Analyzing Potential Mode Effects in the National Crime Victimization Survey

G. Lance Couzens[1], Marcus Berzofsky[1], Christopher Krebs[1]
[1]RTI International, 3040 E Cornwallis Rd, Durham, NC 27709

Abstract

The National Crime Victimization Survey (NCVS) is a nationally-representative survey of the non-institutionalized U.S. population aged 12 and older and utilizes a 7-wave rotating panel design. Prior to 2006, first wave interviews were not used in crime victimization estimation – these interviews were used only to provide a temporal landmark for survey participants to control telescoping, ensuring that the second- through seventh-wave interviews had bounded reference periods. Beginning in 2006, the Bureau of Justice Statistics (BJS) began using first-wave interview data to produce estimates as a way to cut costs without sacrificing estimate precision. This change is significant in many ways, not the least of which is its potential for the introduction of multiple sources of bias. This paper addresses how the authors approached analysis of mode effect bias – one of the potential sources of error in the NCVS resulting from this change. Since first-wave interviews are primarily conducted in-person and second- through seventh-wave interviews are primarily conducted over the phone, the potential for mode effect bias is high. Quantifying this potential source of error is important for understanding how incorporation of first-wave data affects victimization estimates in the NCVS. After accounting for respondents' level of exposure to the NCVS, the difference in victimization rates between in-person and telephone groups is shown to be non-significant, suggesting the apparent mode effect is actually a symptom of respondent fatigue.

Key Words: National Crime Victimization Survey (NCVS), Panel Survey, Survey Mode, Respondent Fatigue, Total Survey Error (TSE)

1. Introduction

The National Crime Victimization Survey (NCVS) utilizes a dual-mode panel design in which respondents are interviewed up to seven times over a three-year period. Unlike the current designs of other nationally representative panel surveys, the NCVS incorporates first-wave interview data in published estimates. In order to address the potential for telescoping bias introduced by this practice, the Bureau of Justice Statistics (BJS) adjusts first-wave victimization data with a bounding factor. This paper attempts to answer two specific questions: (1) is there a mode effect in the NCVS, and (2) in light of the answer to question 1, is the current bounding adjustment appropriate.

1.1 Background

The NCVS is one of two primary sources of national crime statistics in the United States (along with the Federal Bureau of Investigation's Uniform Crime Reporting system) and the only national source that measures both reported and unreported crimes. The NCVS has been used to estimate the nature and frequency of U.S. crime victimization since 1973. Under the current survey design, approximately 90,000 households and 160,000 persons are interviewed in the NCVS each year (Truman, Langton, and Planty, 2013). Sampled persons and households are allocated to panels and rotation groups for the purpose of data collection and are interviewed across seven waves, at six-month intervals, over a period of three years.

Each person 12 years of age or older in sampled households is surveyed individually on questions of personal victimization, and a single person is designated from each household to report on property crime at the household level. According to the survey design, first-wave interviews are conducted in-person, and later-wave interviews take place over the telephone. Though deviations do occur for reasons of nonresponse follow-up and replacement of households (i.e., when new families move into sampled addresses), this dual-mode approach is the prevailing scenario for the majority of respondents[1]. The dual-mode approach is utilized primarily for two reasons: (1) to increase the cooperation rate and (2) to reduce survey costs. Because of the longitudinal aspect of the panel design, it is critical to obtain a household's buy-in for the study so that its members are willing to stay in the panel for all interview waves. Obtaining this cooperation is thought to be more easily done when the interviewer is in-person at the household rather than calling on the phone. Holbrook et al. (2003), for example, showed that in-person interviewing is associated with increased cooperation and engagement as well as decreased dissatisfaction with interview length and suggests that panel retention may be higher for in-person surveys. Once a household agrees to cooperate in the study, however, subsequent interviews are conducted by telephone. The dual-mode approach is therefore thought to maximize participation while reducing costs as much as possible.

