but from the ambiguity of criteria used to determine when distinct victimizations exist or whether events are part of a single crime. For example, an argument between neighbors leads one neighbor to come over to the other's home and threaten to beat him up; as the neighbor is leaving the other's home he turns and hits the man. Is this a single incident involving an assault or two separate incidents, one involving a threatened assault, the other an actual assault? What would be counted if the man returned for the assault later in the day? . . . the next day? . . . the next week? The example illustrates that for some cases the notion of victimizations as discrete events of separate types is difficult to apply.

This lack of discreteness between incidents is exemplified by the special case of "series" victimizations. A series victimization is a set of incidents that occur so frequently and routinely that the respondent cannot distinguish the individual events. For example, a wife is beaten or abused by her husband continuously. When asked to enumerate the separate times that the beatings occurred in the last six months the wife cannot supply the information. Currently, the NCS counts a victimization as a series incident when it has occurred three or more times in the last six months. It is clear, however, that 1) a large portion of the total victimizations in the United States are of this type (perhaps more than 25%), and 2) the ability of respondents to separate distinct events varies, and they may indeed be able generally to distinguish more than three incidents.

Currently, the screener questionnaire provides no mechanism for aiding the respondent to recall distinct events in order to obtain better estimates of the frequency of the crime. The questionnaire procedures may also identify some events as "series" which could actually be described by respondents as individual incidents.

Related to the conceptual problems of criteria for identifying separate incidents is a problem of different meanings of the word "crime," and different thresholds concerning what constitutes a criminal victimization. For example, a white suburbanite is walking down the sidewalk in a black inner city area. A group of teenagers on the corner is blocking the sidewalk and tells the person that they don't like outsiders in their neighborhood. The teenagers do not touch the suburbanite. It is to be expected that people will vary greatly in their recall of this incident and in their classification of the event as a threatened assault. For persons who often walk the neighborhood sidewalks and experience such incidents, it may seem like an everyday event, part of living in that environment, and not a criminal victimization. For someone not familiar with the area, it might be easily recalled, classified as a crime, and reported to the interviewer.

We have revised the screener instrument to aid the reporting of victimizations in several ways: 1) an alteration of the wording of some screener questions and a supplementation to the screener questions 2) the use of a new dating procedure for victimizations mentioned during the screener, and 3) having each respondent answer all screener items, rather than designating a household respondent for certain crimes. Some of the changes attempt to inform the respondents that even events that they might think are too trivial to report should be considered:

Many people do not find it easy to remember things we ask about, so it's necessary to be able to ask a number of specific questions.

The new dating procedure which is now being tested is:

Did this kind of thing happened more than once since _____?

It's important for us to find out as exactly as possible when this incident happened. In what month did it happen?

(If incomplete answer) Could you tell me approximately when it happened?

(If happen more than once) It's important for us to find out as exactly as possible when things like this happen. In what month did this happen the LAST time?

How about the time BEFORE that? In what month did THAT incident happen?

REPEAT UNTIL ALL INCIDENTS HAVE BEEN DATED

We hope that such a procedure will aid the respondent in recalling distinct incidents in order to purify the estimation of total numbers of incidents in analysis.

We will interview all respondents in the sample households, and each will respond to questions which might uncover "household" crimes. We are using this procedure because some of the events which may be considered as happening to the entire family may be particularly salient for only one member, and thus may go unreported by a "household" respondent. We have to provide, however, for the possibility of multiple reports of the same events.

2.2. Scope of Victimization Measurement

The design of the current NCS questionnaire uses the screener instrument to obtain a list of all incidents for a given respondent. For each incident elicited by the screener, the respondent is asked a set of specific questions about the nature of the event (the so-called incident report). Because of this design the screener bears the burden of full enumeration of victimizations experienced by the respondent. To the extent that questions in that part of the NCS instrument do not elicit reports of some crimes, those crimes will not be included in the enumeration. Given these observations, initial efforts at increasing recall of victimizations have concentrated on the screener questionnaire.

One way to increase the completeness of enumeration of victimizations is to increase the scope of crime types covered by screener questions. Several different kinds of crimes have been considered for inclusion: 1) vandalism (e.g., destruction of mailboxes, painting graffiti on structures, driving automobiles across lawns), 2) arson (e.g., setting fire to garage or automobile); 3) intimidation (e.g., threats to make a person do something that they don't want to do), 4) fraud (e.g., failure to deliver services paid for); 5) deliberate poisonings; 6) bombings and bomb threats.

