Potential Survey Error Due to a Panel Design: A Review and Evaluation of the National Crime Victimization Survey

Marcus E. Berzofsky, Andrew Moore, G. Lance Couzens, Lynn Langton, and Chris Krebs

Suggested Citation

Contents

About the Authors

Marcus E. Berzofsky, DrPH, is a senior research statistician in the Social Statistics Program at RTI International.

Andrew Moore, MS, is a research statistician in the Behavioral Statistics Program at RTI.

G. Lance Couzens, BS, is a research statistician in the Social Statistics Program at RTI.

Lynn Langton, PhD, is a senior research criminologist in the Victimization and Resilience Program at RTI.

Christopher Krebs, PhD, is a chief scientist in Applied Justice and Research Administration at RTI.

RTI Press Associate Editor

Ashley Amaya

Abstract

We use a total survey error approach to examine and make recommendations on how to adjust for non-sampling error in longitudinal, mixed-mode surveys. Using data from the National Crime Victimization Survey (NCVS), we examine three major sources of non-sampling error: telescoping, mode effects, and fatigue. We present an assessment of each source of error from a total survey error perspective and propose alternative adjustments to adjust better for this error. Findings suggest that telescoping and fatigue are likely sources of error in the NCVS, but the use of mixed-modes is not. Furthermore, both telescoping and fatigue are present in longitudinal surveys and accounting for one but not the other results in estimates that under- or overestimate the measures of interest—in this case, the rate of crime in the United States.

Introduction

The total survey error (TSE) framework dates back to the 1930s and is one of the predominate approaches for assessing and describing error properties of sample survey statistics (Groves and Lyberg, 2010). Potential sources of sampling and non-sampling error should not be considered separately; instead, survey designers should consider the totality of survey error and select a design that results in the smallest error overall. With this perspective, the TSE framework provides a measure of the accuracy and quality of survey estimates (Biemer, 2010). For those who use the data, as opposed to those who collect or create the data, measuring all aspects of TSE can be a challenge. For example, public use files often only have the final set of respondents. When only a public use dataset is available, the only way to evaluate representation error is through the documentation related to the creation of the final survey weights. Therefore, reviews like this one are forced to focus on the types of error that can be gleaned from the available data.

When considering the TSE framework, the sample design of a survey can dictate the error types that are present. For example, samples from a panel or longitudinal design introduce sources of error that cannot exist or be measured in a cross-sectional or one-time interview survey. In this paper, we examine the types of non-sampling error that can occur in a longitudinal or panel survey. Longitudinal or panel surveys have three major sources of potential non-sampling error that are unique to a longitudinal or panel surveys. These error sources include (1) telescoping error because of an unbounded interview, (2) fatigue, of which there are multiple sources, and (3) mode effect because of the use of different modes across interviews. We then examine if or how estimates can be adjusted to account for each error source. The assessment focuses on the Bureau of Justice Statistics' (BJS) National Crime Victimization Survey (NCVS), which uses a rotating panel design to produce estimates of criminal victimization in the United States.

The research presented in this paper has three broad audience types. First, federal agencies that design the NCVS (i.e., the Bureau of Justice Statistics [BJS])

or other panel surveys can use the methodology described in this paper to alter the current design of their panel surveys. Second, criminologists or other researchers who use the NCVS to estimate crime victimization in the United States can use the findings to adjust estimates of crime victimization. Third, survey methodologists can apply the methods described herein to design or modify other panel surveys.

Sources of Non-sampling Error in the NCVS

The NCVS is a rotating panel survey conducted by the US Census Bureau that interviews households seven times at 6-month intervals (US Census Bureau, 2014). A rotating panel design offers three distinct advantages over a cross-sectional survey: (1) all interviews after the first are bounded by the previous interview, (2) cost savings are realized through higher response rates and the ability to switch to lower cost modes in later waves, and (3) it offers a longitudinal data structure for assessing change over time (Berzofsky & Carrillo-Garcia, 2019). However, each of these advantages is countered by the potential for non-sampling error because of the panel design. The three major sources of error are unbounded interviews and telescoping, mode effects, and fatigue.

Although this paper focuses on the three sources of error most associated with longitudinal or panel surveys, they are not the only sources of error under the TSE framework. Within the broader TSE framework, the types of longitudinal or panel error being examined fall under measurement error (e.g., telescoping and mode effects) and nonresponse error (e.g., fatigue). However, other sources of error—such as coverage error, sampling error, adjustment error, and processing error—are not discussed heavily here because they would apply to the NCVS regardless of the design type selected. In other words, the use of a panel design does not influence the impact that these other error sources will have on survey estimates. Therefore, their influence on the final estimate will be constant across designs. For this reason, we do not focus on them in this paper.

Unbounded Interviews and Telescoping

The first source of error we examine is unbounded interviews, which increase the likelihood of telescoping—when a respondent reports an event that occurred before the reference period as though it happened within the reference period. Telescoping is a specific type of recall bias (Spencer, Brassey, & Mahtani, 2017). Recall bias more generally addresses a respondent's ability to cognitively remember an event. In telescoping, the respondent misremembers when the event occurred and brings it forward in time. Including unbounded interviews, which may contain telescoped events, can erroneously increase estimated rates (Biderman & Cantor, 1984). For this reason, most panel surveys (e.g., Current Population Survey, Consumer Expenditure Survey) exclude the unbounded initial interview from the analysis dataset and the estimation process. This was the case for the NCVS until 2006 when the initial interview wave was included in the analysis datasets and a bounding adjustment was applied to account for telescoping to maintain the respondent sample size (Rand & Catalano, 2007).

Beyond the inclusion of the first interview, there are six additional points during the NCVS when an interview could be unbounded. Table 1 presents the seven sources and the percentage of cases corresponding to each. Unbounded interviews make up 31.5 percent of interviews during a 6-month

Table 1. Percentage of interviews that are unbounded by reason for unbounded interview

Unbounded Source	All Respondents (%)	Unbounded Respondents (%)
Initial interview	14.2	45.1
Replacement household	6.2	19.5
New resident	1.4	4.5
Person aging in	0.5	1.5
Household non-response	2.6	8.4
Person non-response	4.8	15.2
Vacant household that becomes occupied	1.8	5.7
Total	**31.5**	**100.0**

Source: NCVS 2009, Quarters 3 and 4

period, and the biggest source of unbounded interviews (45 percent) is a household's initial wave. Other key sources of unbounded interviews are replacement households, which are households that move into the address and replace the initial tenant while the address is in sample; non-respondents in one wave who respond in the following wave; and those who age into the survey (the minimum age for respondents in the NCVS is 12 years old) during the panel period.

Mode Effects

The second potential source of non-sampling error in a panel survey is mode effects. Often, in panel surveys like the NCVS, a more expensive mode—an in-person interview—is used for the initial interview to help gain the cooperation of the household whereas a less expensive mode, like telephone interviewing, is used in subsequent interviews. Studies on mixed-mode designs have had conflicting results. Holbrook, Green, and Krosnick (2003) found that in-person interviews yield higher estimates than telephone interviews, whereas Cernat (2015) found that in-person and telephone interviews do not yield different results in a panel survey. NCVS interviewers strive to conduct in-person interviews for the first interview with a household or person and telephone interviews for follow-up waves whenever possible. About 44 percent of NCVS interviews are conducted by telephone.

Figure 1 presents violent victimization rates from 2007 to 2013 by mode of data collection for interview waves 2–7 in the NCVS. (Note that the findings in this figure look similar when all interview waves are included.) In-person interviews consistently yield higher victimization rates than telephone interviews, which indicates a possible mode effect. However, the results in Figure 1 do not control for factors like demographic characteristics or the length of time the person has been in the panel.

Fatigue

The final potential source of non-sampling error we examine is fatigue. For this paper, fatigue is used as an umbrella term for multiple types of error that occur because of the repeated (or longitudinal) nature of the survey design. Types of error that fall under this

umbrella include respondent fatigue and interviewer non-compliance. All sources of fatigue have the same impact on the resulting estimates—an under-reporting of the outcome of interest.

Respondent fatigue (also known as panel conditioning) can occur if respondents deliberately choose not to report incidents during subsequent interviews to reduce the interview duration (Hart, Rennison, & Gibson, 2005).

In the NCVS, for example, screener questions are used to identify whether the respondent has experienced a crime. An affirmative response to a screener question then triggers a detailed incident report, which is used to understand the nature of the incident and classify the type of crime. Once a respondent learns how much longer the interview becomes when an affirmative response is given to a screener question, they may choose not to report future incidents in subsequent interviews. However, respondent fatigue can be difficult to identify and quantify because even the act of taking the survey could cause a respondent to change their behavior and potentially avoid future victimizations (Cantor, 2007).

Interviewer non-compliance occurs when an interviewer rushes, does not administer the survey as designed, or uses information learned in past interviews to assume answers in the current wave. Rather than going through the entire survey instrument as intended, the interviewer leads the respondent to an answer or does not provide the opportunity to answer questions.

Figure 1. Violent victimization rates by mode of data collection and year, 2007–2013

Figure 2. Total violent crime rate by interview number

Figure 2 presents the violent victimization rates by interview number—the number of times a respondent participated in the survey regardless of the household's time-in-sample (TIS) (i.e., the number of times a selected address has been in the sample). Victimization rates decrease as a respondent's interview number increases, which may indicate fatigue.

Theoretical Model for Assessing TSE

The NCVS introduces a new sample of about 50,000 households to the panel every six months to replace the households rotating out after being in sample for seven interview waves. Because of non-response, replacement households moving into an address, and persons aging into the sample, the number

RTI Press Publication No. RR-0039-2007. Research Triangle Park, NC: RTI Press.

https://doi.org/10.3768/rtipress.2020.rr.0039.2007

of interviews in which a respondent has participated may differ from the TIS.

Figure 3 presents the theoretical constant rate of being a victim of violent crime along with the NCVS observed rate, across TIS. Because participation in the NCVS is not correlated with victimization risk, the probability of being a victim of crime should not change based on a household's TIS or respondent interview number. In other words, without any non-sampling error the expected or theoretical victimization rate at each interview should be relatively constant. However, Figure 4 shows that the observed victimization rate is higher at TIS-1 and lower across TIS-2 through TIS-7 than the theoretical rate.

The differences between the observed and theoretical model are caused by non-sampling error from unbounded interviews, mode effects, and fatigue. The first interviews for respondents represent a completely unbounded estimate that has no fatigue and minimal mode effects because most interviews are conducted in person—this would be the estimate produced by a cross-sectional survey. The second interview for a respondent represents a bounded estimate with potential mode effects and, assuming respondents become more fatigued with each subsequent interview, the least amount of potential fatigue. Because the likelihood of fatigue increases with each interview, it is hypothesized that the victimization rates at the first and second interviews represent the upper and lower bounds of the true victimization rate.