Though the survey is dual-mode, historically, the majority of in-person interviewing that contributed to victimization estimates resulted from nonresponse follow-up in later collection waves and initial interviewing of replacement households[2]. This is because – prior to 2006 – first-wave interviews were only used to bound the reference period of the second-wave interview. This approach of discarding first-wave data is commonly used in similar surveys where the goal is collection of counts over a fixed reference period[3] and is based on the idea that respondents in the first wave are more likely to make recall errors. This is due to the fact that, unlike in later interviews, respondents in the first wave have no temporal landmark delineating the boundary of the survey reference period and are therefore likely to report more or fewer incidents on average than actually occurred in the past six months. This phenomenon, commonly referred to as telescoping, is widely documented; see, for example, Gaskell et al. (2000) which showed that, among persons exhibiting recall errors, forward telescoping (reporting that an event occurred more

[1] Within the analysis sample, 72% of wave 1 interviews were conducted in-person, and 64% of wave 2-7 interviews were conducted by telephone.
[2] Though individuals and families often move within the three-year lifespan of their panel, their address remains in the sample – replacement households are comprised of the individuals that move into a sampled domicile in or after the second wave of collection.
[3] Such as the Current Population Survey and the Consumer Expenditure Survey

recently than it actually did) may be more prevalent than backward telescoping for certain event types. Gaskell et al. also showed that this tendency may be more pronounced for younger respondents, a characteristic that poses a particular problem for the NCVS, as younger participants report more crime, on average (Truman et al., 2013). Referring specifically to victimization reporting in the National Crime Survey (precursor to the modern NCVS), Biderman and Cantor (1984) describe the tendency among unbounded respondents to forward telescope and the resulting inflationary effect on victimization rates.

Since 2006, crime estimates from the NCVS have incorporated data from unbounded first-wave interviews. This change resulted from the need to cut costs associated with data collection and the desire to do so without sacrificing existing precision standards. Though on its face this design change meets the desired goals – estimates are based on similar numbers of cases with smaller fielded samples – it opens the door for multiple sources of bias, the most obvious of which is telescoping. In order to mitigate this risk, BJS instituted an adjustment factor for first-wave responses (Rand and Catalano, 2007). This so-called bounding adjustment is formulated as the weighted ratio of average wave 2-7 victimization counts, over wave 1 victimization counts. This adjustment factor is estimated using data from the 12-month period preceding the month in which it is to be applied. In application, the factor is used as a multiplier for all wave-1 respondents' victimization counts. Though this adjustment is intended to reduce the risk of telescoping bias, it potentially introduces mode effect bias since wave 2-7 interviews are primarily telephone and wave-1 interviews are primarily in-person.

1.2 Purpose

As part of a larger review of current NCVS methods, BJS is considering a broad spectrum of modifications that address methodological shortcomings from a total survey error perspective. This means it is essential to understand all potential and actual sources of error in the existing design, rather than to focus on (or adjust for) only a single source without understanding how it relates to others. As noted, the inclusion of the first interview wave in the estimation process creates the potential for two types of error resulting from telescoping and a mode effect. If not addressed, each of these error types is likely to affect estimated victimization rates. However, there are other sources of error which may also affect rates. Each of these types of error must be examined and their relationships to one another understood. This paper addresses the individual examination of the mode component and how it relates to the existing adjustment for telescoping. The presented analysis attempts to answer the question of whether or not a mode effect exists in the NCVS, and what impact – if any – it has on the bounding adjustment. Several analytic methods were considered and are presented here in increasing order of complexity and the number of potential confounders controlled for. Other studies are concurrently examining other potential sources of error.

Since the post-2006 NCVS employs two modes for primary data collection and since data from predominantly telephone interviews are used to adjust data from predominantly in-person interviews, any effects on victimization reporting attributable to collection mode would result in biased adjustment factors. This could lead to adjusted first-wave rates that are either too high or too low.