The criteria for addition to the screener included. 1) the ability to form measures that permit the respondent to recall an incident accurately, 2) expected number of victimizations included in the category, and 3) expected effects of the inclusion of new items on the present screener items (e.g., confusion about the distinction in meaning between two crime types). Although it does seem clear that the items added for new crime types should increase reporting of those specific crimes, there is some debate about the effects of the new items on reporting for the old items. There are two opposing hypotheses about the inclusion of new items. First, the use of added questions might stimulate recall of events that would not be obtained by the current screener. This is based on the belief that questions on crime victimization, regardless of their specific content, facilitate the memory of events that fall into one or more of the categories of the screener questions. For example, asking a question about vandalism may stimulate the memory of a break-in to a home that involved

vandalism and burglary. Another hypothesis notes that this effect may be already maximized by the current screener items and that the addition of new items may produce sufficient fatigue among respondents that recall of victimizations would decrease. Given the relatively short length of the current screener (8 to 15 minutes), we have been more persuaded by the argument that more questions may yield more reporting.

Following this rationale, items on vandalism and intimidation were added to the questionnaire, for example:

Since . . . did anyone vandalize or set fire to anything you own?

How about threats? Did anyone threaten to attack or hurt you?

Items on arson (besides that connected with vandalism) and fraud were not added because of difficulty in phrasing the questions and ambiguity about the criteria for determining that a fire was the result of arson or that criminal fraud actually occurred. Items on poisonings and bombings were not included because of suspicions that too few persons would be victims of such crimes to justify their measurement and difficulties in constructing accurate measures about them.

2.3. Failure to Report Victimization

Past reverse record check studies (LEAA, 1972), results of other large scale surveys that ask the respondent to recall distant events (Cannell *et al.*, 1977) and guidance from psychological theory suggest that the one of the major response errors in the NCS is failure to report a victimization. This underreporting occurs, it is thought, both because some victims fail to recall the event and because some events although recalled are not reported.

Several alterations in the screener were made to improve the likelihood that the respondent would recall victimizations. One method of doing this followed the findings of cognitive psychologists that the attributes of objects or events are recalled more easily and more quickly than the object itself. There are results from the NCS that suggest a similar hypothesis. In 1977 a supplement to the NCS questionnaire asked questions about the perceived seriousness of different criminal acts. An experiment was included which placed the attitude supplement before the screener questions in a random half-sample and after the screener in the complement portion of the sample. There was evidence from this experiment that the attitude supplement occurring before the screener acted to heighten recall and reporting of victimizations during the screener. The questions in the supplement may be considered measures of the attributes of victimizations. Despite the fact that they are not focused on specific events experienced by the respondent they serve to stimulate long term memory of personal victimizations.

Following the logic of these hypotheses a set of items was introduced in front of the experimental screener to stimulate memory of the emotional states of the respondent that are often achieved during a victimization experience

There are some situations which can make people really mad or "burned up." We'd like to find out if you've felt that way any time in the last six months -- that is from _____ -- because any of the following situations.

Was there ever a time (since _____) that you felt really mad because you felt someone cheated you?

CONTINUE WITH SEVERAL OTHER ITEMS OF THIS NATURE

Now I'd like to ask you some similar questions about situations which often make people really scared.

Was there ever a time (since _____) that you felt scared because you didn't feel safe out at night?

CONTINUE WITH SEVERAL OTHER ITEMS OF THIS NATURE

Another hypothesis regarding failure to recall victimizations notes that the short length of the current NCS screener questionnaire impedes careful consideration of all possible kinds of victimizations. That is, the sheer number of questions about victimizations, of any sort, may serve to stimulate memory of an incident. Following this logic, some screener questions were expanded to include probes about different subcategories of victimizations falling under a general heading:

Did anyone take or try to take a motor vehicle you owned or were using?

Did anyone use a vehicle or yours without your permission?

Did anyone take or try to take parts or accessories from a vehicle you owned, for instance a battery, hubcaps, tapedeck, tires or motor parts?

Did anyone do any damage to a motor vehicle or bicycle you own?

Did anyone take something you left INSIDE any motor vehicle?

Did anyone try to break into any motor vehicle you owned or were using?

Another hypothesis about recall error flows from the psychological notion of response set. In this case, the vast majority of persons, if responding accurately, will answer "no" to all screener questions. They have not been a victim of any crime. There may be a tendency for respondents to fall into a habit of saying "no"

2.4. Use of Telephone Interviewing Procedures

to the screener questions to such an extent that they do not seriously consider each screener question individually. Because of this lack of attention to each question, "no" answers are obtained on questions that should receive "yes" answers.