Figure 3. Observed and theoretical violent victimization rates by time-in-sample

Figure 4. Observed violent victimization rates and theoretical range for expected victimization rate by interview number

Approach

A two-step approach was used to examine the three major sources of potential non-sampling error in a longitudinal survey:

1. Assess each source of error independently and determine the best correction method, and

2. Develop a combined adjustment method based on the results of step 1.

As a working adjustment factor, it was anticipated that the combined TSE adjustment (ADJ_{TSE}) would take the form

$$ADJ_{TSE} = ADJ_{TELE} \times ADJ_{FAT} \times ADJ_{MODE}$$

where ADJ_{TELE} is the adjustment for telescoping, ADJ_{FAT} is the adjustment for fatigue, and ADJ_{MODE} is the adjustment for the mode effect.

Methods

Data

This study used NCVS public use data files from 2007 to 2013 (see, for example, US Department of Justice, Bureau of Justice Statistics, 2006, 2013). The survey years 2007–2013 were used to develop and assess the alternative adjustments for each source of non-sampling error. In 2006, the NCVS sample went through a decennial redesign, and at the same time, BJS began using TIS1 data with a bounding adjustment in the survey estimates. Thus, 2007 was the first full year, post-redesign, where a bounding adjustment was applied to the NCVS. These years were also included because persons and households could be linked across survey years for the full period.[1]

The use of public use data introduces a couple limitations to our analysis. First, geographic identifiers are greatly limited with only indicator for Census region present on the file. Therefore, any attempt to examine if certain errors are concentrated to particular geographies or interviewers is not possible. Second, the public use file only contains the final adjusted analytic weight. Therefore, any adjustments we propose cannot begin with the design-based weight or some earlier weight before all the final adjustments are applied.

For the development of the fatigue adjustment, it was necessary to use data from addresses that were in sample for all seven waves. Thus, only sampled addresses that were rotated into the sample in 2007 or later and were rotated out of the sample by the end of 2013 (i.e., the selected address
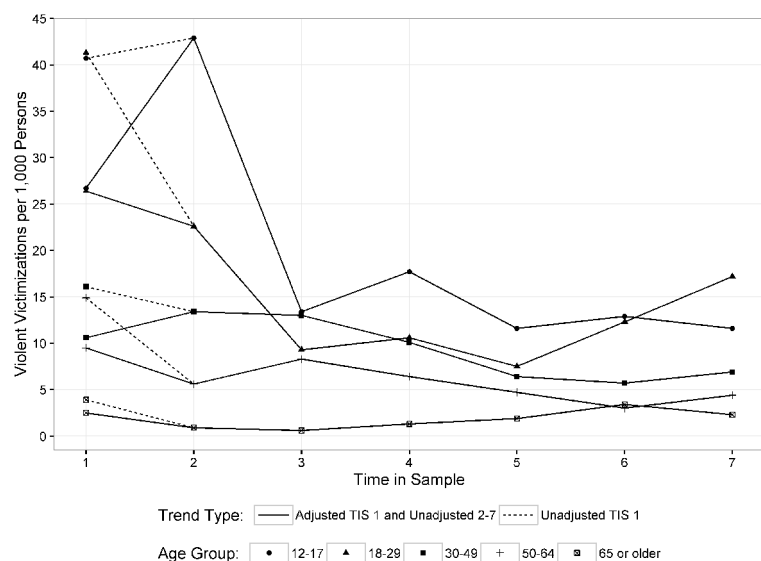
was in the sample for all 7 TISs) were included in the fatigue adjustment analyses; the sample and rotation groups included were Sample 24 rotation groups 5 and 6 and Sample 25. However, for the purpose of producing annual victimization rates, the analyses included all interviews conducted during that survey year as done by BJS when producing estimates.

Assessing Telescoping Error

When TIS-1 cases were included in the annual NCVS crime estimates in 2006, BJS and the Census Bureau recognized that this would increase the annual estimates because of telescoping and higher TIS-1 rates. To account for this, a bounding adjustment was applied to the weighted victimization rate in TIS-1, adjusting it to the average weighted victimization rate reported during TIS-2 through TIS-7. The bounding factor was based on the previous 12 months of data and calculated separately for each month.

The original bounding factor assumed that telescoping does not differ across demographic groups. However, in Figure 5, age is used to illustrate that the rate of telescoping is not the same across different subpopulations. In addition, the original adjustment was only applied to TIS-1 interviews but not to interviews unbounded for the six other reasons presented in Table 1. Thus, the goal of the assessment of telescoping error was to determine (1) if

---

[1] NCVS public use files before 2007 cannot be linked at the household or person level for two reasons: (1) the scrambled ID number for a household changed during the phase-in/phase-out period in 2005 and 2006 when the 2000 Census primary sampling units are rotated in, and (2) before 2006, the scrambled ID numbers were hand entered by interviewers; ID numbers cannot fully be linked because of errors in the keying process.

Figure 5. Bounded and unbounded violent crime rates by time-in-sample and age category

telescoping varies by demographic group and (2) how the inclusion of all sources of unbounded interviews impacts the adjustment factor and resulting estimates.

Assessing the Impact of Telescoping

To compare alternative adjustment factors for telescoping error, three broad classes of factors were defined:

1. Population-based adjustment factor,
2. Class-based adjustment factor, and
3. Model-based adjustment factor.

Population-based adjustment factor. The population-based adjustment factor is similar to the current adjustment factor in that it pools the victimizations from a broad class of crime types (e.g., violent crimes, property crimes) but does not consider demographic characteristics. The population-based adjustment factor differs from the current approach in that the cases considered bounded or unbounded can be varied as described in the section on alternative definitions of an unbounded interview. The definition for the population-based adjustment factor (BF_{PB}) is

$$BF_{PB} = \frac{\sum_{i \in Bounded} AVG(w_i \times numvic_{it})}{\sum_{i \in Unbounded} AVG(w_i \times numvic_{it})}$$

where w_i is the person or household weight for person or household I, and $numvic_{it}$ is the number of reported victimizations for person or household i for crime type t.

Class-based adjustment factor. The class-based adjustment factor builds on the population-based adjustment factor in that it conditions the adjustment factor based on a single characteristic. In other words, the class-based adjustment factor (BF_{CB}) is defined as

$$BF_{CB} = \frac{\sum_{i \in Bounded} AVG(w_i \times numvic_{it})|j}{\sum_{i \in Unbounded} AVG(w_i \times numvic_{it})|j}$$

where $j = 1, 2, \ldots, J$ is the characteristic level (e.g., White for the characteristic of race) for person i, and w_i and $numvic_{it}$ are as previously defined.

Because a person or household will possibly have many characteristics correlated with telescoping, to minimize the number of characteristics included in our analysis, a random forest–based variable importance analysis was used to determine the mostly

highly correlated characteristics.[2] This analysis was conducted for the telescoping of violent crimes and property crimes, respectively. The characteristics included in the random forest models are detailed in Table 2.

Table 2. Characteristics included in random forest analysis by type of crime

Violent crime	Property crime
Age category	• Age category
Gender	• Family structure
Education	• Gender
Household income	• Household income
Mode of interview	• Household tenure
Proxy respondent status	• Number of persons 12 or older in household
Race/ethnicity	• Number of persons under 12 in household
Replacement household	• Outside Access
Urbanicity	• Public housing status • Race/ethnicity • Replacement household • Urbanicity

Furthermore, characteristics with a missing value were imputed. For household income, which has a missing rate around 30 percent, we used the technique developed by Berzofsky et al. (2015). For all other characteristics, none of which had a missing rate greater than 5 percent, we used a conditional stochastic imputation based on the distribution of non-missing values by age, race, and gender within a given year.

Model-based adjustment factor. The model-based adjustment factor further expands the class-based adjustment factor to account for multiple characteristics simultaneously. The dependent variable for the model is the log number of victimizations reported by a person or household. The independent variables for the model are the set of characteristics identified through the random forest analysis that are most correlated with telescoping, an

---

[2] The random forest variable importance analysis is based on random permutations of variable values. If randomly permuting the values of a given characteristic dramatically inflates the MSE of a prediction model, that characteristic is said to be an important predictor of the outcome in question.

indicator for whether the interview was bounded or unbounded, and the two-way interaction between each characteristic and the bounded interview indicator. To illustrate this approach, using only gender as a model characteristic, the model would take the following form:

$$log[E(Y)] = \beta_0 + \beta_1 \times I_{Bounded} + \beta_2 \times I_{male} + \beta_3 \\ \times I_{Bounded} \times I_{male}$$

Where $I_{Bounded}$ is an indicator of whether the interview was bounded, I_{male} is an indicator of whether the respondent is a man, and β_i $i = 1,2,3$ are the corresponding model parameters.

The bounding factor for a set of characteristics is the ratio of the unbounded model parameter estimates and the bounded model parameter estimates. Based on this model the bounding factor for females (BF_{MB-F}) is

$$BF_{MB-F} = RR \left(\frac{Unbounded}{Bounded} \right) = exp \left[\frac{\beta_0}{\beta_0 + \beta_1} \right] \\ = exp[-\beta_1]$$

and the bounding factor for males (BF_{MB-M}) is

$$BF_{MB-M} = RR \left(\frac{Unbounded}{Bounded} \right) \\ = \left[\frac{\beta_0 + \beta_2}{\beta_0 + \beta_1 + \beta_2 + \beta_3} \right] = exp[-\beta_1 - \beta_3]$$

Because the model-based adjustment is based on multiple characteristics, some of the adjustment factors may be considered extreme. We will assess the adjustment factors for outliers and determine if Winsorization (i.e., trimming the upper or lower extreme adjustment factors) is necessary (Hasings, Mosteller, Tukey, & Winsor, 1947).

Alternative Definitions of an Unbounded Interview

For each of the three broad classes of alternative designs, five definitions of a bounded interview were assessed. In each of these alternatives a binary indicator for a bounded interview (I_{bnd}) was defined as the following.