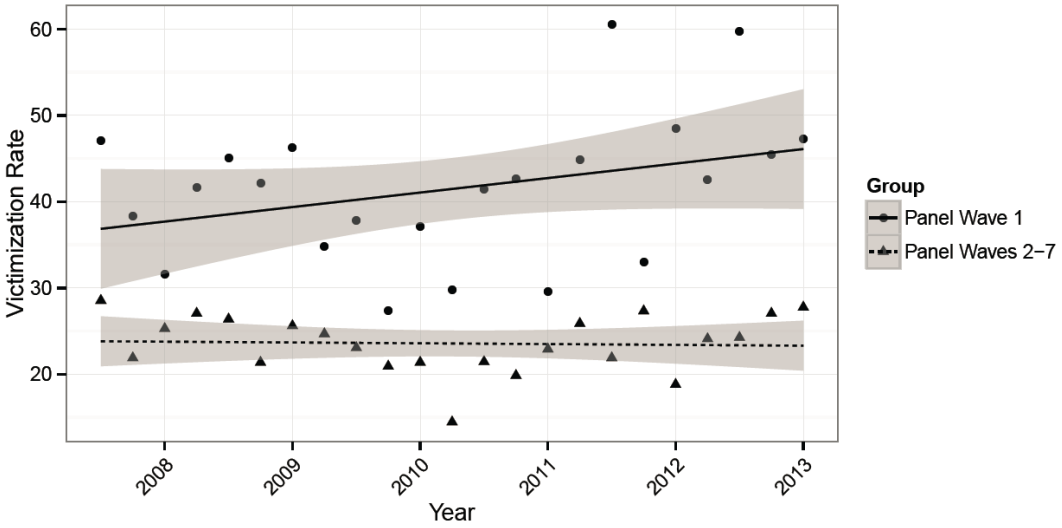

Figure 1: Total quarterly violent crime rates per 1,000 persons by year and panel wave group (1 vs. 2-7)

Figure 1 shows violent victimization rates separately for wave-1 respondents and wave 2 through 7 respondents, illustrating that first-wave victimization reports are consistently higher than those from later waves. Based on the formulation of the bounding adjustment, which benchmarks wave 1 responses to those from waves 2 through 7, this difference is attributable to the first wave being unbounded, and, if true, the adjustment would be appropriate as it would effectively address that single source of error. It is not clear, however, that the difference between wave-1 and wave 2-7 rates is attributable solely to telescoping in the first wave. Figure 2, for example, shows that there is a consistent difference between the rates from in-person and telephone respondents. Since wave 2 through 7 responses come primarily from telephone interviewing, it's possible that the difference seen in Figure 1 is driven – at least in part – by a mode effect. This paper attempts to determine whether or not a mode effect exists in the NCVS as a way to assess the appropriateness of the existing bounding adjustment.

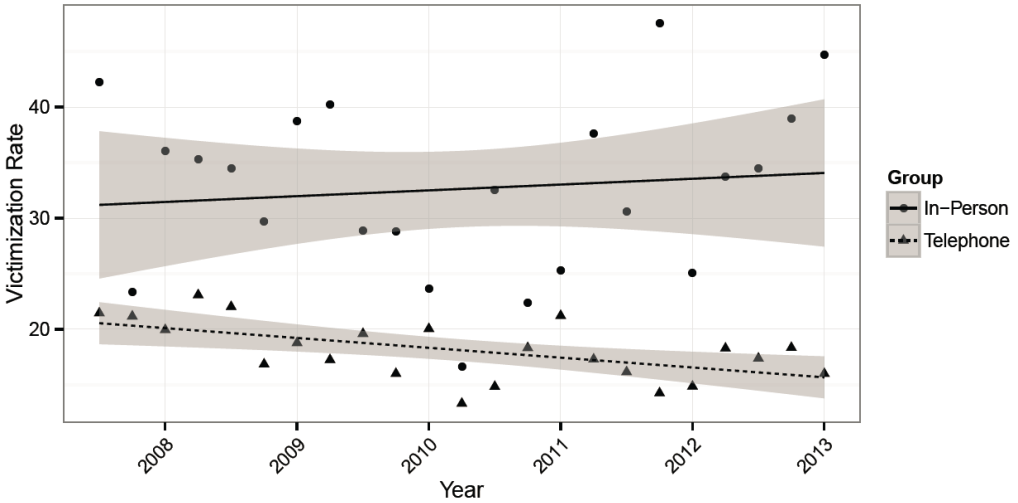

Figure 2: Total quarterly violent crime rates per 1,000 persons by year and mode group (wave 2-7 respondents only)

2. Methods and Findings of Mode Effect Investigation

As an extension to the graphical analysis above, the following sections describe the statistical analyses undertaken for the purpose of determining whether a mode effect exists in the NCVS and, if so, what its impact on victimization rates may be. All analyses employed person-level data from panels and rotation groups having all collection waves in the years 2007-2012 (n=429,250; $n_{telephone}$=251,120; $n_{in\text{-}person}$=178,130), and all statistical analyses utilized estimation tools that account for the complex design of the NCVS.

