# Effects of Unbounded Interviews, Time in Sample, and Recency on Reported Crimes in the National Crime Victimization Survey

Robert E. Fay[1], Jianzhu Li[1]
[1]Westat, Inc., 1600 Research Boulevard, Rockville, MD 20850

**Abstract**
The overall design of the National Crime Victimization Survey (NCVS) has been largely stable for over 30 years. Households in sampled housing units are interviewed for 7 waves, each time collecting data for the previous 6 months. Until recently, the 1st wave has been omitted from the published estimates to exclude reports outside of the intended 6-month reference period. Using publicly available data for 1998-2004, we report on a series of analyses to investigate more current effects of the bounding and find evidence of more general time-in-sample effects. We will also report on the effect of recency in the observed incident reports, where more crimes are reported in the 1st month preceding the interview date than each of the previous 5 months. These phenomena are important in considering a range of design options for the NCVS that would alter the reference period or panel design.

**Key Words:** Response error, recall error, telescoping

## 1. Introduction

Since 1972, the Census Bureau has conducted the National Crime Victimization Survey (NCVS) for the Bureau of Justice Statistics to collect data on the frequency and consequences of crime from the victims of crime. Because many crimes are not reported to the police, the survey provides valuable information on the incidence of crime that would otherwise be unavailable.

From the beginning, when the survey was originally known as the National Crime Survey (NCS), the sample design has retained a set of basic features. The NCS/NCVS is a panel survey of housing units retained in sample for a total of 3 years. Every 6 months, the households occupying the sampled housing units are interviewed to collect information on the occurrence of crime during the preceding 6-month reference period, for a total of 7 waves of interviews or attempted interviews.

The survey design has been shaped by a concern over respondents' ability to accurately recall incidents of crime over an extended reporting period, such as the 6-month period in the current design. Because respondents might recall and report a crime as occurring within the 6-month period when it occurred earlier—an error termed *forward telescoping* in the survey research literature—until recently the NCVS completely excluded the first wave interviews from the published reports. Instead, the first wave interview was used to *bound* the second wave interview, that is, to collect information on crime during the first

.

6-month period so that any incidents reported again in the second wave could be identified and excluded by the interviewer. In other words, the basic strategy is to exclude the unbounded first interview so that the remaining 6 interviews will be bounded. Because of household turnover and other factors, however, the bounding in NCVS has remained imperfect, and the official estimates have always included some unbounded interviews (Rennison and Rand, 2007).

Few surveys adopt the approach of collecting and then abandoning so much data of direct interest to the primary goal of the survey. Empirically, the NCVS data collected during the first wave has implied higher crime rates than the published estimates based on waves 2-7. This evidence is consistent with the standard interpretation that the first wave interview is necessary to bound reporting in subsequent waves in order to exclude forward telescoping. But alternative explanations to be considered here could account for much or all of the difference in the wave 1 rates. Beginning with the 2006 estimates, BJS has begun to include the wave 1 results in the published estimates (Rand and Catalano, 2007), although by reducing the weights on victimization incidents to adjust the wave 1 data to the same level as other waves. How the data from wave 1 should be used in estimation remains an important question.

Because of budget constraints and the increased cost of interviewing, the NCVS sample size has declined by approximately 50% over its history, leaving the survey less sensitive to changes in the crime rate. A panel of the National Academy of Sciences (National Research Council 2008) recommended a thorough review of available design options for the NCVS, including consideration of a 12-month instead of 6-month reference period and alternatives to the 3-year panel design. The research to be reported here addresses part of these objectives by examining evidence from the NCVS on the effect of altering aspects of this basic design.

We begin by reviewing a classic paper on telescoping, Neter and Waksberg (1964b). Although the paper has enjoyed frequent citations, it is rarely summarized in detail in most of the subsequent literature. Because the paper provides an essentially complete context for the methodological issues raised here, we begin with a section providing an extended summary.

We will examine separately two forms of evidence from NCVS. Before doing so, we summarized statistical methods used in both studies. We then report on a study of the relationship between the month of interview and the month of occurrence that respondents report for crimes occurring during the 6-month reference period. Reported crime rates are highest in the month just before the interview. (Most NCVS interviews occur during the first 2 weeks of the month, and respondents are asked to report crimes occurring during the previous 6 full months; e.g., July through December for an interview in January.) Because reported crime is higher in the most recent month, the phenomenon has been occasionally called *recency bias* in the survey literature. Beyond this general pattern, we detect variation in the degree of recency bias according to whether the crime was reported to the police and, to some extent, the wave of interviewing.

Then we report on a longitudinal analysis of the NCVS data to measure time-in-sample effects. Unlike the first study, which analyzes differences within wave, this study examines differences between wave. Not only does the crime rate from first wave interview differ from the others, but the average crime rate detected in the second wave interview differs systematically from the third, etc. We argue that this more general time-

.

in-sample effect cannot be readily separated from forward telescoping to account for the differences between the first and second waves.

The two studies examine the same source of data from two perspectives: The first examines wave effects without accounting for recency bias, while the second focuses on recency bias with reduced attention to accounting for the wave effects. We discuss their joint implications in the concluding section.