1. $I_{bnd} = 1$ *if TIS* >1; $I_{bnd} = 0$ *if TIS* = 1
2. $I_{bnd} = 1$ *if TIS in* 2, 3, 4; $I_{bnd} = 0$ *if TIS* =1

3. $I_{bnd} = 1$ *if TIS* = 2; $I_{bnd} = 0$ *if TIS* = 1
4. $I_{bnd} = 1$ *if BOUNDED* = 1; $I_{bnd} = 0$ *if BOUNDED* = 0 where *BOUNDED* = 1 *if* person/household responded in TIS_k >1 and person/household responded in TIS_{k-1}
5. $I_{bnd} = 1$ *if BOUNDED* = 1 and *TIS in* (2, 3, 4); $I_{bnd} = 0$ *if BOUNDED* = 0 where *BOUNDED* = 1 if person/household responded in $1 < TIS_k \leq 4$ and person/household responded in TIS_{k-1}

Bounding reference types 1–3 use TIS to define an unbounded interview. Bounding type 1 is a population-based adjustment, similar to the bounding factor introduced to the NCVS in 2006. Bounding types 2 and 3 reduce the number of TISs included in the calculation to minimize the amount of fatigue and attrition included. However, the reduced number of TISs decreases the statistical power of these adjustments because of the inclusion of fewer interviews. Bounding reference type 2 uses only TIS-1 through TIS-4 as recommended by Berzofsky and Carrillo-Garcia (2014). Bounding reference type 3 only uses TIS-1 and TIS-2 because TIS-2 has the least amount of fatigue. Bounding reference types 4 and 5 define an unbounded interview based on whether the respondent was interviewed in the prior wave rather than the household's TIS. These bounding reference types are theoretically more accurate but a larger departure from the approach introduced to the NCVS in 2006. Bounding reference type 4 uses all 7 TISs to incorporate as many cases as possible. Bounding reference type 5 only includes cases from TIS-1 to TIS-4 for similar reasons as bounding reference type 2.

Assessing the Impact of Fatigue

Fatigue, which can cause respondents to suppress or reduce the reporting of victimizations, is assumed to be driven by exposure to the survey or a respondent's interview number.

It is difficult to know with certainty if a respondent who previously reported a victimization reported fewer or no victimizations in a subsequent wave because they knew the survey would take longer with an affirmative response or because they were not victimized. Couzens, Berzofsky, and Krebs

RTI Press Publication No. RR-0039-2007. Research Triangle Park, NC: RTI Press.

https://doi.org/10.3768/rtipress.2020.rr.0039.2007

(2014) tried to distinguish between these two situations and was unable to identify definitively when a respondent who previously reported victimizations reported fewer incident in subsequent interviews because of fatigue.

Rather than trying to isolate fatigued respondents, another option is to apply an adjustment to all respondents who reported a victimization and who could potentially be fatigued. Based on the theoretical framework, the total number of victimizations reported by random samples of the population should be roughly consistent over the same time period, regardless of the sample's TIS. Thus, a fatigue adjustment would inflate victimization weights in waves 2 through 7 to be consistent with model-estimated benchmark victimization levels that factor out the effect of fatigue because of survey exposure. Creating this inflationary adjustment factor is based on modeling the shape of the fatigue curve across interviews 2–7 and extrapolating that to interview 1, at least among bounded cases. This provides a fatigue-adjusted interview 1 rate that takes respondent characteristics into account and that is not affected by telescoping. Using the ratio of this estimated interview 1 rate over the interview number–specific rate estimated using the same model, we are able to reverse the deflationary effects of fatigue while accounting for the ways in which respondent characteristics affect the shape of the fatigue curve.

To adjust each respondent's number of victimizations to account for fatigue, a model-based approach similar to the model-based telescoping adjustment was employed. Under this approach, a separate model was estimated for person crimes and household crimes. The dependent variable was the number of reported victimizations (including zero). Therefore, for each victimization type a Poisson model was used to regress the number of victimizations reported to the interview number (as a continuous variable), the square of the interview number, person or household characteristics associated with fatigue, and the interaction between those characteristics and interview number and the square of the interview number.

This method generated a separate fatigue adjustment factor for each combination of the independent variables. An individual's fatigue factor is the rate

ratio of the estimated unfatigued victimization count over the estimated fatigue-affected count. For example, if the model contained interview number, the square of interview number, and gender (female reference level), then our model would be as follows:

$$
\begin{aligned}
log[E(Y)] = {}& \beta_0 + \beta_1 * WaveNum + \beta_2 * WaveNum^2 + B_3 \\
& * I_{male} + \beta_4 * WaveNum \\
& * I_{male} + \beta_5 * WaveNum^2 * I_{male}
\end{aligned}
$$

For a female respondent at wave number 4 the fatigue adjustment factor (FF) is

$$
\begin{aligned}
FF = \left(\frac{Unfatigued}{Fatigued} \right) &= exp \; \frac{\beta_0 + \beta_1 \times 1 + \beta_2 \times 1^2}{\beta_0 + \beta_1 \times 4 + \beta_2 \times 4^2} \\
&= exp \left[\beta_1 \times (-3) + \beta_2 \times (-15) \right]
\end{aligned}
$$

For a male respondent at wave number 7 the FF is

$$
\begin{aligned}
FF = RR \; &\left(\frac{Unfatigued}{Fatigued} \right) \\
= exp \; &\left[\frac{\beta_0 + \beta_1 \times 1 + \beta_2 \times 1^2 + \beta_3 + \beta_4 \times 1 + \beta_5 \times 1^2}{\beta_0 + \beta_1 \times 7 + \beta_2 \times 7^2 + \beta_3 + \beta_4 \times 7 + \beta_5 \times 7^2} \right] \\
= exp \; &\left[\beta_1 \times (-6) + \beta_2 \times (-48) + \beta_4 \times (-6) + \beta_5 \times (-48) \right]
\end{aligned}
$$

Because the model-based adjustment is based on multiple characteristics, some of the adjustment factors may be considered extreme. We assessed the adjustment factors for outliers and determined if Winsorization was necessary.

Assessing the Impact of the Mode Effect

To assess the impact of mode effect and to determine an appropriate adjustment, two modeling approaches were implemented: (1) Poisson model of the number of victimizations by mode type and (2) propensity score balancing (both matching and reweighting).

Poisson model. A Poisson model was developed to estimate the effect of mode on the number of reported victimizations, controlling for demographic characteristics and interview number.

Propensity score matching. Because mode is tied to non-response and survey exposure, and is, therefore nonrandom, propensity score models were used to balance the analysis sample on observed characteristics across mode groups. Propensity score matching and inverse treatment probability weighting were used to balance the analysis sample on person-

and household-level characteristics and survey weights and to estimate adjusted and unadjusted victimization (Poisson) models within each interview number.

Developing a Combined TSE Adjustment

Once potential adjustments were developed for each of the three sources of error, the next step was to implement a process for assessing which of the resulting adjustments best aligned with the TSE theoretical model. A three-step evaluation was developed in which adjustment options that did not meet minimum standards were removed from consideration. The three-step process included the following:

Evaluation step 1: Determine if, using the total adjustment factor (product of telescoping adjustment factor and *FF*), victimization rates are consistent across interview wave by type of crime.

Evaluation step 2: Determine if victimization rates based on the total adjustment factor are consistent across interview number for detailed demographic characteristics.

Evaluation step 3: Determine if annual victimization rates across time, by type of crime, differ by adjustment factor.

Because some of the adjustment factors require up to 36 months of retrospective data, the evaluation of the combined adjustments used data from 2008 to create the adjustment factors and data from 2011 to 2013 to produce victimization rates.

Evaluation Step 1

Based on the theoretical model, the goal for the revised adjustment factor was to produce victimization rates that are consistent across interview number. Theoretically, at a given point in time, the victimization rates from any interview wave should be nearly identical when estimated for the same population characteristics. Using data pooled across the 3-year periods, the current and proposed adjustment factors were computed by interview number for violent (i.e., rape and sexual assault, robbery, and assault) and property (i.e., burglary, motor vehicle theft, and other theft) crime. Because

the estimates were time invariant, pooling all 3 years maximized the statistical power to assess differences.

In this evaluation step, the victimization rates produced by each adjustment factor were assessed for consistency across interview number. For each type of crime, model-adjusted F tests of linear trend and time-variant fixed effects were computed to determine if the rates differed by interview wave after applying the adjustment factor. Adjustment factors that did not meet the criteria—a near constant crime rate across interview wave—were dropped from consideration for the final factor.

Evaluation Step 2

Although the current bounding adjustment factors for persons and households result in aggregate TIS-1 estimates in line with the average of TIS-2 through TIS-7, the adjustment factors across different subpopulations (e.g., young persons, persons living in urban areas) may over- or under-estimate the required adjustment because telescoping and fatigue are not necessarily constant across all groups. Using the adjustment factors that met the evaluation criteria in Step 1, victimization rates by key household and person characteristics, TIS, interview number, and type of crime were computed. The household and person characteristics assessed included both characteristics used in one or more of the adjustment factors and characteristics not used in any adjustment factors. The factors analyzed are shown in Table 3.

In this evaluation step, the victimization rates produced by each adjustment factor were reviewed

Table 3. Characteristics included and excluded from adjustment factors based on random forest analysis

Characteristics included in adjustment factors	Characteristics not included in adjustment factors
Age category	Gender
Education level	Marital status
Household income	Region
Household tenure	Reporting to police
Race/ethnicity	Victim of intimate partner violence
Urbanicity	Victim of violent crime with a weapon
	Victimizations with an injury

to assess if the rates are consistent for each level of a household or person characteristic by type of crime. Consistent with Step 1, χ^2 tests for each type of crime by characteristic level will be computed, and adjustment factors that do not meet the criteria will be dropped from consideration.