2.1 Unadjusted Poisson Regression

The first question to be answered in the search for a mode effect is: do victimization rates differ significantly by mode group? To address this question, unadjusted Poisson regression was used. Under this approach, victimization counts are modeled as a function of assigned mode group alone. The mode effect was first estimated using data from all waves, with wave 1 data incorporating the bounding adjustment. The resulting beta coefficient for in-person interviewing (telephone omitted) was positive and highly significant (p<0.001). To remove any potential influence of the bounding adjustment itself, this basic model was also estimated using data only from waves 2 through 7 – the resulting beta for in-person interviewing remained highly significant (p<0.001).

2.2 Adjusted Poisson Regression

Though the simple bivariate analysis showed a significant relationship between mode group and the level of victimization reported, no effort was made to control for potential confounders. To extend these basic results and control for confounding effects, several household- and person-level characteristics known to be associated with victimization were incorporated into the wave 2-7 model. These covariates included gender, race/ethnicity, age group, marital status, educational attainment, urbanicity, household income, employment status, and survey year. Again, the coefficient for in-person interviewing was highly significant (p<0.001).

The inability to 'control away' the apparent mode effect on victimization reporting is important because it suggests that the socioeconomic and demographic composition of the sample itself is not driving the significant influence of mode group through confounding. While this parameterization may be adequate in a cross-sectional design context where subjects are only interviewed once, it does not directly address subjects' prolonged exposure to the NCVS longitudinal design. Given that respondents' levels of exposure could be directly or indirectly confounded with mode and victimization, the next iteration of the multivariable analysis incorporated a covariate representing each respondent's panel-progression through the three years of inclusion in the NCVS sample. This measure, referred to as time in-sample (TIS) is an integer ranging from 1 to 7. Though this had no effect on the significance of mode, a more direct measure of individuals' survey progress (interview number) did. Interview number measures the level of survey exposure as an accumulation of interviews actually participated in and is person- rather than cohort-specific. For example, a person that participated only in waves 1 and 3 would have a TIS value of 3 in the wave 3 record, but would have an interview number value of 2, reflecting the instance of nonresponse in the second wave. Using

interview number and the previously listed socioeconomic and demographic covariates in a victimization model estimated over wave 2-7 respondents yielded a non-significant coefficient for in-person interviewing ($p>0.05$).

2.3 Propensity-Score Balancing

Findings from the multivariable analysis suggest that interview modality itself may not be driving the apparent mode effect, but rather that it may be the result of respondents' extended exposure to the NCVS. To determine whether this is the case, an additional source of potential confounding needed to be addressed. In the NCVS, the mode in which a respondent's interview wave is conducted is not random – mode assignment is tied to nonresponse and household replacement. In-person interviewing in waves 2-7 and telephone interviewing in the first wave are used as part of new household initial interviewing and secondary nonresponse follow-up collection.

Ideally, to assess the impact of mode, the analysis sample would be balanced both on observed and unobserved characteristics. In an experimental setting this is achieved through randomization – experimental subjects are assigned to one group or another according to a completely random process that ensures – at least in expectation – that no one group favors characteristics that could influence the outcome of interest. NCVS mode groups obviously cannot be said to be in balance in this way, though through the use of propensity score methods, the influence of mode can be assessed using a sample that is balanced on observed characteristics, including those of the survey design itself (DuGoff et al., 2014).

2.3.1 Propensity-Score Matching

In order to form a balanced analysis sample and control for the influence of prolonged survey exposure, in-person interviewing propensity models were estimated within interview number-defined groups. The propensity scores generated through these models were then used to pair respondents across mode groups. Each in-person respondent was matched to a telephone respondent based on similarity of propensity scores. By forming pairs in this way, the final analysis sample contained equal numbers of in-person and telephone respondents, and these two groups were similarly composed with regard to the covariates used in the propensity model. Since the propensity models were stratified by interview number and included covariates for the person and household characteristics previously listed as well as the survey weight, subsequent analytic model were balanced on relevant socioeconomic and demographic characteristics as well as unmeasured time-variant characteristics and characteristics associated with the design, but that are not available on public-use datasets[4].