## 2. The Survey of Residential Alternations and Repairs

Neter and Waksberg (1963, 1964a, 1964b) published three papers based on a 1960-1961 experiment conducted by the Census Bureau to study reporting effects in a consumer survey of resident home owners on their expenditures on alterations and repairs. Each of the papers analyzed a different aspect of the study. Neter and Waksberg (1963) examined whether designating a preferred household respondent (husband, wife, husband and wife together, vs. unspecified respondent) affected the survey outcome. (Note: we have replaced their term "heads" with "husbands.") The design also allowed them to test whether any conditioning of responses occurred, that is whether systematic differences could be detected between a first and second interview conducted under identical conditions (Neter and Waksberg 1964a). Finally, they analyzed the effect of telescoping, comparing results of unbounded and bounded interviewing (Neter and Waksberg 1964b). By far, the last paper is the most cited of the three. (In September, 2010, Google Scholar listed 160 citations compared to 2 for the first and 6 for the second.) By providing carefully documented experimental evidence for telescoping, Neter and Waksberg (1964b) arguably wrote the classic paper on the subject. But the three papers taken together account for the complexity of their careful experimental design and complete some arguments that were somewhat unclear in Neter and Waksberg (1963b).

The experimental design was based on a rotating panel survey of the U.S. population of eligible homeowners of 1-4 housing unit properties. Each sampled housing unit was interviewed 4 times, 3 interviews 1 month apart and a fourth following 3 months later. Each month, 2 panels were started. In one respondents were asked in their first interview to report on expenditures for the last 6 months and in the other to report on expenditures for the last month. More precisely, the interviews were conducted in the first part of a month to report on the immediately preceding whole months. The first interviews were consequently unbounded. The second and third interviews required 1-month bounded recall, and the last interview required 3-month bounded recall. For 6-month unbounded and 3-month bounded recall, the interviewer first asked about all events in the period, then asked the respondents to allocate them to specific months.

The target size of the segment (ultimate sampling unit) was 4 eligible households, and households were randomly assigned to the targeted respondent groups. The first 2 panels were started in February, 1960. One panel was asked about expenditures in January, 1960 and the other for August 1959 – January 1960. Two panels started each month thereafter for a total of 30 panels.

In analyzing the data, all three papers noted unfortunately large standard errors for total expenditures. Instead, they analyzed the number of jobs broken by job expenditure, classifying jobs under $20 as small and those of $100 or more as large. In some cases

.

(e.g., Table 1, p. 53 in Neter and Waksberg 1964a) they divided the data into finer categories.

To control for possible variation in expenditure by month, they formed comparisons of different panels across the same set of months. For example, from March 1960 until the end of the study in March 1961 (collected in April 1961 interviews) they could compare the unbounded 1-month recall with the average of the bounded 1-month recalls from the second and third interviews.

Neter and Waksberg (1964a) began their paper by reviewing previous studies of conditioning in expenditure surveys but found no consistent pattern in the results. They noted that many of the previous studies confounded conditioning with other effects, such as period effects. To study conditioning in the 1960 Census Bureau study, they primarily focused on the comparison of the second and third interviews, which were conducted under identical conditions of 1-month bounded recall, except for their difference for time in sample. They found an overall conditioning effect amounting to an approximate 9% drop in reported jobs overall. They reported that the drop for jobs under $10 was statistically significant, and that drops appeared for most other job classes that were not individually significant. They concluded “From the combined evidence from all the data, it appears that there was a definite conditioning effect for the smallest jobs, and that probably there were conditioning effects also for some of the other job-size classes.”

They also reported, based on pair-wise tests using a 2-standard error criterion, that conditioning was significant for the procedure of designating wife as the respondent and for the unspecified group, but not for the husband procedure and the husband and wife procedure. Taking this approach to testing, they concluded that the conditioning effect was probably restricted to the first two groups. But the large standard errors (e.g., Table IV in Neter and Waksberg 1963, p. 51) admit several alternative interpretations, including that the conditioning effect was uniform but that there was a small main effect of respondent designation procedure instead.

Neter and Waksberg (1963) analyzed the effect of the respondent designation procedure across the experimental conditions. As with conditioning, their review of past expenditure studies did not surface a consistent pattern of findings. In their own study, they found almost no effect on the reported data, except for their interpretation that conditioning effects appeared in the third interview for wives and for the undesignated group.

Neter and Waksberg (1964b) primarily addressed two issues: telescoping and the length of the recall period. They remarked that telescoping had often been mentioned in previous studies but rarely studied systematically, and that effects of the length of the recall period had typically been confounded with other effects in previous studies.

They described the bounding procedure in the study in the following way (pp. 19-20):

> At the beginning of a bounded interview, which is a second or later interview with a household, the interviewer tells the respondent the expenditures which had been reported during the previous interview, and then asks for the additional expenditures made since then. During and after this interview, the interviewer checks to make sure that no duplication of expenditures occurs. This utilization of earlier expenditures reports also

.

serves to avoid any tendency on the part of respondents to think that recent expenditures were made prior to the current reporting period.

This procedure incorporates measures both to exclude previously reported cases and, in a partially offsetting manner, to assure respondents that recently occurring incidents were not previously reported.

As a measure of telescoping, they compared the 1-month unbounded first interviews to the 1-month bounded second and third interviews. They recognized that the evidence of conditioning affecting differences between the second and third interview raised the possibility that conditioning could also play some role in differences between the first and second. In other words, the effect of telescoping and conditioning was confounded in their study. They proposed some bounds on the effect of conditioning, based on the observed patterns in the data. Their arguments are too involved to summarize here, but they are reasonably convincing that a telescoping effect occurred, particularly for large jobs.