Evaluation Step 3

The third evaluation step determined which adjustment factor would be optimal for assessing annual victimization rates over time. Annual estimates were produced for 2011, 2012, and 2013. Annual victimization rates were produced using each of the adjustment factors that remained after Evaluation Step 2. Victimization rates were produced for each type of crime and compared with the published NCVS victimization rate trends and victimization rates over time for which no adjustment factor was applied. Although the "true" trend of victimization is not known, an adjustment factor that produced dramatically different looking trend lines from the current approach was strongly scrutinized. After reviewing the trend line from each adjustment factor, the adjustment method that produced an acceptable annual rate trend across the three years and had the best properties in Evaluation Steps 1
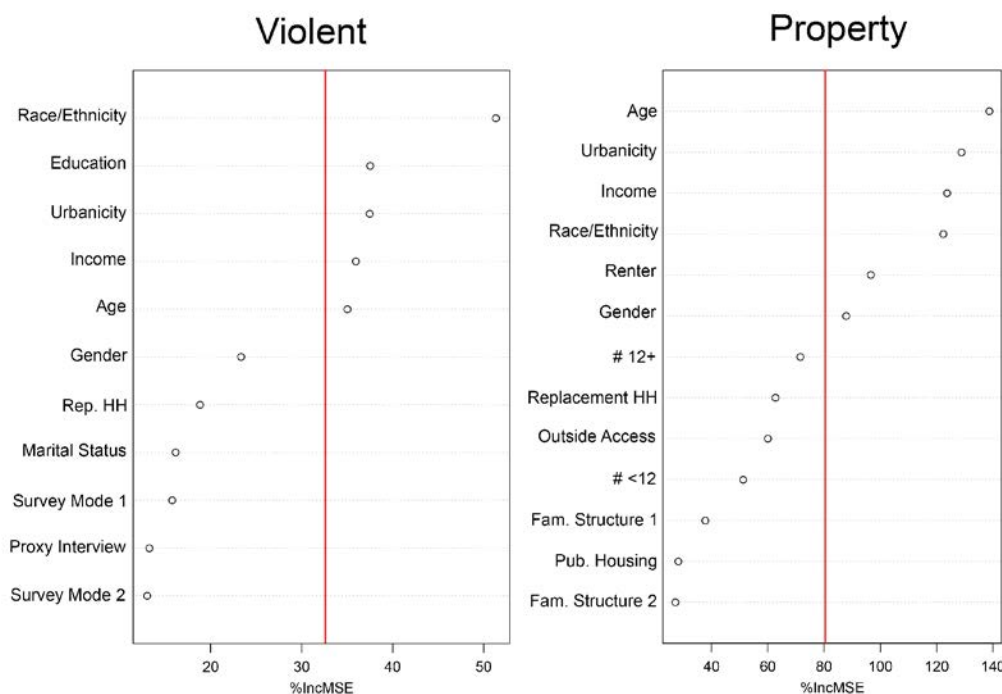
and 2 was recommended as the revised adjustment method.

Results

Telescoping Error

Figure 6 presents the results of the random forest analysis that was used to narrow down the characteristics used in the class-based and model-based adjustment factors. For both violent and property crime, we examined the results to see if there was a set of characteristics that clearly induced larger mean square error (MSE) increases than the other characteristics. For violent crime, a 35 percent MSE breaking point was chosen based on a subjective review of how the MSEs for each characteristic were clustered. The 35 percent MSE breaking point identified race/ethnicity, education, urbanicity, household income, and age for incorporation into the class-based and model-based adjustments. For property crime, an 80 percent MSE breaking point was chosen based on a subjective review of how the MSEs for each characteristic were clustered. This 80 percent MSE breaking point identified age urbanicity, household income, race/ethnicity,

Figure 6. Percent MSE for person and household characteristics based on random forest analysis for violent and property crimes

household tenure, and gender for incorporation into the class-based and mode-based adjustments.

For each alternative adjustment, a factor was produced based on all violent crimes and all property crimes. Considering the five bounding reference types and the results of the random forest analysis, 70 different alternative adjustments were produced. Each of these adjustment factors were reviewed for extreme values. No extreme values were identified for any of the telescoping adjustments.
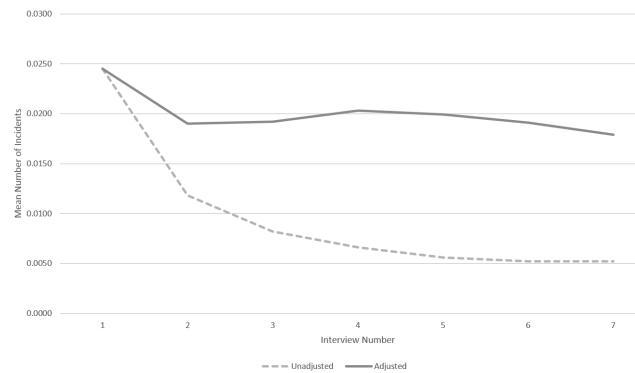
Fatigue Error

For the fatigue adjustment, the first step was to determine which characteristics to include in the model. The same characteristics identified during the random forest analysis (Figure 6) were used in the fatigue model.

Figure 7 presents the mean number of incidents reported by interview number before and after applying the fatigue adjustment. As expected, the number of incidents in interview numbers 2–7 was greater after the adjustment, but lower than the number incidents in interview number 1, which is subject to telescoping. The adjustment was successful in correcting the number of incidents to be similar in interview numbers 2–7.

The fatigue adjustment factors for each respondent were reviewed for extreme values. Because extreme values were identified and could result in respondents from having too much influence on the estimates, a 5 percent Winsorization was applied. As exemplified in Figure 8, the Winsorization had a small impact on the overall victimization rate but had varying impact by domain category. For instance, estimates for persons 12–17 were largely impacted whereas those 65 and older were negligibly impacted. Notably, when more data are

Figure 7. Mean number of incidents by interview number, adjusted and unadjusted by fatigue adjustment
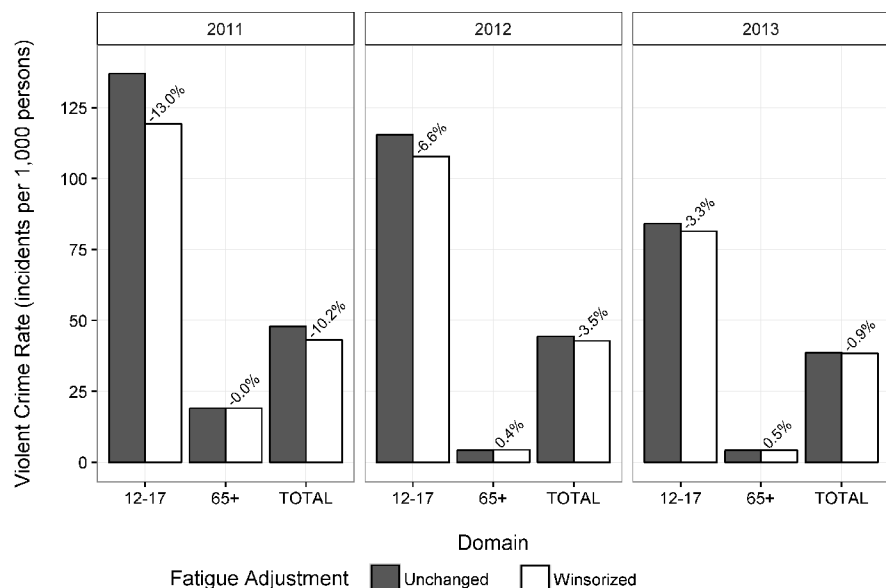

used to estimate the fatigue models (2011 had the fewest data points and 2013 the most), the impact of Winsorization decreases substantially. This illustrates the need for as many data as possible when fitting fatigue adjustment models.

Mode Effect Error

Figure 9 presents the violent victimization rates by mode after controlling for interview number. After taking interview number into account, the apparent mode effect disappears (see also Couzens et al., 2014). Moreover, as shown in Figure 10, the

Figure 8. Impact of 5% Winsorization in the fatigue-adjusted victimization rates on the violent crime rate for all persons, 12–17-year-olds, and persons 65 or older by year, 2011–2013

victimization rates at each interview number are similar by mode, but the percentage of interviews conducted in each mode differs by interview number. Therefore, the apparent mode effect is largely a function of telescoping and fatigue and does not need to be accounted for separately in the combined TSE adjustment.

Combined TSE Adjustment

Given the findings of the independent reviews of the three major sources of error in the NCVS, assessment of the combined TSE adjustment consisted of 70 possible adjustments. Because no adjustment for mode was necessary, the combined TSE adjustment was modified to

$$ADJ_{TSE} = ADJ_{TELE} \times ADJ_{FAT}$$

Evaluation Steps 1 and 2

Evaluation step 1 assessed victimization rates by type of crime across TIS and interview number. A successful adjustment would produce a near straight line across TIS and interview number. However, as shown in Figure 11, the initial results led to victimization rates that were lower in TIS-1 than the remaining TISs. After an investigation, it was determined that applying the telescoping adjustments to victimization counts that had not been adjusted for fatigue potentially over-corrected for telescoping. Failing to account for fatigue in TISs 2–7 caused the apparent telescoping error to be artificially inflated. Therefore, the adjustment factor over-corrected for telescoping in TIS-1. Based on this

Figure 9. Violent victimization rate by data collection mode after controlling for interview number and demographic characteristics
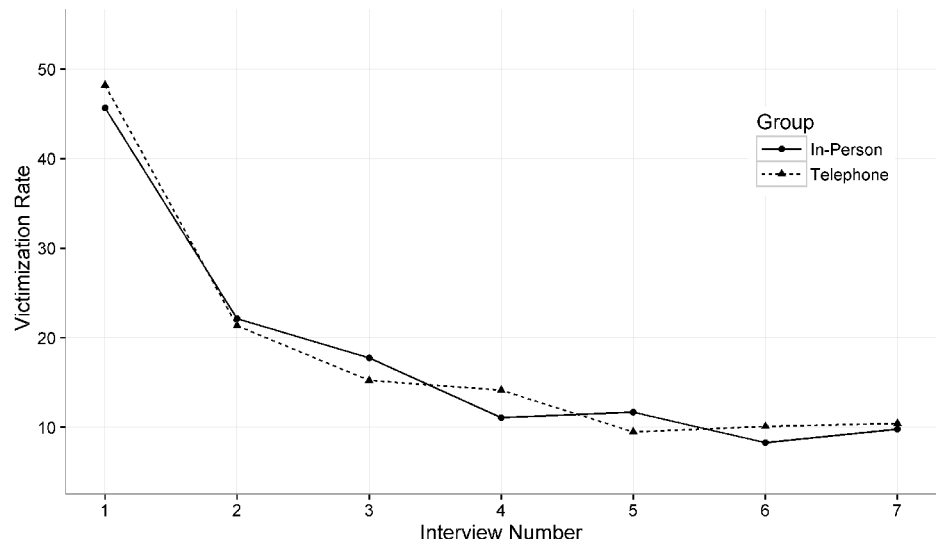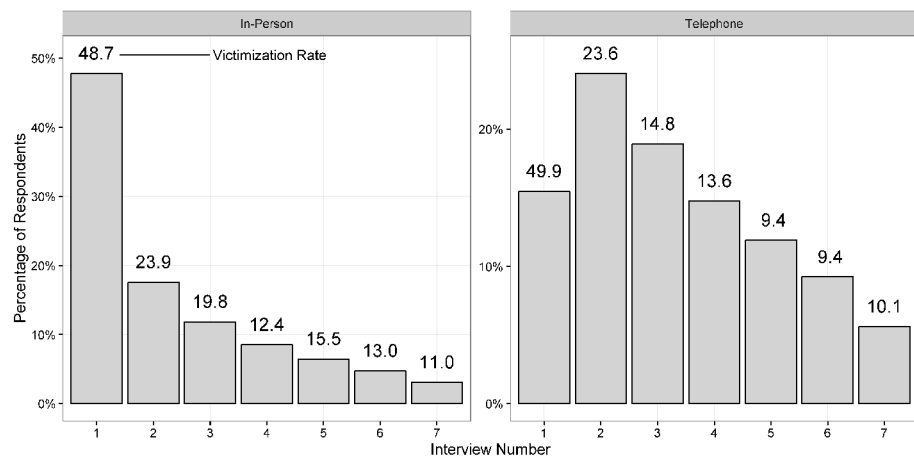