Once the analysis sample was formed, unweighted Poisson regression was used to assess the mode effect on victimization reporting both with and without the covariates from the propensity model. None of the resulting models contained a significant coefficient for in-person interviewing ($p>0.05$ in all cases).

[4] Any individual or household characteristic related to the sample design (e.g., geographic location) has the potential to affect analytic weights and may be captured by proxy in propensity models, despite not entering the models explicitly.

2.3.2 Inverse Treatment Probability Weighting

Inverse treatment probability weighting (ITPW) is an alternative to propensity score matching that makes use of the entire sample in the final analytic models. Where unmatched telephone cases are discarded in the matched pairs approach, all cases are used in ITPW. Each respondent is weighted according to its in-person mode propensity score; in-person respondents with low scores and telephone respondents with high scores receive higher weights. With the exception of these weights, the parameterizations of both the propensity and analytic models are the same across the two approaches, and, as was the case in the matched pair models, no mode effect was uncovered (p>0.05 in all cases).

3. Discussion

Though a high level comparison of victimization rates between in-person and telephone respondents shows a consistent and significant difference, a more careful look reveals that this difference disappears when analytic methods control for subjects' prolonged exposure to the survey. This finding strongly suggests that the difference in rates between the two mode groups is not the result of a mode effect, but rather that the difference is the result of respondent fatigue.

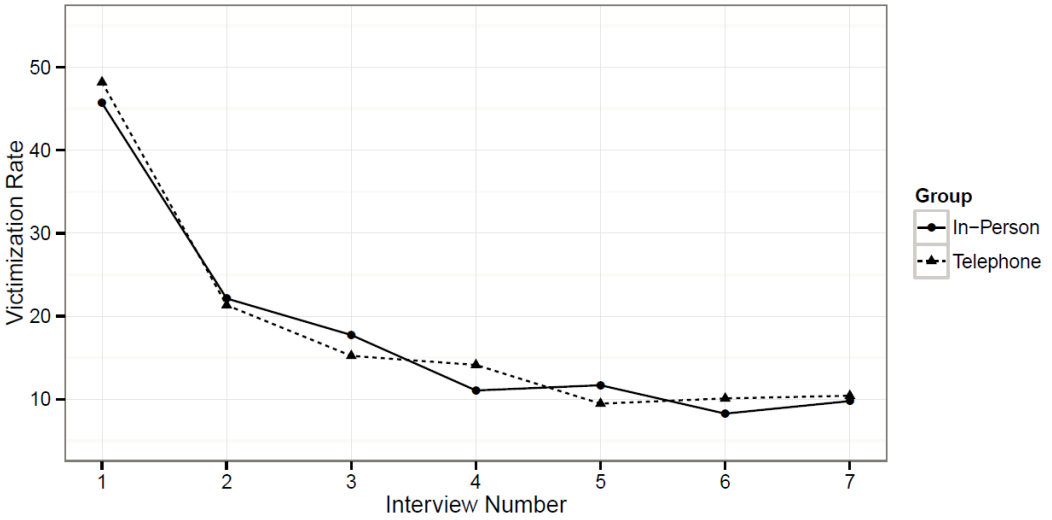

Figure 3: Total model-adjusted violent crime incident rates per 1,000 persons by interview number

In Figure 3 it is clear that when victimization rates are presented across interview numbers rather than by survey year, differences between the two mode groups evaporate. Though Figure 3 helps illustrate that there isn't a mode effect on victimization reporting, it also helps to explain why rates differ by mode group when survey exposure is not controlled for. Although there is no significant difference between in-person and telephone rates within any given interview number, rates are clearly dropping over the three years respondents are in the survey. Some of the decline – especially from the first to second interview – can be attributed to telescoping, but rates also drop by half from the second to seventh interview. It is also the case that in-person interviewing becomes less prevalent as interview number increases. This is likely attributable to a variety of reasons,

the most obvious of which is the inverse association between nonresponse propensity and willingness to participate. Taking these two pieces of information together, it is clear that rates can differ significantly by mode when no mode effect is present.