In order to work against this tendency toward a "no" response set, two changes in the screener instrument were introduced. The first ends each question set with a question that verifies that nothing has happened for those who have answered "no" to each preceding question. For example,

So you would say that no one has gotten into your home, any motor vehicles or other places where you keep things since _____?

Another hypothesis about response errors notes that one attribute of a victimization is the place where it occurred. Following this observation, it is argued that improved recall may be obtained if the instrument asks about different locales in which a victimization could have taken place. For example,

Since _____, was something you own taken from a place you were staying -- like a friend's house, hospital, or motel?

Another source of underreporting of victimizations occurs because respondents are hesitant to report to an interviewer an incident that they remember. Such a tendency has been noted in a variety of survey situations and is often associated with questions that have clearly "socially desirable" responses among those offered. One such case may be the measurement of rapes or attempted rapes, an incident which the respondent may be embarrassed to report to an interviewer. Separate from the effects of social desirability are possible fears that respondents may have about reprisals for reports. Two circumstances prompt this suspicion: 1) Some violent acts that occur within a household have high probabilities of repetition. The respondent may fear that the offender within the family may learn of the report to the interviewer. 2) Evidence from group interviews we conducted in Detroit with elderly respondents suggests that they do not report crimes to police because of fears that the neighborhood youth suspected as offenders will seek revenge for the report. Both of these are fears that may reduce reports of several crime types. The revised NCS screener employs some instructions to the respondent in an attempt to increase reporting of these events:

Now I'd like to ask you about any times since _____, when someone attacked or hurt you, tried to hurt you, or threatened to hurt you. We'd like to know even about things that seem small or unimportant to you.

A substantial proportion (over 75%) of current NCS interviews are administered by telephone. However, Census interviewers receive little training in telephone survey techniques, and the instrument is not altered for delivery in the telephone mode. One of the attractions of telephone surveys is the lower cost of that medium relative to personal interviews. One proposal for a revised NCS is the use of even a greater amount of telephone interviewing (see discussion of dual frame designs in the later sections of this paper).

Part of the questionnaire development activities at the Survey Research Center have addressed questions of adapting an NCS instrument to telephone usage. This has taken two different forms: 1) the adjustment of questions that require visual observation either by the respondent (in a response card) or an interviewer (in observations of housing structure type) and 2) the use of a computer assisted telephone questionnaires to introduce new measurement forms. The most radical changes in measurement are those in the second category.

The use of computer assistance in a questionnaire permits more complex flow between questions in the instrument. In the NCS instrument we have made several changes:

- a) the introduction of separate question sets about safety measures taken in the home dependent on whether they are residents of a single family house, a condominium or townhouse, or an apartment.
- b) A rewording of the screener questions for those respondents who reported being afraid or angry about certain situations (e.g., "You mentioned that you were afraid because you thought someone might take something from your home . . . Since _____ .").
- c) The reminder to the interviewer of a short description of a particular incident (e.g., "Automobile stolen on April 16th") so that they can begin the incident report with the focal incident clearly in mind.

2.5. Testing of Response Error Hypotheses

Significant changes to the NCS questionnaire have been made during this research. These changes have been guided by previous observations concerning possible causes of underreporting and telescoping in the NCS data. Many of the revisions we have noted above are the result of conceptual and methodological contributions of redesign consortium members, and the questionnaire may be seen as a collective product of the consortium. In order to test these changes in the instrument a record check study is planned using a sample of victims identified on police reports in the Peoria, Illinois city police department. The survey will assign to random half-samples a revised version of the questionnaire or the current NCS

form. The criteria for evaluation of the alternative forms will be 1) the rate of matches between survey reports and the police reports among the victim sample and 2) the rate of overall reporting among those respondents assigned the two forms (this follows the argument that failure to report is the major response error).

3.0 Sample Design Alternatives for the National Crime Survey

A second set of activities being conducted at the Survey Research Center concentrates on improving the precision of NCS estimates through changes in the sample design. To date this work has examined previous investigations into stratification and the rotating panel design in order to form suggestions for future changes in design and offer design alternatives to the current NCS. The discussion in this paper addresses stratification, the rotating panel design, double or two phase sampling, and multiple frame designs. Each section briefly sketches the major issues and discusses possible changes in the sample design.