Figure 10. Percentage of interviews conducted and victimization rate by mode and interview number

finding, we modified the combined TSE adjustment to be

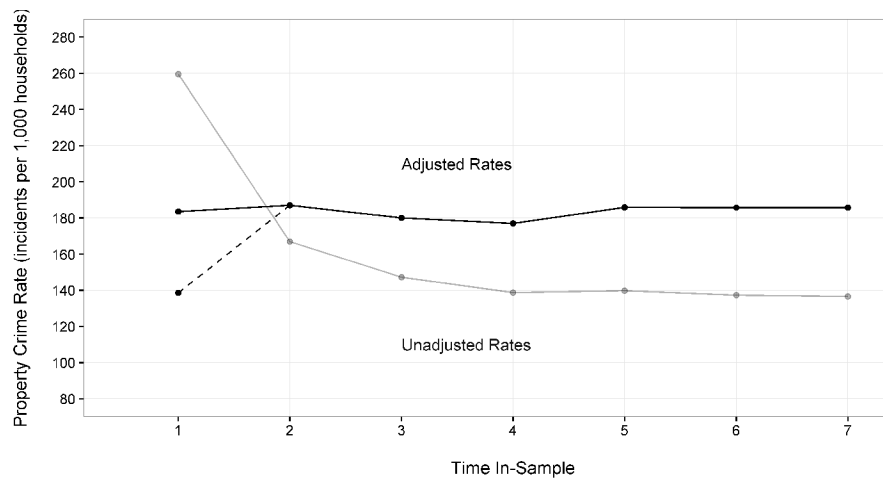$$ADJ_{TSE} = ADJ_{TELE|FAT} \times ADJ_{FAT}$$

As shown in Figure 11, once this adjustment was made the victimization rates across TIS became more level. The adjustment for fatigue is implicitly conditioned on telescoping because it was based on bounded cases only.

After adjusting the formula for the combined TSE adjustment, the overall victimization rates for violent and property crimes were compared for each of the five bounding types. As shown in Figure 12, the

adjusted rates for both violent and property crimes were similar for bounding types 1–3 (the types based on TIS) and for bounding types 4 and 5 (the types based on whether an interview is bounded). This suggested that the focus of the next evaluation steps could be limited to bounding type 1 (TIS-1 vs. TIS-2 through TIS-7) and type 4 (unbounded vs. bounded from all TISs). These two types were chosen over the other bounding types because they incorporated all available data, which increases the power of the adjustment factors.

Figure 11. Property crime victimization rates based on combined TSE adjusted before (dashed line) and after conditioning the telescoping adjustment on fatigue-adjusted incident counts by TIS
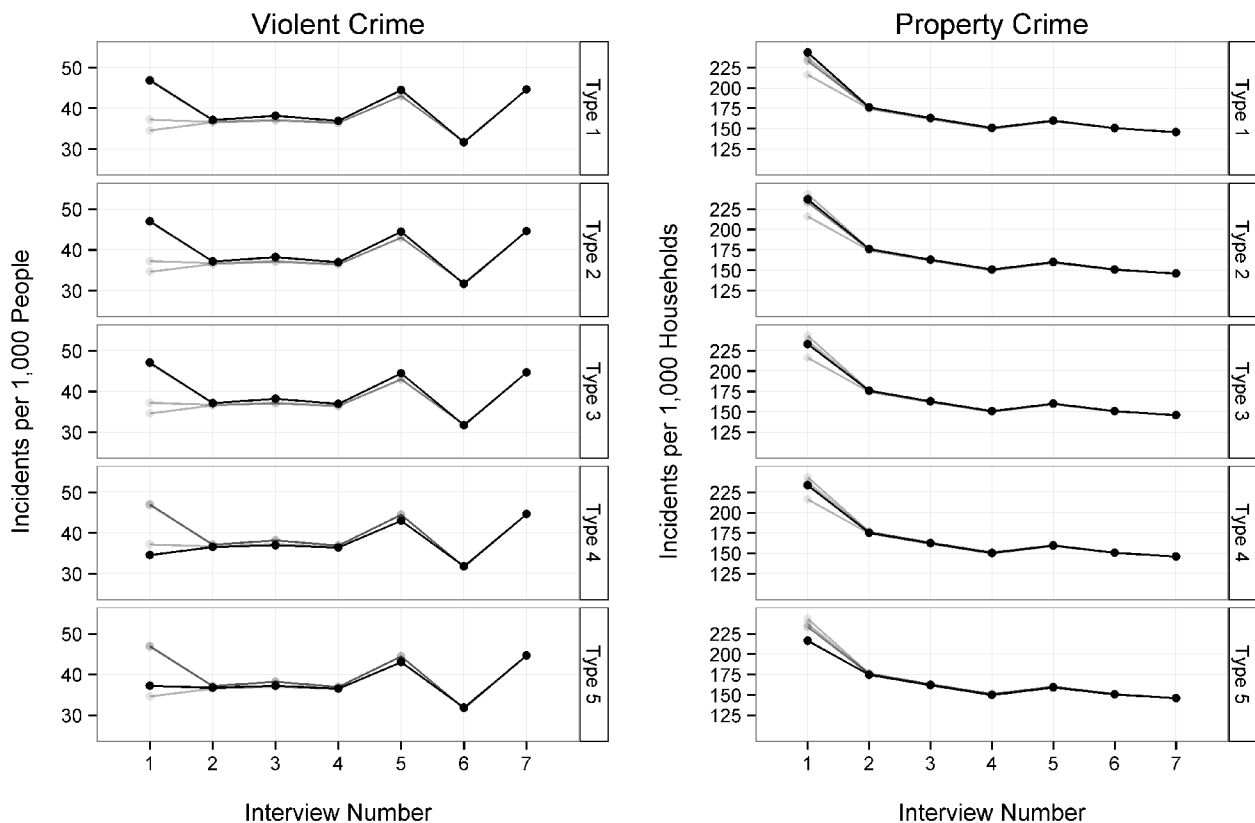
In Figures 13 and 14, the model-based telescoping adjustment and bounding types 1 and 4 are used to compute NCVS victimization rates for the major crime types by interview number. The figures show the unadjusted estimates (light gray lines) and the adjusted estimates (black lines). Across both bounding types and for all crime types, the adjustment resulted in trend lines that were closer to the theoretical model than the unadjusted estimates were. However, even with the adjustment

Figure 12. Violent and property victimization rates by interview number and bounding type

the lines are not completely flat. To determine if the difference in the slopes between the adjusted trend line and the theoretical trend line, an adjusted F test was computed. The adjusted F tests for linear trend generally found the victimization rate trends to be flat (i.e., slope = 0), but that was sometimes a result of wide confidence intervals (CIs) at later interview numbers. In general, the adjustment factors performed better for property crimes than violent crimes. Because the adjusted trend lines appeared similar for bounding types 1 and 4 both were included in Evaluation Step 2, but the other three bounding types were no longer considered.

The combined TSE adjustments based on the population-based, class-based, and model-based adjustments also performed similarly. Therefore, all three telescoping adjustment types were considered in Evaluation Step 2, which involved producing victimization rates by person and household characteristics of interest, interview number, and bounding type. The Evaluation Step 2 analysis showed that the adjustment performed well, except in cases where the sample size for a category was small. This

held true across characteristics of interest that were and were not included in the models for violent and property crime (see the Appendix for Figures A.1, A.2, A.3, and A.4, which show rates of victimization across demographic and incident characteristics).

As with Evaluation Step 1, the combined TSE adjustments based on the population- and class-based telescoping adjustments (not shown) performed similarly to the TSE adjustments that used the model-based telescoping adjustment. Therefore, all three options were considered in Evaluation Step 3.

Evaluation Step 3

Figures 15, 16, 17, and 18 provide examples of comparisons among published victimization rates for 2011, 2012, and 2013 and victimization rates based on several alternative combined TSE adjustments. Across violent crime, serious violent crime, general property crime, and motor vehicle theft, the BJS published rate (i.e., the estimate based on the current bounding adjustment) was the lowest and the unadjusted rate (i.e., the estimate generated without any bounding or fatigue adjustment) was the second lowest rate