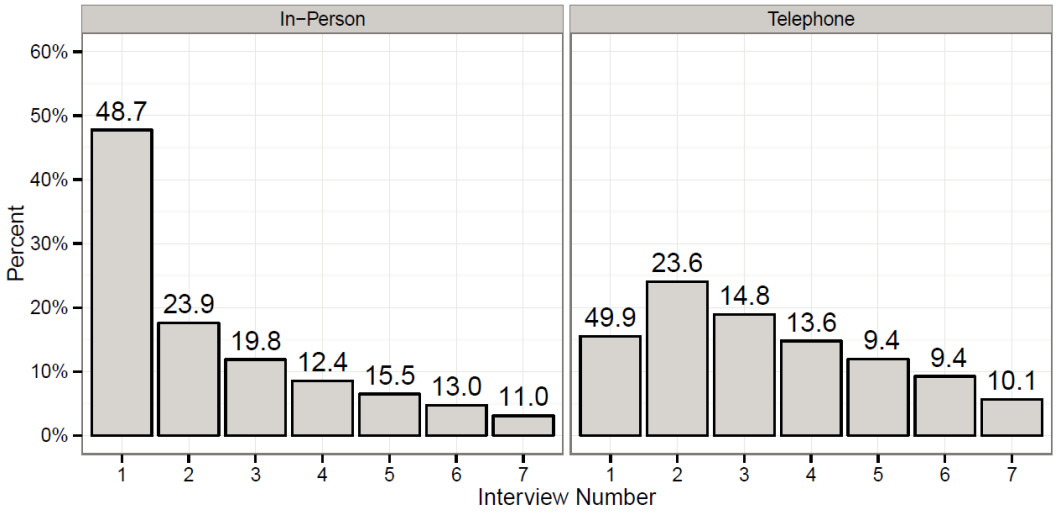

Figure 4: Interview number distribution by mode group, with interview number- and mode-specific violent victimization rates per 1,000 persons presented above bars

Figure 4 shows the interview number distributions by mode group along with the violent victimization rates associated with each group. Since the telephone group contains proportionally many more late-wave respondents, and since late-wave victimizations are relatively low, rates for in-person respondents are higher, even when first-wave responses are excluded. So, while this analysis has shown that there is no detectable mode effect in the NCVS, it has also shown that respondent fatigue has a strong deflationary effect on victimization rates, making the current implementation of the bounding adjustment inappropriate from a total survey error perspective, as it incorporates this bias in the adjustment of first wave responses. The fatigue effect is therefore no less troubling than a true mode effect would have been and requires further research to understand its drivers, its impact on various types of estimates, and appropriate mitigation strategies for reducing its influence moving forward.

Acknowledgements

The authors would like to thank BJS for sponsoring this research and for the comments provided by Mike Planty and Lynn Langton. However, we would like to note that the views expressed in this paper are those of the authors only and do not reflect the views or position of BJS or the Department of Justice.

References

Biderman, A.D., Cantor, D. (1984). *A Longitudinal Analysis of Bounding, Respondent Conditioning and Mobility as Sources of Panel Bias in the National Crime Survey.* In: American Statistical Association 1984 proceedings of the Section on Survey Research Methods. American Statistical Association, Washington, D.C.

DuGoff, E. H., Schuler, M., and Stuart, E. A. (2014). *Generalizing Observational Study Results: Applying Propensity Score Methods to Complex Surveys.* Health Services Research, 49(1):284–303.

Gaskell, G.D., Wright, D.B., O'Muircheartaigh, C.A. (2000). *Telescoping of Landmark Events: Implications for Survey Research*. Public Opinion Quarterly 6477- 89

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). *Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias.* Public Opinion Quarterly, 67(1), 79-125.

Rand, M. and Catalano, S. (2007). *Criminal Victimization, 2006: Technical Notes,* Government Printing Office, U.S. Bureau of Justice Statistics, Washington, DC.

Truman, J. L., Langton, L., and Planty, M. (2013). *Criminal Victimization, 2012,* Government Printing Office, U.S. Bureau of Justice Statistics, Washington, DC.