They also studied the effects of internal telescoping within 3-month bounded recall, where jobs were disproportionately reported in the month just prior to the interview than the previous two months, relative to controls based on 1-month recall. Telescoping was particularly pronounced for large jobs. They found that 3-month recall resulted in very little net loss for large jobs compared to the findings from 1-month recall, but that there were statistically significant losses of small jobs with 3-month recall. For smaller jobs, the losses were more substantial. In hindsight, the consequences of internal telescoping resemble forgetting curves (e.g., Rubin and Wetzel, 1996), and the results for small jobs were likely more driven by forgetting than internal telescoping.

Their analysis also used the data from unbounded 6-month recall to study the effect of reporting load—the load on memory required to recall the larger number of jobs over a 6-month interval than a 1-month interval.

The design also permitted them to study the effect of recall length by comparing the bounded 3-month recall to the bounded 1-month recall interviews. They noted that, without a source for external validation of the reports, the experimental comparison only measured the combined effect of forgetting and cognitive load, that is, the additional difficulty experienced by respondents trying to report on a larger number of occurrences. Their analysis again had to consider the possible effects of conditioning on the comparison.

In their discussion section, the researchers concluded that the substantial telescoping effect established for large expenditures, combined with a conditioning effect apparently pronounced only for small jobs, made the choice of a rotating panel design attractive. They recommended that the design use the first interview solely for bounding purposes and not estimation. This strategy has been the one adopted by the NCVS until recently. Neter and Waksberg noted, however, that a different balance between the sizes of the telescoping and conditioning effects might favour a different design or estimation approach, particularly if the cumulative effects of conditioning were large. Additionally, for the goal of measuring total expenditures, they suggested that the 3-month recall appeared sufficiently accurate to justify its use in place of 1-month recall.

.

The Academy report (National Research Council, 2008) highlighted the importance of telescoping in considering any design modification to the NCVS. By comparison, the report mentioned possible time-in-sample effects only once (Table 4-1, p. 84) and did not use the term *conditioning* at all.

The analyses presented in the next sections examine a parallel set of concerns, but uses different terminology from Neter and Waksberg's choices. The term *recency* is analogous to their *internal telescoping* and forgetting, and our use of *time-in-sample* refers to the combined effect of *external telescoping* (that is, telescoping between interview waves, the more typically studied form of telescoping) and *conditioning.*

## 3. Methods

### 3.1 The NCVS Design

As noted in the introduction, the NCVS has a rotating panel design. The ultimate sampling unit is a segment, generally of size 4, of housing units. The designated mode for the first interview of a household is personal visit, but subsequent interviews are now largely conducted by telephone if the household agrees. New households moving into sampled housing units are targeted for an initial personal visit, and their unbounded data have always been included in the estimates.

The NCVS universe comprises the household population age 12 and over. The survey uses self-response to collect data on personal crime. Most personal crime is categorized as violent crime—rape, robbery, aggravated assault, and simple assault—but the much smaller category of personal theft—purse snatching and pocket picking—is also included. We restrict our analysis to violent crime. An adult member of the household serves as the household respondent to answer questions about property crimes, including burglary, auto theft, and other theft.

Both the household-respondent and self-response questionnaires are divided into (1) a screening phase designed to identify the crimes or potential crimes in scope for the survey, and (2) an incident phase to record details for the individual incidents. Except for special provisions for series crimes involving multiple incidents of the same crime, respondents were asked the incident questions for each crime or potential crime identified in the screening phase.

### 3.2 Data Source

All of the data sets used in the analyses were obtained from the National Archive of Criminal Justice Data (NACJD), which is housed by Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan. The data are publicly available to researchers agreeing to use the data solely for research.

We used the annual bounded data sets for waves 2-7 with the collection year weights (the terminology employed in the documentation). These weights can be used to reproduce results in BJS's primary publications, including the Criminal Victimization series. We used years 1996-2005 for the recency study. For the time-in-sample study, we used a smaller set of years 1998-2004 that permitted a longitudinal file to be constructed on the basis of the scrambled unit identifiers and that was unaffected by any changes in the redesign of the sample following the 1990 census.

.

For the recency study, we also used the unbounded data sets containing data for waves 1-7 in order to obtain data for wave 1. The files were only available from NACJD for 1999-2004.

### 3.3 Statistical Analysis

The NACJD files contain two codes provided to facilitate design-based variance estimation. A pseudostratum code divides the data set into groups that can be treated as strata for the purposes of variance estimation. Each pseudostratum is divided into two half-samples by a variable called "secucode." The Census Bureau provided the assignments based on their internal knowledge of the design, and codes were assigned across time to fully reflect the aspects of the rotating panel design. In particular, an assignment of a sample household to a pair of codes was maintained consistently across years during the period we analyzed longitudinally, 1998-2004. Our analyses were performed in SAS®, with PROC SURVEYMEANS, PROC SURVEYLOGISTIC, and related procedures. Although our graphs do not display the standard errors, we checked that the principal findings suggested by them are statistically significant.

