DE GRUYTER OPEN

Journal of Official Statistics, Vol. 33, No. 2, 2017, pp. 551–570, http://dx.doi.org/10.1515/JOS-2017-0026

Effect of Missing Data on Classification Error in Panel Surveys

Susan L. Edwards[1], *Marcus E. Berzofsky*[2], *and Paul P. Biemer*[3]

Sensitive outcomes of surveys are plagued by wave nonresponse and measurement error (classification error for categorical outcomes). These types of error can lead to biased estimates and erroneous conclusions if they are not understood and addressed. The National Crime Victimization Survey (NCVS) is a nationally representative rotating panel survey with seven waves measuring property and violent crime victimization. Because not all crime is reported to the police, there is no gold standard measure of whether a respondent was victimized. For panel data, Markov Latent Class Analysis (MLCA) is a model-based approach that uses response patterns across interview waves to estimate false positive and false negative classification probabilities typically applied to complete data.

This article uses Full Information Maximum Likelihood (FIML) to include respondents with partial information in MLCA. The impact of including partial respondents in the MLCA is assessed for reduction of bias in the estimates, model specification differences, and variability in classification error estimates by comparing results from complete case and FIML MLCA models. The goal is to determine the potential of FIML to improve MLCA estimates of classification error. While we apply this process to the NCVS, the approach developed is general and can be applied to any panel survey.

Key words: Survey error; full information maximum likelihood; measurement error; Markov latent class analysis; national crime victimization.

1. Introduction

Social and behavior science researchers often collect data using questionnaires or instruments consisting of items that purport to measure some underlying construct that is difficult to measure accurately. For example, it is well known that employment status is difficult to measure because it relies on misunderstood concepts such as “looking for work,” “temporary layoff” versus “job termination,” “temporary work” versus “permanent employment,” and so on (see Biemer 2004). Employment classifications are typically based on responses to a series of questions that must be combined to categorize an individual as “employed,” “unemployed,” or “not in the labor force.” Because of the fine

[1] RTI International, 3040 East Cornwallis Road, Research Triangle Park, NC 27709, U.S.A. Email: sedwards@rti.org
[2] RTI International, 3040 East Cornwallis Road, Research Triangle Park, NC 27709, U.S.A. Email: berzofsky@rti.org
[3] RTI International, 3040 East Cornwallis Road, Research Triangle Park, NC 27709, U.S.A. Email: ppb@rti.org
Acknowledgments: The authors would like to thank the National Science Foundation (NSF) for sponsoring this research under award number 1229222. However, we would like to note that the views expressed in this article are those of the authors only and do not reflect the views or position of NSF.

distinctions among these categories or classes, misclassifications that lead to unstable and biased estimates of the class sizes are not uncommon.

A mixture modeling technique called Markov Latent Class Analysis (MLCA) can be used in panel surveys to correct the estimates for misclassification bias. It models wave-to-wave transitions and treats inconsistencies between the data and the model as measurement error or other model errors. MLCA provides estimates of the probabilities of misclassifying people in each labor category, the Wave 1 class probabilities, and the probabilities of transitioning from class to class across waves that have been corrected for misclassification.

A common problem in panel surveys that may limit this analysis is that some respondents fail to respond at one or more panel waves, resulting in an incomplete longitudinal record. This incompleteness poses a problem not only for MLCA but also for standard longitudinal modeling techniques that delete observations with missing time points and analyze only records with no missing values (referred to as case-wise deletion; see, for example, Allison 2001). Two different, although somewhat equivalent, modeling approaches are available to address this missing data problem: imputation and Full-Information Maximum Likelihood (FIML) estimation. One key difference between the two is that imputation replaces the missing values in the record with model-derived values to obtain a complete record that can then be used in a full data set estimation process. FIML, the focus of this article, obtains parameter estimates by maximizing the incomplete data likelihood using completely observed and partially observed cases; that is, all available (full) information. *Multiple* imputation (see, for example, Schafer and Graham 2002; Little and Rubin 2002) is an extension of single imputation that multiply-imputes each missing value to facilitate the computation of imputation variance. It has been shown in Allison (2012) that FIML is equivalent to multiple imputation in the limit as the number of imputations per missing value approaches infinity.