Figure 13. Adjusted (black line) and unadjusted (gray line) victimization rates by type of violent crime and interview number by bounding type 1 and 4 based on the model-based telescoping adjustment
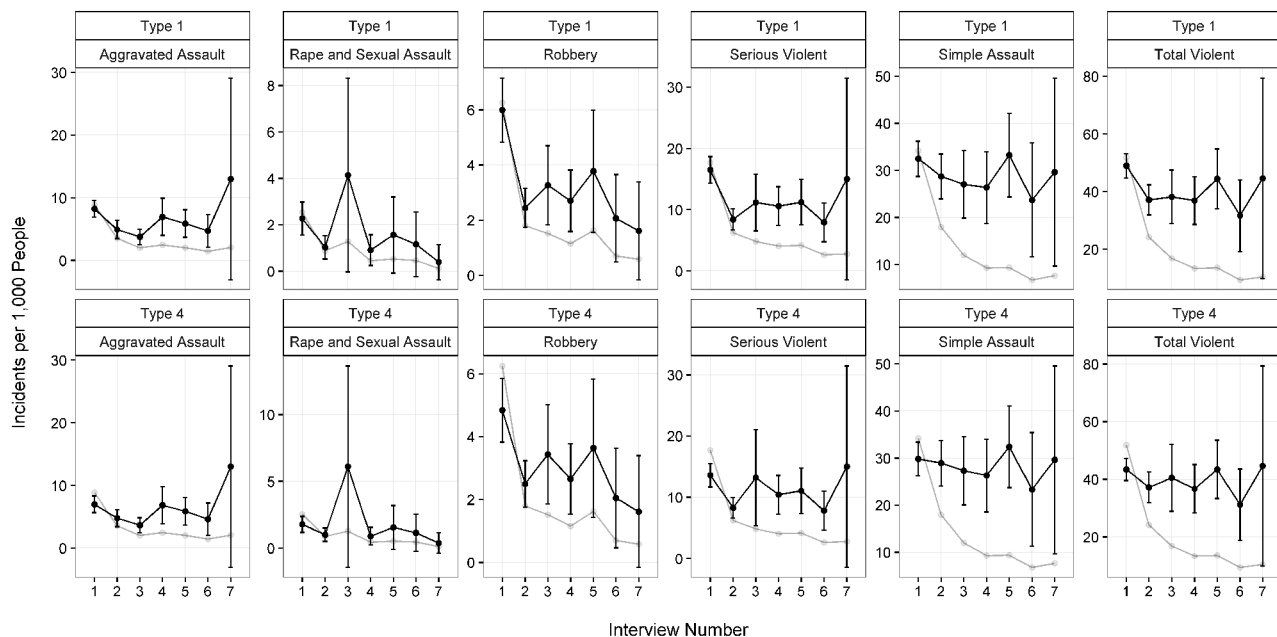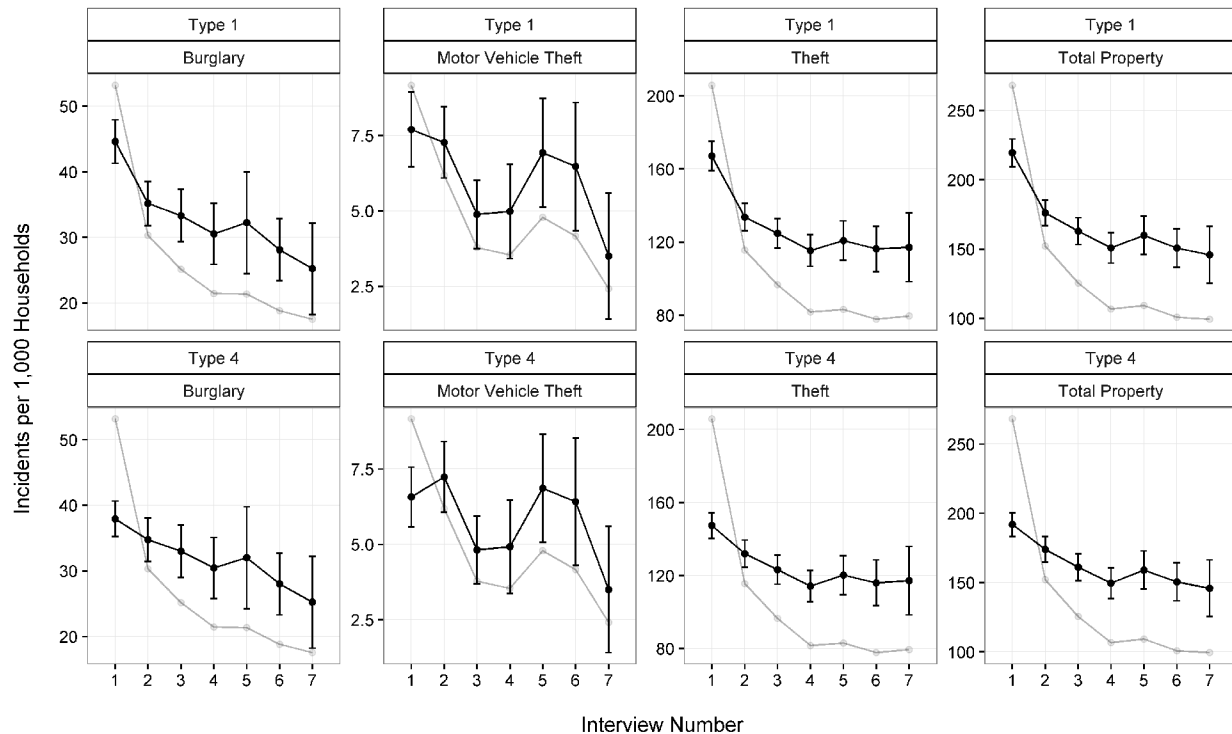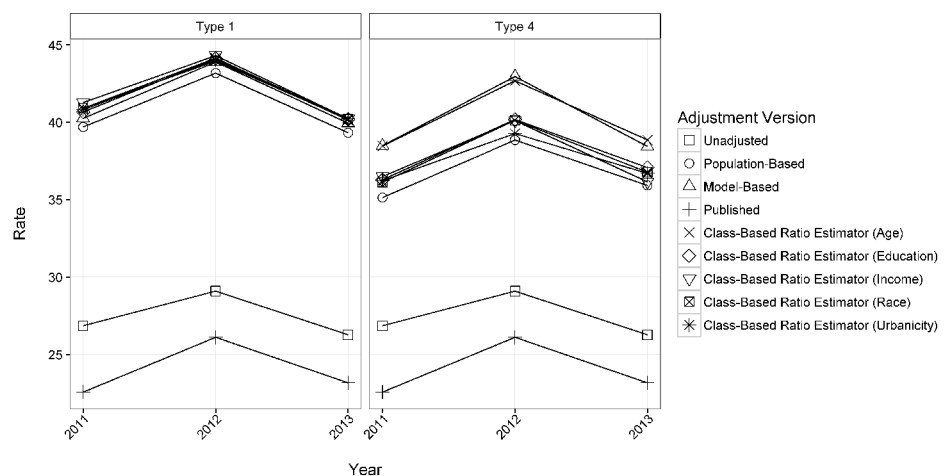
Figure 14. Adjusted (black line) and unadjusted (gray line) victimization rates by type of property crime and interview number by bounding type 1 and 4 based on the model-based telescoping adjustment

for all survey years. The victimization rates based on the combined TSE alternative adjustments were about 1.5 to 2 times larger than the published victimization rate for each type of crime. However, the order of the alternative adjustments (from highest victimization rate to lowest) was not always the same. For example, for violent (Figure 15) and serious violent crime (Figure 16), the model-based telescoping adjustment produced one of the highest victimization rates, but for property crime (Figure 17) and motor vehicle theft (Figure 18) several of the class-based adjustments produced higher rates. Although the magnitude of the adjusted victimization rates varied across bounding reference types 1 and 4, the rankings

Figure 15. Violent crime victimization rates by select combined TSE adjustment factors, bounding reference type, and survey year, 2011–2013

of which adjustment produced the highest rates were largely consistent by crime type.

For the most part, the trend line of the alternative adjustments (i.e., the slope of the victimization rates over time) followed the trend of the published

estimate. This implies that the alternative adjustments only impacted the magnitude of the estimate.

Discussion

Review of Error Because of the NCVS Design

At the onset of the analysis it was hypothesized that longitudinal survey designs lead to three potential sources of non-sampling error: (1) telescoping error, (2) fatigue, and (3) mode effect error.

Data from the NCVS show that unbounded interviews result in significantly higher victimization rates than bounded interviews. Based on this finding, we feel certain that respondents in unbounded interviews engage in telescoping. Therefore, an adjustment for all unbounded interviews is appropriate.

Although it was not possible to determine which respondents or interviewers were becoming fatigued (Couzens & Berzofsky, forthcoming), applying the theoretical model to NCVS data suggested the presence of fatigue in interviews 2 through 7. Not adjusting for fatigue while adjusting for telescoping only applies a downward adjustment without any compensating upward adjustment. From a TSE perspective, fatigue

Figure 16. Serious violent crime victimization rates by select combined TSE adjustment factors, bounding reference type, and survey year, 2011–2013
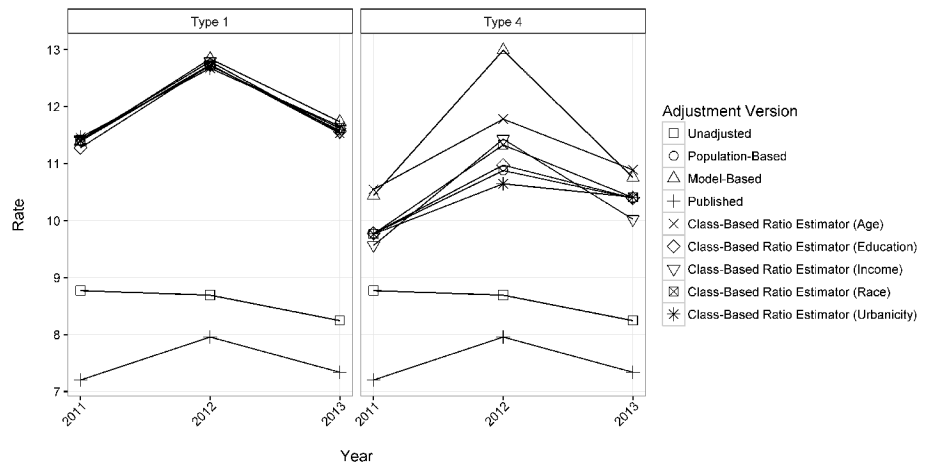

Figure 17. Property crime victimization rates by select combined TSE adjustment factors, bounding reference type, and survey year, 2011–2013
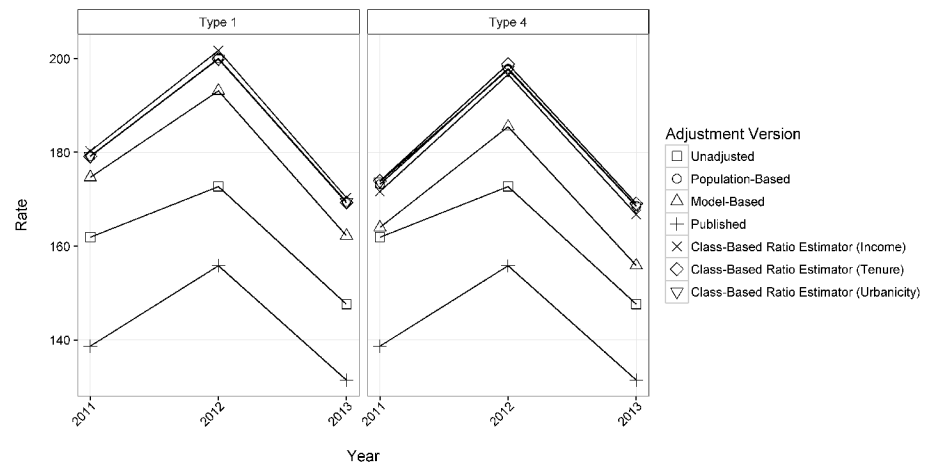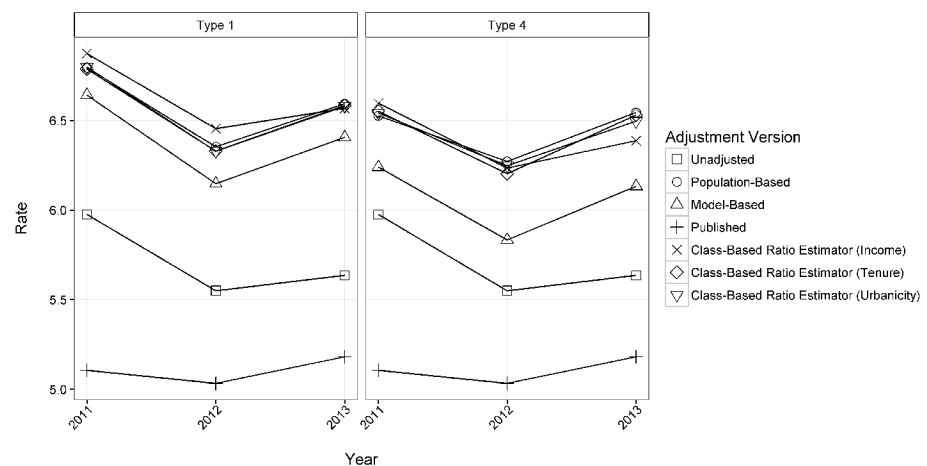

Figure 18. Motor vehicle theft crime victimization rates by select combined TSE adjustment factors, bounding reference type, and survey year, 2011–2013

RTI Press Publication No. RR-0039-2007. Research Triangle Park, NC: RTI Press.

https://doi.org/10.3768/rtipress.2020.rr.0039.2007

should be adjusted for based on the same rationale as telescoping.