## 4. Recency

Most NCVS interviews occur within the first 2 weeks of the month, and respondents are asked about crimes occurring during the preceding 6 months. Averaged over time, accurate reporting should produce monthly estimates each about 1/6 of the total for the period; yet, analogous to Neter and Waksberg's findings, the proportion is particularly higher in the immediately preceding month. Figure 1 shows this pattern for total crime, violent crime, and property crime. Note that the recency effect is greater for violent crime than property crime, and that a small but statistically significant upswing in reported violent crime appears between the fifth and sixth months.

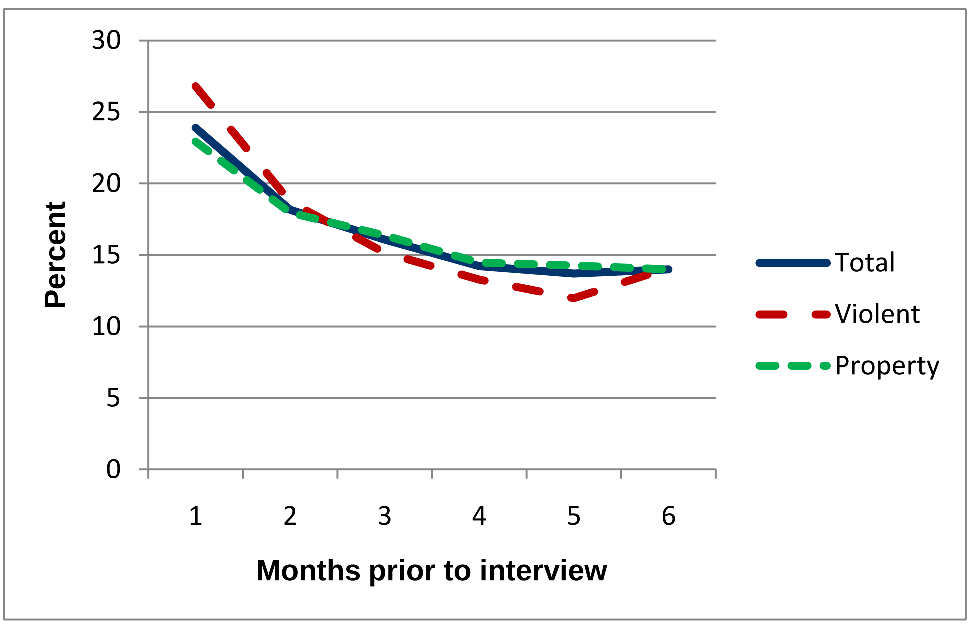


**Figure 1:** Percent distribution of crime by months prior to interview, NCVS 1996-2005.

.

As part of the incident phase, NCVS asks respondents whether the crime was reported to police. Overall, respondents report that only roughly half are, although rates vary by the type of crime. (Auto theft is the most consistently reported.) Figures 2 and 3 show the monthly distribution classified by whether the respondents stated that the crime was reported. In both cases, the crimes reported to police show a partial lessening of the recency effect. In contrast, respondents place almost 30% of all violent crimes not reported to the police in the immediately preceding month.