Equally as important as the choice of approach is the assumption that is made for the missing data mechanism itself. Assuming that the data are Missing Completely At Random (MCAR) will lead to bias inferences if response propensities are correlated with the classification error probabilities, which seems common (see, for example, Vermunt 1997; Hess et al. 2013). For example, Biemer (2004) showed that, in the Current Population Survey, people who misreport unemployment may tend to be nonrespondents whose information is often collected by proxy response. Likewise, people who under-report victimizations or who provide erroneous information about their victimizations may also be more likely to fail to respond at some panel wave.

This article demonstrates the importance of compensating for nonresponse in the Latent Class Analysis (LCA) of panel survey data, particularly when making inferences about the measurement components of the model. It shows the importance of including observations that contain missing values on some variables, not only for variance reduction, but also to reduce the bias. We will also show how it is possible to model data that are Missing At Random (MAR) using MLCA combined with FIML models.

Thus, the focus of this article is to explore the effects of methods for compensating for wave nonresponse on the classification error rates in each panel survey wave under the alternative assumptions about the nature of missing data. For this purpose, data collected between 2007 and 2013 from a long-standing national panel survey with indicators of violent and household-level crime victimization, the National Crime Victimization Survey (NCVS) (U.S. Department of Justice 2015), will be used to fit MLCA models for two types of victimizations: property crimes and violent crimes. Missing data will be modeled simultaneously in an MLCA model under MCAR and MAR missing data assumptions to address two key aims:

(1) Demonstrate the importance of using full information in modeling the structural and measurement components of an MLCA model by determining the effect that missing data have on the MLCA model determined to best fit the data.
(2) Evaluate the effects of alternative assumptions about the missing data mechanism (i.e., MCAR or MAR) on the estimates of misclassification and prevalence.

The remainder of this section provides a brief overview of MLCA models and the basic FIML approach to compensate for nonresponse. Section 2 describes the study data and modeling approach used to address the key aims of this article. In Section 3, the final MLCA model under each missing data mechanism is presented, along with estimates of classification error and crime victimization prevalence over time under MCAR and MAR missing data assumptions. The article concludes in Section 4 with a discussion of the differences across these models, their impact on classification error, thoughts on which model is most appropriate for the NCVS, and ideas for future analysis in this area. Although we apply this process to the NCVS, the approach we develop is general and can be applied to any panel survey.

1.1. Methods for Assessing Measurement Error in Panel Data

Markov Latent Class Models (MLCMs) adjust a panel survey's substantive estimates for the effects of misclassification and, as a byproduct of this process, produce estimates of the "response probabilities". In this application, response probabilities are referred to as classification error parameters because of the interpretation that the latent variable is the true classification. Rather than relying on external realizations of the true or "gold standard" values to estimate measurement error, MLCMs assume a model of the population structure and the measurement distribution parameters to provide maximum likelihood estimates of the parameters of this model. This approach was first introduced with cross-sectional data by Paul Lazarsfeld (1950) as LCA. In 1973, a modification of LCA, MLCA, was proposed by Wiggins (1973) to extend LCA techniques to panel data. Since then, MLCA methodology has been further developed by Poulsen (1982), Van de Pol and De Leeuw (1986), Van de Pol and Langeheine (1990), Dias and colleagues (2008), and Di Mari and colleagues (2016).

Using the notation in Biemer (2011), let X and Y denote two arbitrary random variables having values x and y, respectively. Denote $\Pr(X = x)$ by π_x^X and $\Pr(Y = y|X = x)$ by $\pi_{y|x}^{Y|X}$. Extensions of this notation to three or more variables are straightforward. The MLCM assumes that observations on a latent categorical variable X are subject to classification errors. These models require a minimum of three time points with each time point consisting of a latent variable and an indicator of that latent variable. Let the variable X_t denote the true value of the latent variable (X) at time t and let the observed value Y_t be an indicator of X_t. For purposes of this article, X_t and Y_t are assumed to have the same number of categories for all time points t. However, extensions to situations where the number of latent and manifest classes differ are straightforward.

The general MLCM contains two components: (1) the *structural component*, which describes the interdependencies between the X_t and the model covariates (referred to as grouping variables because they are categorical variables), and (2) the *measurement component*, which describes the interdependencies among the observations Y_t at each wave $t = 1,...,T$ and their interactions with X_t and other model covariates. Later in the article, a model employing four panel waves will be used in the analysis. However, to simplify the exposition, fix the ideas and establish the notation, here we present the model for three panel waves (i.e., $T = 3$) – the minimum number of panel waves for a MLCM to be identifiable. Extensions to four or more waves are straightforward.