Although an initial descriptive review of victimization rates by data collection mode indicated a mode effect (Figure 1), a more thorough assessment resulted in findings more in line with Cernat (2015) than Holbrook et al. (2003). In other words, the apparent mode effect was really a function of the mixed-mode protocol and the predominate use of in-person interviews in circumstances where the estimates were unbounded rather than differences in the victimization rates. Therefore, if telescoping and fatigue are already addressed, a further adjustment for mode is not necessary.

Determining the Best Adjustment

This evaluation shows the current NCVS adjustment results in underestimated victimization rates. However, it is less clear how to adjust the victimization rates appropriately. The results of Evaluation Step 3 suggested that the alternative combined TSE adjustment factors lead to a wide range of adjusted estimates. In determining the most appropriate factor for the NCVS and other longitudinal surveys, it is necessary to consider (1) the bounding reference period, (2) the telescoping adjustment method, and (3) the legitimacy of a fatigue adjustment.

In terms of the bounding reference period, Evaluation Step 1 found that the reference period comparing TIS-1 with the other TISs (type 1) and the bounding reference period that compares any unbounded interview with bounded interviews regardless of TIS (type 4) behaved similarly to the adjustments based only on data from TIS 1–4 (Types 2 and 5) or TIS 1–2 (Type 3). For this reason, adjustment types 1 and 4 were considered for further analysis as they were based on the most data and allowed for more refined adjustments in further steps, particularly for the class- and model-based approaches. Type 1 is consistent with the approach used in the current NCVS bounding adjustment whereas type 4 is more theoretical in determining that all unbounded interviews should be treated the same. As seen in Evaluation Step 3, the annual victimization rates based on the type 4 bounding reference period are

generally lower than the estimates based on type 1. This is because more interviews have the downward telescoping adjustment applied. From a TSE perspective, if all bounded cases in TIS-2 through TIS-7 are upwardly adjusted for fatigue it follows that any unbounded cases should be adjusted as well. Therefore, the recommendation is to use bounding reference period type 4.

In terms of the telescoping adjustment method, it is less clear which is most appropriate. Because alternative combined TSE adjustments lead to a wide range of victimization rates and there is no known truth for the victimization rate—because it includes both reported and unreported victimizations—among the adjustments that meet our evaluation criteria in Evaluation Steps 1 and 2, the selection of the best telescoping adjustment is a bit subjective.

To assess the legitimacy of the fatigue adjustment, the analysis relied on the theoretical model rather than attempting to identify which respondents experienced fatigue. Because crime victimization is a rare event, lower numbers of victimizations in subsequent interviews do not necessarily indicate fatigue. Furthermore, some have argued (Cantor, 2007) that as respondents learn the NCVS instrument they answer the screener more accurately, reducing the number of erroneous victimizations reported. However, given the additional interview length that occurs when a respondent reports a victimization and decrease in the average number of incidents as interview number increases, it is highly plausible that some level of fatigue occurs.

Based on these considerations, it is likely that the NCVS is underestimating the rate of crime. Therefore, we recommend implementing an adjustment that accounts better for TSE (both telescoping error and fatigue) or, because there is uncertainty around the exact level of the adjustment, one possible option is to not apply an adjustment at all. Under either case, the resulting victimization rate is higher than the currently estimated rate. The unadjusted victimization rates were higher than the estimates with just the bounding adjustment and lower than any of the combined TSE adjusted victimization rates. This implies that the fatigue adjustment in the combined TSE adjustment was larger than the

telescoping adjustment. Therefore, if there is some question about the fatigue adjustment, the unadjusted estimates may be an acceptable alternative.

Limitations

The analysis relied solely on public use data files available through the National Archive of Criminal Justice Data. Because it was necessary to link households and persons across survey years, this meant that the analysis was limited to 2007 and later because IDs were re-scrambled in 2006. In addition, because of the reliance on public use data we focused our analysis on error sources we could directly measure ourselves. This led us to focus on errors caused by the panel design. Other error sources in the TSE framework exist in the NCVS, but they were out of this paper's scope. Furthermore, because of the complex nature of both telescoping and fatigue and the need for a 36-month reference window to support the model-based approaches it was only possible to compare 3 years of victimization rates (2011–2013). Including additional years of data could have allowed for more refined and precise models of telescoping and fatigue, which may have resulted in more conclusive trends.

Another limitation is that the proposed adjustments are done in an ad hoc manner with the post-survey adjusted public use file. Although it may be useful for survey methodologists or researchers using the

data to make their own individual adjustments as needed, given the number of error sources in any survey, expecting a user to decide which adjustments to make is too onerous. Instead, it would be better if these adjustments could be made by the survey owners (e.g., BJS in the case of the NCVS) prior to the creation of the public use file. Additionally, any adjustments made should be documented to allow user identification of data alterations and the impact of those alterations on the estimates.

Conclusions

When considering post-survey adjustments to account for non-sampling error, a TSE perspective should be used. Applying this perspective to the NCVS sample design, we determined that telescoping error and fatigue are likely sources of error whereas a mode effect caused by the mixed-mode design is not causing error in the survey estimates. Because it is not possible to identify which respondents were susceptible to telescoping or fatigue, we developed a model to show how victimization rates should behave across interview waves. The model was then used to develop adjustment methods to account for the sources of error and assess how well the adjustments met the model expectations. However, even with the model there is still some subjectivity that comes into determinations about the appropriate adjustment.

References

Berzofsky, M. E., & Carrillo-Garcia, I. A. (2014). Impact of changing the number of interview waves in the National Crime Victimization Survey. In *Proceedings of the Joint Statistical Meetings, Survey Research Methods Section* (pp. 2645–2659). Alexandria, VA: American Statistical Association; https://www.amstat.org/membersonly/proceedings/2014/data/assets/pdf/312516_89508.pdf

Berzofsky, M. E., & Carrillo-Garcia, I. A. (2019). Determining the optimal number of interview waves in a panel survey with application to the National Crime Victimization Survey. *Science Journal of Applied Mathematics and Statistics*, *6*(5), 135–147.

Berzofsky, M. E., Creel, D. V., Moore, A. S., Smiley McDonald, H. M., & Krebs, C. P. (2015). *Imputing NCVS income* (NCJR Library Collection; NCJ 248563). Washington, DC: National Criminal Justice Reference Service.

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. Paper presented at the Survey Research Methods Section, American Statistical Association, Alexandria, VA.

Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly*, *74*(5), 817–848. https://doi.org/10.1093/poq/nfq058

Cantor, D. (2007). A review and summary of studies on panel conditioning. In S. Menard (Ed.), *Handbook of longitudinal research: Design, measurement, and analysis* (pp. 124–137). Amsterdam: Academic Press.

Cernat, A. (2015). The impact of mixing modes on reliability in longitudinal surveys. *Sociological Methods & Research*, *44*(3), 427–457. https://doi.org/10.1177/0049124114553802

Couzens, G. L., & Berzofsky, M. E. (forthcoming). Assessing the impact of respondent fatigue in the National Crime Victimization Survey. In *Proceedings of the Joint Statistical Meeting, Survey Research Methods Section*. Alexandria: American Statistical Association.

Couzens, G. L., Berzofsky, M. E., & Krebs, C. P. (2014). Analyzing potential mode effects in the National Crime Victimization Survey. In *Proceedings of the Joint Statistical Meetings, Survey Research Methods Section* (pp. 2915–2923). Arlington: American Statistical Association; https://www.amstat.org/membersonly/proceedings/2014/data/assets/pdf/312744_89876.pdf

Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly*, *74*(5), 849–879. https://doi.org/10.1093/poq/nfq065

Hart, T. C., Rennison, M. C., & Gibson, C. (2005). Revisiting respondent "fatigue bias" in the National Crime Victimization Survey. *Journal of Quantitative Criminology*, *21*(3), 345–363. https://doi.org/10.1007/s10940-005-4275-4

Hasings, C., Mosteller, F., Tukey, J. W., & Winsor, C. P. (1947). Low moments for small samples: A comparative study of order statistics. *Annals of Mathematical Statistics*, *18*, 413–426.

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 respondents satisficing and social desirability bias. *Public Opinion Quarterly*, *67*(1), 79–125. https://doi.org/10.1086/346010

Rand, M., & Catalano, S. (2007). *Criminal victimization, 2006* (Bureau of Justice Statistics Bulletin; NCJ 219413). US Department of Justice, Office of Justice Programs. Retrieved from https://www.rainn.org/pdf-files-and-other-documents/News-Room/press-releases/2006-ncvs-results/NCVS%202006-1.pdf

Spencer, E. A., Brassey, J., & Mahtani, K. (2017). *Catalogue of bias collaboration: Recall bias*. Retrieved from https://www.catalogueofbiases.org/biases/recall-bias

US Census Bureau. (2014). *National Crime Victimization Survey: Technical documentation* (NCJ 247252). Retrieved from https://www.bjs.gov/content/pub/pdf/ncvstd13.pdf

US Department of Justice, Bureau of Justice Statistics. (2006). *National Crime Victimization Survey: Unbounded data, 2000* (ICPSR04445-v1). Conducted by US Dept. of Commerce, Bureau of the Census. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor].