3.1. Stratification in the NCS.

There are two alterations of stratification features of the NCS design which may provide more precision for NCS estimates per unit cost of data collection. First, the characteristics used to stratify sampling units in the current NCS design were not for the study of criminal victimization but rather for the purposes of the Current Population Survey. These general purpose variables were used for stratification at a primary as well as a secondary stage of selection. Second, the allocation of the sample across strata needs to be examined more closely in the redesign. Currently the allocation is proportional selecting observations at the same rate across all strata at both primary and secondary levels of selection. Results from the NCS indicate that there may be some gains to the NCS estimates if a disproportionate allocation were implemented.

3.1.1 Stratifying Variables.

The primary sampling units in the NCS consist of the approximately 1,900 primary sampling units of the 1970 Current Population Survey. Characteristics that were used in the stratification were Standard Metropolitan Statistical Area status (i.e. SMSA and nonSMSA), rate of population change, proportion of population in urban areas, proportion of population employed in manufacturing, principal industries, average per capita retail trade and proportion of nonwhite population. These characteristics were chosen for improving the precision of employment estimates in the Current Population Survey; they are not necessarily the most appropriate for a victimization survey.

The size of gains that might be achieved through alternate stratification variables in the NCS is currently being examined by the Census Bureau. At present we can only note that for the current survey design a larger than

expected proportion of the total variance of estimates for the NCS is attributed to between primary sampling units within collapsed strata. Part of the reason for this larger than expected contribution is the method of collapsing strata for variance estimation purposes; some share of the larger contribution may also be attributable to ineffective stratification at the primary stage.

Other variables more highly correlated with crime or victimization are available for the primary sampling units from Population Census materials, Uniform Crime Reports, and other sources. For example, measures of the transiency of the population, the proportion of population in public housing, or the proportion of households with children ages 12 to 18 can be expected to be more highly correlated with victimization experience than those variables currently being used. Similarly, the Uniform Crime Reports (UCR) on police reported criminal incidents, available in the County and City Data Book (U. S. Bureau of the Census, 1978) and elsewhere by county or county-like units, may be a source of effective stratification data.

Some concern has been expressed that because the UCR data are based on police crime reports, they may not accurately reflect the actual victimization experience in a given area. To the extent that the Uniform Crime Reports are incorrect, due to poor or incorrect counts or due to counts by location of events rather than residence of victims, the use of UCR data will reduce the effectiveness of the stratification. However, stratification of primary sampling units by UCR data will not bias NCS results even though the UCR data may be inaccurate themselves.

Second stage sampling units in the NCS consist of Enumeration Districts (ED's) within NCS primary areas. For stratification, ED's were grouped explicitly within the primary areas into strata; implicit stratification was used by sorting ED's into a geographic order for systematic selection. The explicit stratification of ED's is on the basis of type of place, or CBUR code:

C: Central city of an SMSA.
B: Urbanized area in an SMSA not in C.
U: Urbanized areas not in C or B.
R: All other places.

Alternative schemes for stratification of ED's possibly more appropriate for the NCS, and similar to those for primary units, can be suggested. It is unlikely that UCR data will be available for ED's, but other characteristics may be considered in addition to or as substitutes for the CBUR strata and geographic ordering presence of public housing projects, median income, or the proportion of single person households. Certainly one would expect that some ED's will be in "high crime" areas within the NCS primary areas. Even if existing data are not adequate to identify these areas, some subjective attempts to identify "high crime" ED's and place them in a separate stratum may yield gains in precision for NCS estimates as well.

3.1.2. Allocation to Strata

The usual model to determine the allocation of the sample across strata depends on the amount of variability within strata for a characteristic of interest and the cost within strata of data collection. The optimal allocation requires larger samples from strata with larger variance or with smaller data collection costs. Generally, large differences among strata in the variance or the cost of data collection are necessary before optimal allocation results in sizeable gains in precision. On the other hand, the optimal allocation is actually somewhat insensitive to small deviations of the sample sizes from the optimal values.

Although large gains in precision are possible under an optimal allocation for one characteristic, serious losses may also be experienced for other characteristics and estimators of interest. For example, an optimal allocation for estimating burglary can lead to much smaller levels of sampling error for that type of crime, but for assault, which may require a different optimal allocation, losses in precision, even compared to precision achieved in simple random sampling, may be experienced. Some caution is thus needed in applying optimal allocation strategies to multipurpose surveys. Compromise allocations may be found among several variables of interest that result in retaining part but not all of the gains from an optimal allocation. Finally, gains for the total sample estimates achieved under an optimal allocation cannot be expected for subclasses of the sample; the smaller the subclass the fewer the gains that are retained for estimates for the subclass.