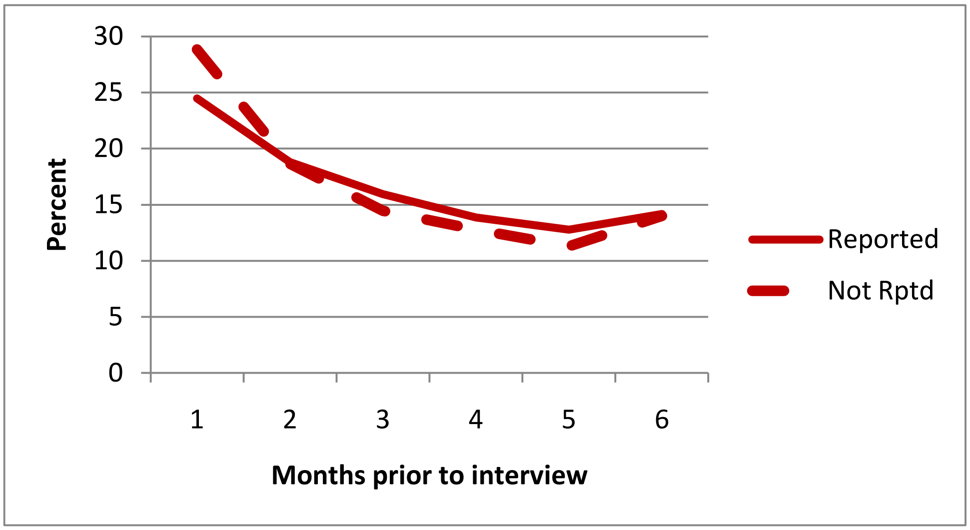


**Figure 2:** Percent distribution of violent crime by whether crime was reported to the police, NCVS 1996-2005.

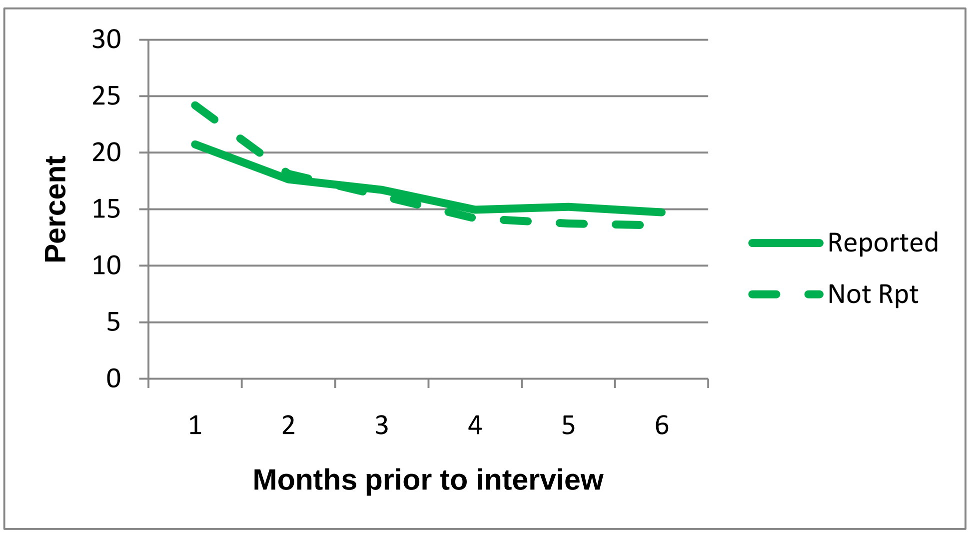


**Figure 3:** Percent distribution of property crime by whether crime was reported to the police, NCVS 1996-2005.

Using the unbounded data set for 1994-2004, we compared the reported distribution by wave. The reduced number of years and division of the data into waves increased the standard errors of the proportions, and the results include some visible wiggle compared to the preceding graphs. Figures 4 and 5, for total crime and total crime reported to police, compare wave 1 results to 3 of the 6 remaining waves, to avoid excessive clutter in the display. Over waves 2-7 there is barely any progression, but wave 1 is distinctly different, showing a flattened distribution.

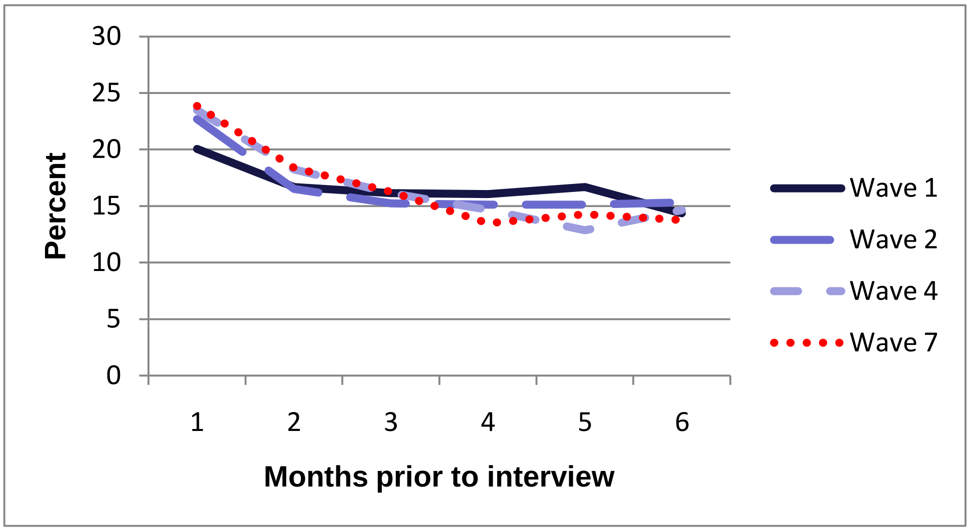


**Figure 4:** Percent distribution of total crime by wave, NCVS 1999-2004. Waves 2, 4, and 7 are presented, but waves 3, 5, and 6 overlap with them, leaving wave 1 the most distinct of the set.

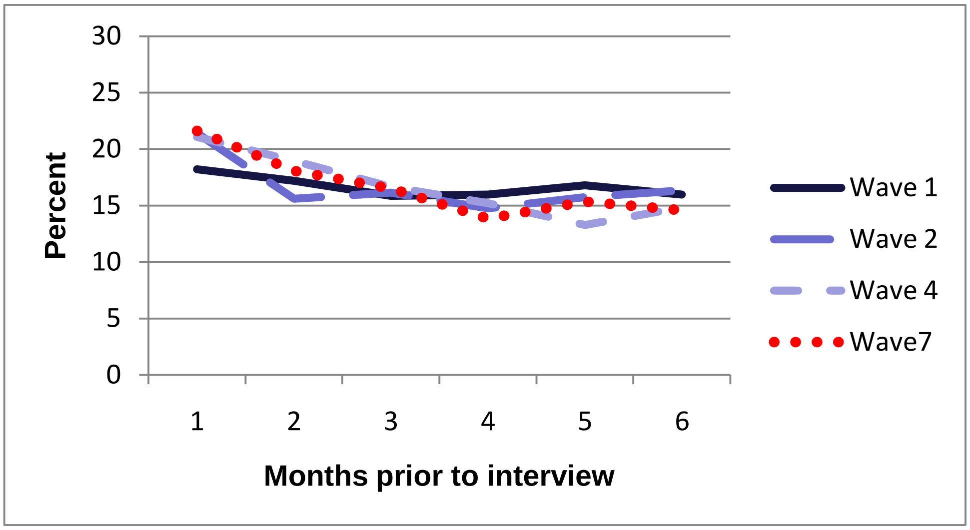


**Figure 5:** Percent distribution of total crime reported to police by wave, NCVS 1999-2004. As in Figure 4, waves 2, 4, and 7 are presented, but waves 3, 5, and 6 overlap with them, again leaving wave 1 the most distinct of the set.