The standard MLCM assumptions for three waves are as follows:

1. *First-Order Markov Property*. $\pi^{X_3|X_1X_2}_{x_3|x_1x_2} = \pi^{X_3|X_2}_{x_3|x_2}$ (i.e., a unit's latent state at Wave 3 (X_3), given its state at Wave 2 (X_2) is independent of its state at Wave 1 (X_1)).
2. *Independent Classification Errors (ICE)*. $\pi^{Y_1Y_2Y_3|X_1X_2X_3}_{y_1y_2y_3|x_1x_2x_3} = \pi^{Y_1|X_1}_{y_1|x_1}\pi^{Y_2|X_2}_{y_2|x_2}\pi^{Y_3|X_3}_{y_3|x_3}$ (i.e., classification errors for the three indicators are mutually independent across waves).
3. *Time-Invariant Classification Errors*. $\pi^{Y_t|X_t}_{y_t|x_t} = \pi^{Y|X}_{y|x}$ for $y = y_t, x = x_t, t = 1,2,3$; classification errors for the indicator Y_t are assumed to be the same for all waves $t = 1,2,3$.
4. *Group-Homogeneous Error Probabilities*. $\pi^{Y_t|X_t}_{y_t|x_t}$ for $t = 1,2,3$ is the same for all units in class $X_t = x_t$ (i.e., within the same latent class, individuals in the same class have equal misclassification probabilities).

Thus, the likelihood kernel for an MLCM with three time points with latent variables X_1, X_2, and X_3 with corresponding indicators Y_1, Y_2, and Y_3 and a single grouping variable G can be expressed as:

$$\mathcal{L}(\pi) = \pi^{GY_1Y_2Y_3}_{gy_1y_2y_3} = \pi^G_g \sum_{x_1,x_2,x_3}\left(\pi^{X_1|G}_{x_1|G}\pi^{X_2|GX_1}_{x_2|Gx_1}\pi^{X_3|GX_2}_{x_3|Gx_2}\right)\left(\pi^{Y_1|GX_1}_{y_1|Gx_1}\pi^{Y_2|GX_2}_{y_2|Gx_2}\pi^{Y_3|GX_3}_{y_3|Gx_3}\right) \quad (1)$$

where $\pi^G_g \sum_{x_1,x_2,x_3}(\pi^{X_1|G}_{x_1|g}\pi^{X_2|GX_1}_{x_2|gx_1}\pi^{X_3|GX_2}_{x_3|gx_2})$ is the structural component of the model and $\sum_{x_1,x_2,x_3}\pi^{Y_1|GX_1}_{y_1|gx_1}\pi^{Y_2|GX_2}_{y_2|gx_2}\pi^{Y_3|GX_3}_{y_3|gx_3}$ is the measurement component of the model with $\pi^{Y_t|GX_t}_{y_t|gx_t}$ representing the classification error probabilities at time t with $t = 1,2,3$.

The likelihood kernel presented in (1) can be expressed succinctly using Goodman's (1973) notation for hierarchical models, whereby the model terms for the structural, measurement and nonresponse (if applicable) components are specified in braces using only the highest order interactions. For example, in (1), the structural component can be expressed as a log-linear model $\{GX_1\ GX_1X_2\ GX_2X_3\}$, or as a modified path model as $\{X_1|\mathrm{G}\ X_2|X_1G\ X_3|X_2G\}$, and the measurement component as $\{GX_1Y_1\ GX_2Y_2\ GX_3Y_3\}$ or $\{Y_1|X_1G\ Y_2|X_2G\ Y_3|X_3G\}$. Thus Goodman's notation for the likelihood kernel presented in (1) may be expressed either as the log-linear form: $\{GX_1\ GX_1X_2\ GX_2X_3\}\{GX_1Y_1\ GX_2Y_2\ GX_3Y_3\}$ or the modified path model form: $\{X_1|\mathrm{G}\ X_2|X_1G\ X_3|X_2G\}\{Y_1|X_1G\ Y_2|X_2G\ Y_3|X_3G\}$. Goodman's notation will be used throughout the rest of the article because it is more succinct. Figure 1 graphically depicts this model in the form of a path diagram.