3.2. NCS Rotating Panel Design.

Essentially four "samples" or panels of approximately 72,000 housing units each were designated for use of the NCS during the course of the 1970's and early 1980's. Each panel is designed to be used for a three year period during which interviews would be conducted at each occupied housing unit seven times, or once every six months. Each panel of 72,000 housing units is divided into six rotation groups. Every six months one of these rotation groups is removed from the survey and replaced by a new rotation group from the next panel. Over the course of a three year period, a single panel enters the survey gradually, replacing old panel rotation groups one at a time.

Each rotation group is further divided into six "rotation group panels." Each month during a six month period, one of the rotation group panels is assigned to interviewers for data collection so that by the end of a six month period each occupied housing unit within a rotation group has been interviewed once. Thus, every six months 36 rotation group panels comprising approximately 2,000 housing units each (i.e., approximately 72,000 housing units are assigned for interviewing in the NCS).

The rotation scheme is complicated by a feature called bounding, a device designed to control the telescoping of victimization events

forward from one reference period into the next. The first of the seven visits to a housing unit is a bounding interview in which data are collected but do not contribute to NCS estimates. The bounding or first interview data serve only as a reference for the second interview to prevent respondents from reporting events that occurred prior to the second reference period as occurring during that reference period; bounding interview data do not contribute to NCS estimates. Six subsequent interviews are made to the housing unit at six month intervals.

At the same time as the seventh and final interview (i.e., bounding plus six interviews) for a rotation group panel, a new rotation group panel is being assigned for a bounding interview to replace and outgoing panel six months later. Each month there are actually seven rotation group panels being assigned: six rotation group panels being interviewed for the first through sixth bounded interviews, and one new bounding interview rotation group panel. Of the 42 total rotation group panels interviewed every six months, data from 36 are used to form NCS estimates.

At the time of interview, each respondent is asked about victimization events occurring during the previous six months. For a given reference month, data collection is not completed until six months later when the last rotation group panel is asked about event occurring up to six months previously.

Although the rotating panel design provides administrative convenience for balancing monthly interviewer workloads, the primary sampling design purpose is to afford overlap of sampling units between months, quarters, and years. In principle, overlap can achieve gains in precision for estimates of change in level or rates between time periods. However, the amount of overlap between any two time periods is complicated by a combination of the reference period and the rotation schedule.

In terms of sampling units or rotation group panels, the overlap can be specified fairly specifically. For any given reference year, there are forty-seven (47) distinct rotation group panels contributing reports. Between two adjacent reference years, thirty (30) of those 47 panels will be identical (although the contributions to reports for each year of the overlap panels are not necessarily the same), an overlap of 64%. To the degree that reports by occupied housing units in overlap panels are correlated across years, the precision of estimates of change will be improved. Between adjacent reference months, 35 of 36 panels reporting about those months overlap. Between adjacent reference quarters, the overlap is 35 and of 38 panels. Estimates of change for these reference periods should also have improved levels of precisions over independent samples. The amount of improvement is, however, difficult to examine because of the complex relationship between reference period and rotation schedule.

The importance of the panel design to the NCS extends beyond the issues of precision for measures of change between different time periods. Several nonsampling error issues are

affected by the NCS rotating panel design, and may be more important that the sampling issues for the accuracy of the NCS estimates. The following brief summaries indicate some of the issues and their importance to NCS estimates:

- a) There is increasing evidence that the six month reference period in the NCS is too long. The recall loss errors induced by the six month reference period may be so large that the sampling error measures only a small portion of the total error in NCS victimization estimates. A shorter reference period may lead to a smaller, but more accurate NCS. Further study of recall errors for victimization events is needed.
- b) Panel bias studies for the NCS indicate that for a six month reference period, and the current NCS rotation scheme, the optimum tenure for a panel member housing unit should be between four and five interviews - not the current six interviews. A redesigned NCS questionnaire, changes in the length of reference period, and other design modifications will have implications for the tenure of housing units in the NCS.
- c) Simultaneous consideration of shorter reference periods, panel bias, and telescoping bias (i.e., the incorrect recall of the time of occurrence of an event) is necessary. Changes in reference period or tenure or even in the need for a bounding interview cannot be made independently of the other design features.

3.3. Double Sampling.

Sometimes information needed for stratification is not readily available for all sampling units before selection. Further, the cost of measurement of these stratifying characteristics for each sampling unit is often relatively small compared to the cost of other measurements which the survey is designed to make. In such a case, a two phase or double sampling design is employed. In the first phase of the sample selection, a large sample is selected and certain relatively inexpensive measurements are made on all sample elements, including stratifying characteristics. In the second phase, the large sample is stratified with respect to these inexpensive measurements and a small stratified sample is selected from which the more expensive measurements will be made.