We will return to these results in the discussion section but a few comments are in order at this point. When the NCVS is averaged over this group of years, the resulting large data set supports some general observations. Researchers might reasonably assume that violent crimes are associated with higher affect (in the psychological sense) on average than property crimes, yet the recency effect is somewhat more pronounced. On the other

.

hand, crimes reported to the police are presumably on average more important to the victim, yet the recency effect for these crimes is less, not more, than for crimes not reported. These are not, perhaps, surprising findings, but they illustrate that the accuracy of reports is not tied in a simple way to the importance of the crime, but that other factors may influence the recency effect. For example, the additional elaboration of details involved in reporting a crime to police, which may result in some followup over time, may help to bind a detail such as the date of occurrence in memory. We will return to issues of memory in the discussion section.

## 5. Time-in-Sample

As previously noted, we constructed a longitudinal data set based on matching NAJCD data on the basis of scrambled control numbers across 1998-2004, using the bounded data set containing waves 2-7. These years were selected because 1998 was the first full year of the 1990 census-based design for NCVS. In constructing the longitudinal data set, we wish to avoid complexities from the deletion of 1980 census-based sample units. Between 1998 and 2004, the system of scrambling control numbers was stable, so that housing units could be matched across years. In 2005, a new system was introduced.

Codes on the files indicate whether the household was new in that wave. In wave 2, it is possible to determine whether the household was also interviewed in wave 1, even though the wave 1 data were not included on the files we analyzed.

A preliminary analysis showed that the relative number of waves that a household had occupied the housing unit since its first interview was more important a predictor of the level of reported crime than was the number of times the housing unit had been in sample. We refer to this relative numbering as the *relative wave* or *relwave* of the household.

We matched households according to household number and other identifiers on the files. We did not incorporate person-level matching to establish a longitudinal file of persons. We divided households into three groups defined longitudinally:

1. Group 1, the same initial household was interviewed in all 7 waves. By definition, the data file contains information on waves 2-7, but no relwave 1 interview would be available.
2. Group 2, other initial households interviewed fewer than 7 times. If the first interview of the household occurred after the first wave, then an unbounded relwave 1 interview would be available.
3. Group 3, replacement households at the housing unit, identified by a household number of 2 or higher on the file. The first interview of these households is always unbounded and available for analysis.

Having defined the longitudinal groups and established the relwave of each household at the respective interview waves, we analyzed an indicator of whether the household reported any crime during the 6-month period with logistic regression. Households in group 1 contributed 6 observations each to the logistic regression, and households in groups 2 and 3 usually contributed more than 1 but typically than 6. (A group 2 household that was not interviewed in the first wave but in all subsequent waves could contribute 6 observations.) The use of design-based analytic tools (specifically, PROC SURVEYLOGISTIC in SAS) and the consistent assignment of the variance codes across

.

years properly captured the effect on the variance of any dependence between these observations.

The coefficients of relwave in the logistic fit to households experiencing crime are given in Figure 6. The constant term and terms for each year (1998, 1999, etc.) were also included. Additional models also incorporating terms for household composition and for mode did not substantially alter the results shown in Figure 6.

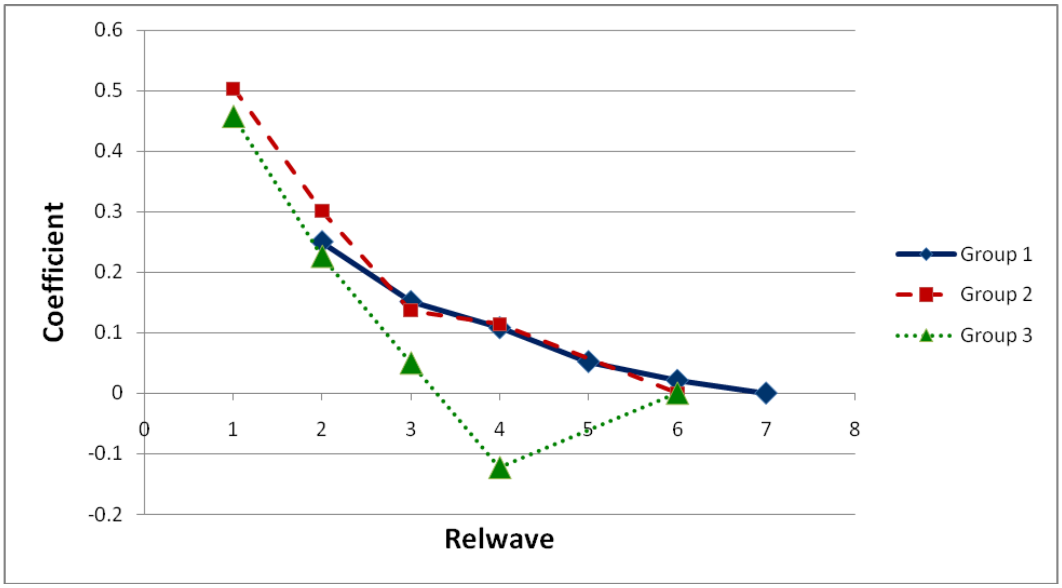


**Figure 6:** Logistic regression coefficients for relwave by group for the prediction of total crime. Group 1, households interviewed in all 7 waves while they were in the panel, is the largest group, but their reporting for wave 1 was unavailable on the files used in our analysis, so there is no basis to estimate a coefficient for wave 1. Wave 7 is the reference category for Group 1. Groups 2 and particularly 3 are smaller, and the reference group was set the combination of waves 5, 6, and 7, and this is shown by plotting a coefficient of 0 at relwave 6 for these two groups. The coefficient for group 3 at relwave 4 is not different from 0 by a statistically significant amount.