US Department of Justice, Bureau of Justice Statistics. (2013). *National Crime Victimization Survey* (ICPSR35164-v1). Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/ICPSR35164.v1

Appendix

This appendix contains additional figures related to Evaluation Step 2 for the determination of the most appropriate combined TSE adjustment.

Figure A.1. Adjusted (black line) and unadjusted (gray line) violent victimization rates by household income, interview number, and bounding type
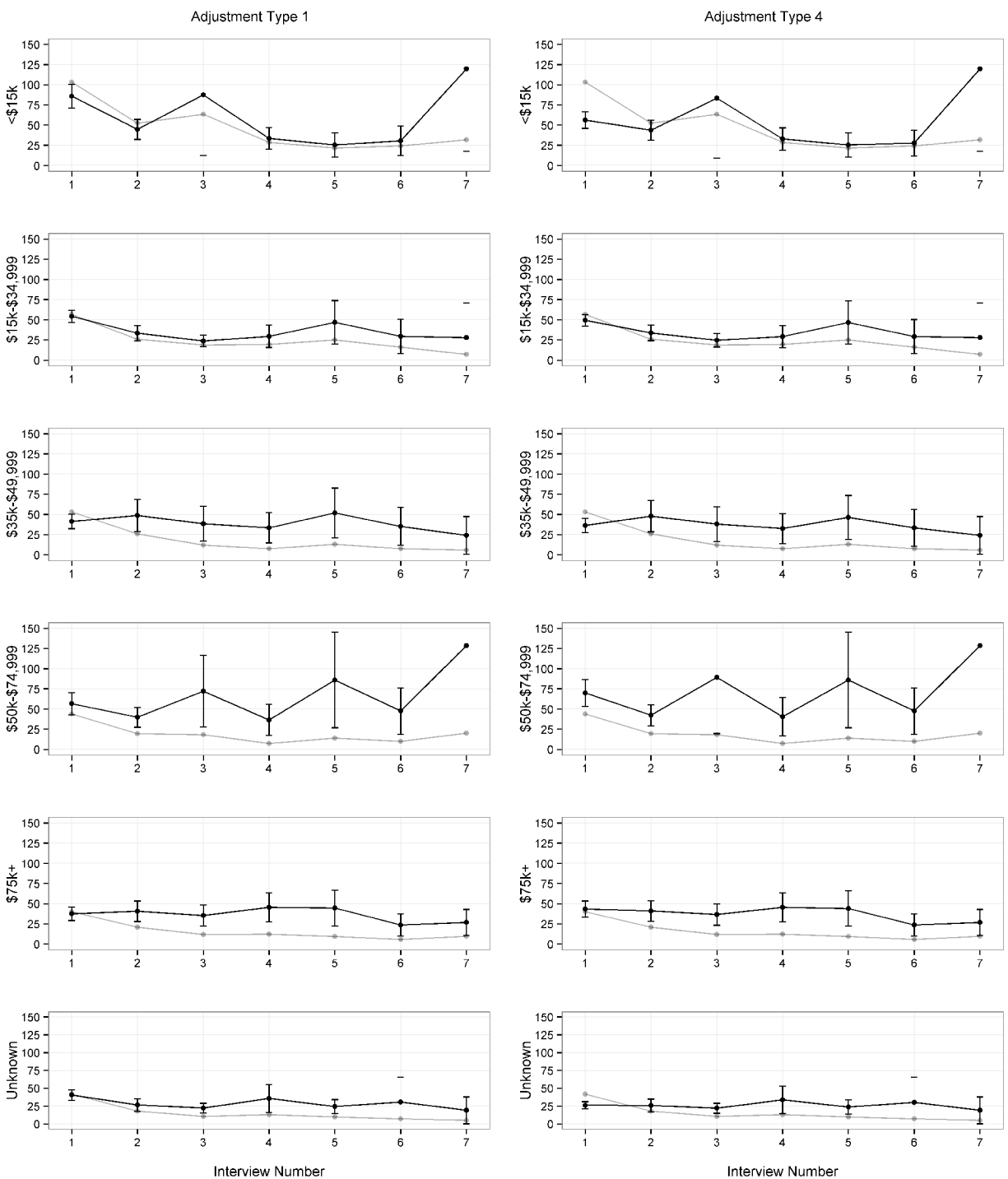
Figure A.2. Adjusted (black line) and unadjusted (gray line) violent victimization rates by presence of a weapon, interview number, and bounding type
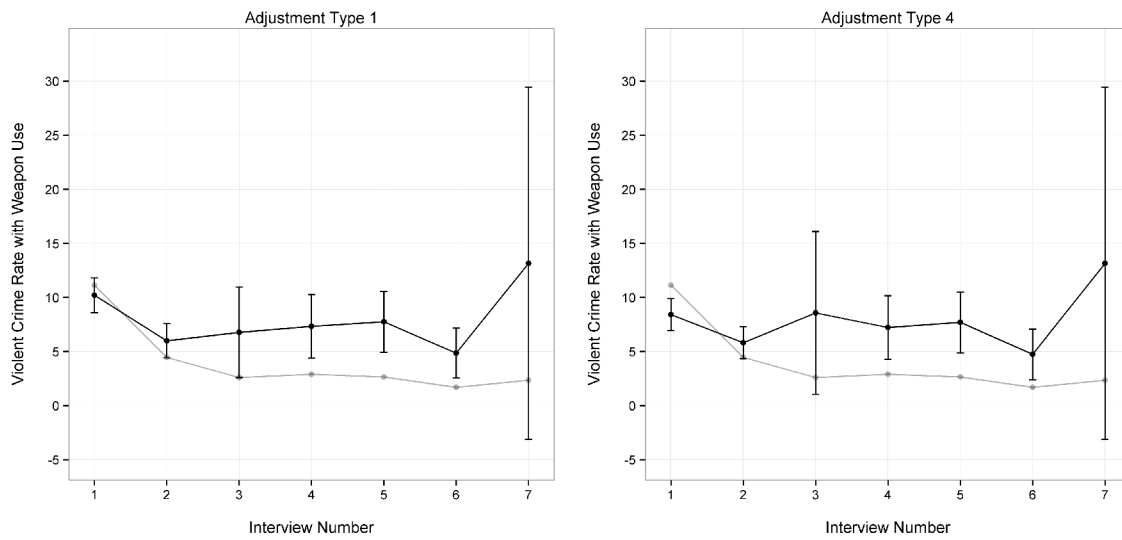
Figure A.3. Adjusted and unadjusted property victimization rates by urbanicity, interview number, and bounding type
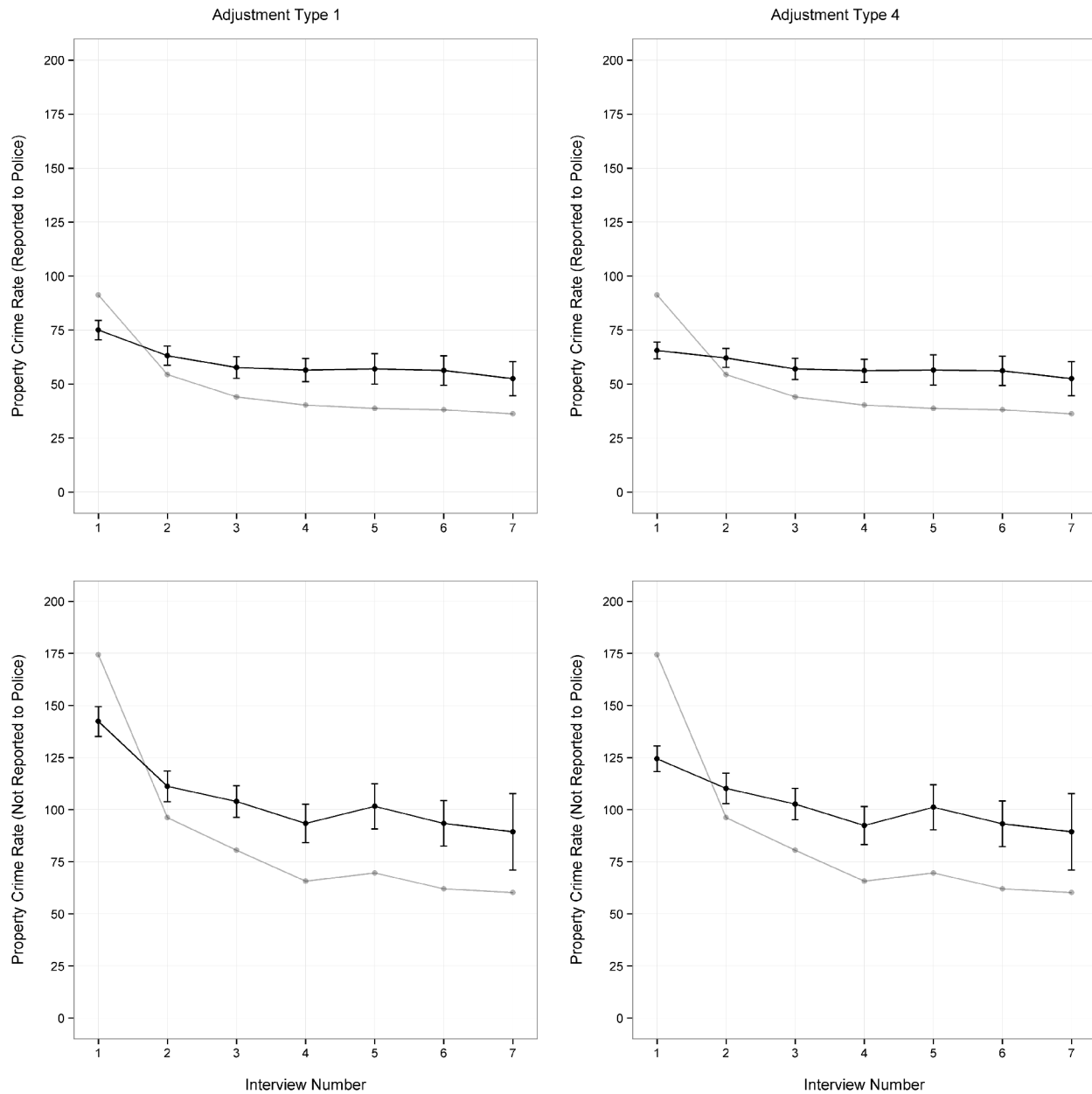
Figure A.4. Adjusted and unadjusted violent victimization rates by reporting to police status, interview number, and bounding type