For the NCS it is difficult to find the data needed to stratify some sampling units, such as households or persons with respect to victimization experience prior to sample selection; one might view the bounding interview as an inexpensive measurement compared to the six panel interviews conducted for each sample address during the NCS. A two phase sampling scheme might stratify households for the six panel interviews by the victimization experience reported on the bounding interview; victim and

nonvictim segments households, or persons could be subsampled at different sampling rates, and only subsampled units followed during the next six interviews. If victimization experience over time is correlated for one or more of these units, there may be some improvement in precision for fixed survey costs, and there would be more victims in the survey to study.

The multipurpose nature of the NCS has serious implications for achieving optimal allocations and gains in precision for different types of estimates using a two phase design. For example, burglary may have different optimal first to second phase size allocations, as well as stratum allocations than would personal assaults. Sizeable gains in precision for some measures may be attenuated or lost altogether when considered in the context of a multipurpose NCS.

There are several other features of two phase sampling for the NCS that should be considered.

- a) The two phase design is more complicated to administer since first phase materials must be returned to be processed, analyzed, stratified, and subsampled, and then sent back for follow-up. The complexity extends to estimation and analysis as well. The increased cost of such complexity must be examined in terms of the potential gains in precision.
- b) It is not clear how estimates of variance should be calculated if the two phase design were implemented within second stage selections of the current NCS sample design.
- c) Stratification of first phase selections into potential victim and nonvictim groups based on the bounding interview probably will tend to lose its effectiveness after a few interviews. Other stratification schemes besides following victims and nonvictims may be better predictors of future victimization experience. The alternate stratification characteristics may be demographic, geographic or community features. Longitudinal information concerning victimization experience over the full seven interview series would be necessary to study the problem further.
- d) The particular sampling unit to be followed over time has not been specified. Segments of addresses, households, or persons could be stratified into victimization classes and followed. For different types of crimes, different units could achieve greater gains in precision. For instance, one could argue that for burglary following segments of addresses would produce greater gains in precision than following persons because of important geographic characteristics of burglary events.

Consider the following notation:

	Total number of elements	Sample mean
Frame F_1 , only	N_1	\bar{x}_1
Frame F_1 overlapping w. F_2	N_{12}	\bar{x}_{12}
Frame F_2 only	N_2	\bar{x}_2
Frame F_2 overlapping w. F_1	N_{21}	\bar{x}_{21}

It is not clear whether a two phase strategy would be improve the relative precision of NCS estimates. Preliminary consideration indicates that the gains achieved are small and would not lead to significant improvements in the precision of NCS estimates per unit cost. Further investigation of the issues in two phase sampling for NCS will require estimation of key parameters from existing NCS data.

3.4. Multiple Frame Designs.

A combination of frames is sometimes used to increase the cost efficiency of a sample design. A principal difficulty with multiple frame sampling is the overlap among frames - the simultaneous presence of sample elements in two or more frames. When a sample element has two or more listings in a single frame some weighting is necessary to avoid biasing the sample estimates. Similarly, sample elements in two or more frames in a multi-frame sample design can cause biased sample estimates. A variety of approaches can be taken to eliminate or adjust for the overlap problem, but basically only one approach will be reviewed here. In particular, the sample overlap can be post-stratified to create artificially unique listings by weighting overlap members to reflect membership in a sample from a particular frame. That is, those elements that are members of the same combination of frames are treated as members of the same post stratum.

Consider the case of two overlapping frames, F_1 and F_2 , with overlap portion between the frames denoted F_{12} . Let x_i denote the value of characteristic X for the i th element of a simple random sample selected independently from each of these two frames. Let f_1 denote the sample elements selected from F_1 only, f_{12} those from F_1 that also are in F_2 (the F_1 overlap), f_2 those from F_2 only and f_{21} those from F_2 that are also in F_1 (the F_2 overlap). Define new values for the sample elements as follows:

$$Y_{1i} = \begin{cases} x_i, & \text{if the } i\text{th sample element is} \\ & \text{in } f_1, \\ px_i, & \text{if the } i\text{th sample element is} \\ & \text{in } f_{12} \end{cases}$$

and

$$Y_{2i} = \begin{cases} x_i, & \text{if the } i\text{th sample element is} \\ & \text{in } f_2, \\ qx_i, & \text{if the } i\text{th sample element is} \\ & \text{in } f_{21}, \end{cases}$$

where $p + q = 1$. Thus, after the samples have been selected, those elements in the overlap (f_{12} or f_{21}) are arbitrarily weighted by the constants p and q to create non-overlapping "post-strata".