Although the only results comparing relwave 1 and 2 are available from groups 2 and 3, the data are consistent with the largest drop between adjacent waves occurring at this point. But the drop from relwave 2 to relwave 3, consistent across all 3 groups, is also significant. In the abstract, the situation here is similar to the one encountered by Neter and Waksberg, but the evidence for a significant time-in-sample effect separating relwave 2 from relwave 3, in particular, and from the average of the remaining waves. This consistent evidence of conditioning is stronger and more consistent than the findings for the consumer expenditure survey. Applying the logic that Neter and Waksberg used previously, the observed differences between wave 1 and 2 confounds a possible telescoping effect with a conditioning effect, whose magnitude is roughly suggested by the wave 2 to wave 3 differences.

Before the general discussion, we note that a possible extension of this work would be to match the longitudinal wave 2-7 data to wave 1 data from the unbounded file, where possible. By way of explanation, we did not undertake this effort initially because it would reduce the number of available observations. It was also unclear how to match the files, because the lengths of their scrambled identifiers differed. We subsequently

.

determined how to make this match, and acknowledge that this work could contribute to analyzing the differences between relwave 1 and relwave 2, particularly for group 1.

# 6. Discussion

## 6.1 Implications of the Findings

The NRC report recommended that BJS consider the effect of changing from a 6-month to a 12-month recall period. The research reported here does not provide a decisive answer to this question, but we believe that the findings provide relevant information. The recency curves shown in Figures 1-3 suggest a gradual falling off in reporting after the first two or three most recent months. The results suggest that a 12-month recall period would result in somewhat decreased reporting, with the impact probably somewhat greater for crimes not reported to police and, it appears, for violent crime. Stated another way, the data do not exclude a 12-month recall period as a possibility. The next step might be to conduct a designed experiment, such as the one reported by Bushery (1981), that would compare recall periods experimentally, imbedded in the NCVS itself. Even if the experiment showed a modest decrease in crime for 12-month compared to 6-month recall (a finding similar to his), high data collection costs might justify 12-month recall. In the same way, his earlier findings showed some loss from a 6-month recall compared to 3-month, but the choice of 6-month recall was seen at the time as a practical necessity.

The evidence presented here would encourage a re-examination of the issue of telescoping and the role of the bounding interview. The time-in-sample results challenge the notion that it is possible, on the basis of the NCVS data alone, to establish a method to determine the true level of crime on an absolute basis. The prevailing assumption for many years was that the use of the first interview as a bounding interview assured the meaningfulness of the resulting estimates formed from the remaining 6 waves. But even under this assumption, the results show that the average of the remaining 6 waves would be less than the expected value from the second wave alone, which is distinctly higher than the others. Following Neter and Waksberg's recognition that, in their terms, telescoping and conditioning could be confounded in their contribution to differences between waves 1 and 2, we suggest that the logic leads us to look for a point somewhere between the expected wave 1 and wave 2 results as arguably the point at which biases from telescoping and conditioning are offset. To be clear, however, we recognize that BJS certainly has good reasons to maintain the *status quo* for now, until other changes in the design are implemented in the future.

## 6.2 Extending the Literature Review

In place of a standard literature review, we took the more unusual step of summarizing a single milestone study. The study anticipated the major issues that resurfaced in our study of the NCVS data, even though some of our findings and interpretations differ from theirs. The subject of telescoping, time-in-sample effects, and their interaction deserves a more extensive review, and we will provide a sketch of what might emerge from additional effort.

The crime survey has benefited from a history of vigorous methodological research. The 1980s were particularly a golden era, when the survey was the original focal point for the initial attempts at collaboration and interdisciplinary discussion with cognitive psychology. The research, conducted both inside and outside of the Census Bureau,

.

resulted in a major redesign of the questionnaire, fielded in 1992 (National Research Council, 2008; Rennison and Rand, 2007). The redesigned questionnaire increased reporting for most types of crime, although by varying amounts, and created a notable break in the series. The strikingly different performance of the new questionnaire raises questions on whether past research, based on the old questionnaire, continues to apply under the new questionnaire. We consequently focused our research on the performance of the NCVS under the new questionnaire, but we recognize the value in reviewing parallels in studies conducted for the NCS under the old questionnaire.

A specific challenge in reviewing the literature is the apparent disagreement on what a bounding interview should do. The previously quoted description of bounding in Neter and Waksberg (1964b)—where the interviewer reviews items from the previous interview with the respondent before asking about the current period—is different from the procedure implemented by the NCVS, where the interviewer is solely responsible for intercepting duplicated reports. We are unaware of any argument why these two different procedures should produce the same result (and would be somewhat surprised to find such an argument), but we are also unaware of any study comparing the two procedures.