An estimate from the two frame sample of the total value of characteristic X in the population is the Hartley estimator (Hartley, 1962),

$$x_H = N_1 \bar{x}_1 + p N_{12} \bar{x}_{12} + q N_{21} \bar{x}_{21} + N_2 \bar{x}_2$$

The values of p and q are determined by minimizing the variance of x_H subject to fixed costs of data collection in the two frames. Note that the case where $p=0$ and $q=1$ (or $p=1$ and $q=0$) corresponds to a screening approach in which the overlap is eliminated by excluding sample elements from one of the two samples that are also contained in the other frame. The choice between a post-stratified estimator x_H , where p is not zero or one, and a screening method depends on the cost of obtaining information used to include (or exclude) sample elements because they are part of the overlap.

3.5. Joint Use of Telephone and Area Frames.

One particular formulation of this problem is of interest to the NCS - one of the two frames is completely contained within the other, and the smaller frame is considerably cheaper to use than the larger frame. A two frame approach is preferred to using the smaller and cheaper frame alone in order to improve the coverage of the smaller frame.

The current NCS sample is based on a household frame. A telephone sample would be contained within the current frame, but would not itself cover the population completely. Hence, two independent samples could be selected, one from the current household frame (F_1) and one from a telephone frame (F_2), and the estimates computed using an estimator such as x_H .

There are two comparisons of procedures and estimators that are of interest when one frame is contained in the other. First, it is necessary to determine whether the two frame estimator offers an improvement in precision for fixed costs over an estimator based on the single large frame. Second, if the two frame procedure is more precise, it is of interest to know whether the screening method can be used to improve precision for fixed cost as well.

Using rough estimates of relative costs, it appears that the use of a household-telephone two frame sample would result in sizeable gain in precision per unit cost over the current single frame design. The discussion of these gains is limited to a consideration of operating costs and does not consider the developmental costs necessary to introduce a telephone interviewing system, or the overhead costs

associated with maintaining survey operations for two separate data collection modes. Certainly a shared development and operation of such a system by several agencies would reduce the burden to any one agency alone; but those issues are outside our consideration in the present discussion.

Suppose a sample of households is to be selected from the current household frame (F_1) and an independent sample from the telephone frame (F_2). The Hartley estimator can then be expressed as

$$x_H = N_1 \bar{x}_1 + p N_{12} \bar{x}_{12} + q N_{21} \bar{x}_{21}$$

An optimum value of p is needed to minimize the variance of x_H . Consider the following set of assumptions:

- a) The proportion of households that have telephones is 0.93.
- b) The cost of completing a telephone interview is one-half the cost of a personal-visit household interview i.e., $c_2 = (1/2)c_1$
- c) The variance of characteristic X in the household frame and the telephone household frame is the same, i.e. $\text{Var}_1(X) = \text{Var}_2(X)$.

If there are to be n_1 household interviews and n_2 telephone interviews, denote the total variable cost of the survey as $C = c_1 n_1 + c_2 n_2$. For fixed variable costs, the value of p that provides minimum variance for x_H is $p = 0.26$ (and $q = 0.74$). Further, it can be shown that with this value of p , a 28% reduction in variance can be achieved for the dual frame design with a Hartley estimator over the estimator from household sample frame design alone. The reduction can be achieved by having a final sample composed of approximately 25% from the household frame and 75% from the telephone frame. Households with telephones in the household frame would be included in the estimate \bar{x}_{12} , and the weight $p = 0.26$ would be applied to each observation. All telephone frame sample elements are included in \bar{x}_{21} , with weight $q = 0.74$ applied to each.

Suppose that the cost of a household interview is $c_1 = 40$ and for a total variable cost of $C = 6.0 \times 10^6$, $n_1 = 150,000$ households are visited. For the same fixed costs and $c_2 = 20$, the dual frame design using the Hartley estimator can select 58,000 households from the household frame and sample 183,000 telephone households. The increased complexity of using this two frame procedure would probably not offset these gains. On the other hand, the same sample size of 150,000 households may be retained but allocated as $n_1 = 36,000$ households from the household frame and $n_2 = 114,000$ households from the telephone frame. The total cost in this case would be $C = 3.7 \times 10^6$, a 38% reduction. The savings could be invested in the development of a telephone interviewing system and other uses important to BJS. Note that such a telephone interviewing system could also be phased in gradually until the optimum

combination of telephone and household interviews is achieved.