In a previous paper, one of us (Fay, 2009), argued that survey research had fallen behind relevant developments in basic science, using research on memory as the example. But catching up does not appear to be a simple task, as this application illustrates. For a time, interdisciplinary work between survey research and the basic science research on memory received heightened attention, for example, the publication by Bradburn, Rips, and Shevell (1987) in *Science*, or the edited volume by Schwarz and Sudman (1994), which attracted a number of generally well known researchers outside the direct sphere of survey research. Skowronski et al. (1994) reviewed the particular difficulty in dating events. Work on theories of telescoping (Bradburn, Huttenlocher, and Hedges, 1994; Huttenlocher, Hedges, and Prohaska, 1988; Rubin and Baddeley, 1989) emerged. A more extensive literature review would clarify the relationship between the different theories of telescoping and evaluate to what extent they have been tested subsequently.

But the basic science of memory has moved forward substantially in the last 15 to 20 years. As one example, a recent paper by Daselaar et al. (2008) applies fMRI to study the involvement of different areas of the brain in autobiographical recall. Significantly, recall of autobiographical information, in the sense it is now generally understood, takes a relatively long time relative to most cognitive tasks, and for this reason fMRI, with good spatial but poor temporal resolution, could be applied to measure both, a novel use of this technology. The paper did not examine the specific issue of the recall of dates, but it represents a paradigm by which researchers in the near future might be able to study the complexity of recalling dates. Closer ties between survey researchers and other disciplines could accelerate these developments.

## Acknowledgements

The findings and views expressed in this paper are solely those of the authors. They do not necessarily reflect those of the Bureau of Justice Statistics or of Westat.

We want to thank Mike Brick, David Cantor, and Sharon Lohr for their consultation and contributions at different stages of this research. We thank the Bureau of Justice Statistics for the original support for this work.

.

## References

Bradburn, N.M., Huttenlocher, J., and Hedges, L. (1994),"Telescoping and Temporal Memory," in Schwarz, N. and Sudman, S. (eds) (1994), in *Autobiographical Memory and the Validity of Retrospective Reports,* New York: Springer-Verlag, pp. 203-215.

Bradburn, N., Rips, L.J., and Shevell, S.K. (1987), "Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys," *Science*, 236, 157-161.

Bushery, J. M. (1981a), "Recall biases for different reference periods in the National Crime Survey," in *Proceedings of the Survey Methods Research Section,* Alexandria, VA: American Statistical Association, pp. 238-243.

Daselaar, S.M., Rice, H.J., Greenberg, D.L., Cabeza, R., LaBar, K.S., and Rubin, D.C. (2008), "The Spatiotemporal Dynamics of Autobiographical Memory: Neural Correlates of Recall, Emotional Intensity, and Reliving," *Cerebral Cortex,* 18, 217-229.

Fay, R.E. (2009), "Why is Survey Research 20 Years Behind?" in *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association, pp. 844-852.

Huttenlocher, J., Hedges, L., and Prohaska, V. (1988), "Hierarchical Organization in Ordered Domains: Estimating the Dates of Events," *Psychological Review,* 95, 471-484.

Lynch, J.P. and Addington, L.A. (eds) (2007), *Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR,* Cambridge University Press, New York, NY.

National Research Council (2008), *Surveying Victims: Options for Conducting the National Crime Victimization Survey,* Robert M. Groves and Daniel L. Cork, (eds.), Panel to Review the Programs of the Bureau of Justice Statistics, National Academy of Sciences, National Academy Press, Washington, DC.

Neter, J. and Waksberg, J. (1963), "Effects of Interviewing Designated Respondents in a Household Survey of Home Owners' Expenditures on Alterations and Repairs," *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 12, 46-60.

______ (1964a), "Conditioning Effects from Repeated Household Interviews," *Journal of Marketing*, 28, 51-56.

______ (1964b), "A Study of Response Errors in Expenditures Data from Household Interviews," *Journal of the American Statistical Association*, 59,18-55.

Rand, M. and Catalano, S. (2007), "Criminal Victimization, 2006," NCJ 224390, issued by the Bureau of Justice Statistics, Dec. 2007.

Rennison, C.M. and Rand, M. (2007), "Introduction to the National Crime Victimization Survey," in Lynch and Addington (eds.) (2007), *Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR*, pp. 17-54.

Rubin, D.C., and Baddeley, A. (1989), "Telescoping is not Time Compression: A Model of the Dating of Autobiographical Events," *Memory and Cognition,* 17, 653-661.

Rubin, D.C., and Wenzel, A.E. (1996), "One Hundred Years of Forgetting: A Quantitative Description of Retention, *Psychological Review*, 103, 734-760.

Schwarz, N. and Sudman, S. (eds) (1994), *Autobiographical Memory and the Validity of Retrospective Reports,* New York: Springer-Verlag.

Skowronski, J.J., Betz, A.L., Thompson, C.P., Walker, W.R., and Shannon, L. (1994), "The Impact of Differing Memory Domains on Event-Dating Processes in Self and Proxy Reports," in *Autobiographical Memory and the Validity of Retrospective Reports,* New York: Springer-Verlag, pp. 217-231.