Given that sizeable gains in precision can be achieved using a dual household-telephone frame design and the post-stratified Hartley estimator, it is of interest to determine whether further gains in precision for fixed cost can be achieved using a screening method, i.e., setting $p=0$ and $q=1$ in x_H by conducting full interviews in the household frame sample (F_1) only at nontelephone households. The relative gains of such a procedure depend on the relative costs of screening for telephone households in the household frame sample. The screening method would contact a sample of households from the household frame, but conduct full interviews only in those without telephones. Let c_s denote the cost of determining whether a household has a telephone in the household frame. The screening method would yield even greater gains in precision over the two-frame Hartley estimation approach only if $c_1 - c_s > c_2$. Since the major portion of the cost of household interviews is not due to interviewing, but arises from the effort to contact household residents, determining whether a household has a telephone will cost nearly as much as doing the complete interview, whether a telephone is present in the household or not.

If only a single interview is to be conducted, then c_s , the screening cost, will be nearly the same as c_1 , the full interview cost, and $c_1 - c_s < c_2$; the screening method would not be more precise than the dual frame Hartley estimator method for fixed costs.

The need for a longitudinal component in the NCS design introduces a complicating element for a screening method. If the panel consists of the same addresses of non-telephone households and the same telephone numbers followed over three years, the nature of the sample could be quite different at the end of the three years than at the beginning - when screening for non-telephone households was conducted. Non-telephone households may acquire telephones. Telephone households may move or have service disconnected. Further, it is not clear what the implications of longitudinal designs are for telephone samples. These issues require further investigation, in the context of a rotating panel design, before serious consideration can be given to designing and implementing a dual frame household telephone design for the NCS.

4. Conclusions

The National Crime Survey is an example of a large scale survey whose design has not yet been optimized within the resources available. This optimization requires careful attention to survey cost components, to the impact of the design on sampling variance, and to the causes of nonsampling errors. The work on the NCS redesign by the consortium of organizations is attempting to link these concerns into a coordinated effort to improve the NCS. The work described in this paper is speculative because no tests of the new design features have yet been conducted. As the work of the consortium

and the Survey Research Center continues, these tests will be performed.

It is likely that the questionnaire development activities will be accomplished through a series of pilot surveys with experimental variation in questionnaire form assigned to different subsamples. There will probably be simultaneous tests of a personal interview and a telephone interview version of the questionnaire; these tests may suggest that different versions of the questionnaire are needed in the two modes in order to minimize response errors. Some of the ideas tested in the instruments will be impractical or inefficient for use in a production version of the NCS. For this reason a final test of the recommended revised NCS instruments has been suggested before the consortium submits its final report.

The activities in terms of alterations of sample design do not as easily yield themselves to pretesting. Instead, models estimating the impact on total error through change in sampling design will be examined. Parameter estimates for these models will be supplied by reanalysis of existing data, examination of Census Bureau technical reports, and by utilization of data from surveys of a similar design. Whenever possible, the pilot tests will incorporate altered sample designs in order to make the sample design part of the experiment.

There are several design aspects both in questionnaire and sample design that have large impact on the nature of the administrative structure for the NCS. The use of a dual frame design demands a staff supervising both of the interviewing groups, groups that require different interpersonal skills and different training to be effective in their job. There are a variety of administrative formats that are

possible in such a design ranging from a single centralized telephone interviewing staff to several regional telephone interviewing staffs, to a complete dispersal of the telephone interviewing so that the same interviewers would conduct both telephone and personal interviews. These alternative administrative structures have implications on the magnitude of nonsampling errors and of the total cost of the NCS. Understanding the interrelationships of cost and error is one of the major goals of the NCS Redesign Consortium.

References

- Cannell, C.F., L. Oksenberg, and J. Converse (1977). "Striving for response accuracy: experiments in new interviewing techniques." Journal of Marketing Research.
- Hartley, H.O. (1962). "Multiple Frame Surveys," Proceedings of the Social Statistics Section of the American Statistical Association, pp. 203-206.
- Law Enforcement Assistance Administration (1972). San Jose Methods Test of Known Crime Victims. Statistics Technical Report No. 1. Washington, D.C.: U. S. Government Printing Office.
- U. S. Bureau of the Census (1978). County and City Data Book, 1977: A Statistical Abstract Supplement. Washington, D. C.: U. S. Government Printing Office