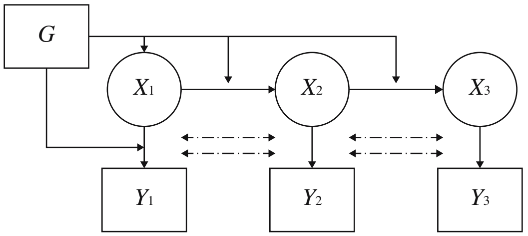

Fig. 1. Illustration of a Markov latent class model with one grouping variable, G. Double arrow denotes equivalence of the response probabilities.

At the tth wave, an indicator of the event (Y_t) is collected, which is a representation of the true value or latent construct X_t. In addition to measurement error, the indicators at Waves 2 and 3 are also subject to attrition (wave nonresponse) and item nonresponse. In Figure 1, circles represent the latent variables, squares represent manifest variables, and arrows denote relationships. An ignorable nonresponse mechanism, defined in more detail below, is assumed for the model.

1.2. Methods for Accounting for Wave Nonresponse in MLCA

When wave nonresponse exists in the indicators or item nonresponse exists in the grouping variables, then the exclusion of cases with one or both types of nonresponse may introduce bias into the model results. When dealing with nonresponse, it is important to understand the nonresponse mechanism and account for it appropriately. Nonresponse is often classified according to one of three missing data mechanisms: MCAR, MAR, or Missing Not At Random (MNAR), also referred to as "nonignorable." Originally defined by Rubin (1976), MCAR occurs when the missing data do not depend on the observed or unobserved data; MAR is less restrictive in that the missing data depend on only the observed data; MNAR is the least restrictive mechanism where the missing data depend on the unobserved data.

Recent work with cross-sectional data suggests benefits of using FIML techniques over listwise deletion, listwise deletion with reweighting, and hot deck imputation to fit a single hypothesized model. FIML methods provide better estimates of variance and are recommended when nonresponse is more than 5% and missing is dependent on the outcome (Allison 2012; Iannacchione 1982). FIML has shown promising results in LCA under a MAR and MNAR missing data mechanism to estimate inmate victimizations over complete case analysis, suggesting that respondents with missing indicators are more likely to be victims (Berzofsky, Biemer, Edwards 2015).

FIML can maintain unbiased inferences on the estimates (Graham 2009; Little and Rubin 2002; Enders 2010). For categorical data analysis, FIML approaches are similar to those developed to handle continuous data – partially observed information is used when fitting log-linear models under the assumption of a multinomial sampling distribution (Vermunt 1997). FIML can handle an MCAR, MAR, or MNAR missing data mechanism (Fay 1986; Vermunt 1997). However, handling MNAR missing data requires knowledge of the MNAR mechanism that is unobservable; this requirement leaves the researcher to formulate a model for the MNAR mechanism for which methods of testing have not been developed (Enders 2010).

In 1982, Fuchs (1982) extended the methodology of FIML to estimate the parameters of a saturated log-linear model using the Estimation-Maximization (EM) algorithm when nonresponse is MAR (Vermunt 1997), thus providing the fifth assumption for the models presented in this article:

5. *Nonresponse is Ignorable.* Nonresponse at each wave is "MAR" in the sense of Little and Rubin (2002).

Thus, the likelihood kernel for the MLCM detailed in Equation 1 can be modified to include dichotomous (0/1) response indicators R_1, R_2, and R_3 that correspond to indictors Y_1, Y_2, and Y_3, respectively, under a MAR mechanism, as follows:

$$\begin{aligned}\mathcal{L}(\pi) = {} & \pi^{GY_1Y_2Y_3}_{gy_1y_2y_3} \\ = {} & \pi^{G}_{g} \sum_{x_1,x_2,x_3} \sum_{r_1,r_2,r_3} \left(\pi^{X_1|G}_{x_1|G} \pi^{X_2|GX_1}_{x_2|Gx_1} \pi^{X_3|GX_2}_{x_3|Gx_2} \right) \left(\pi^{Y_1|GX_1}_{y_1|Gx_1} \pi^{Y_2|GX_2}_{y_2|Gx_2} \pi^{Y_3|GX_3}_{y_3|Gx_3} \right) \left(\pi^{R_1R_2R_3|GY_1Y_2Y_3}_{r_1r_2r_3|gy_1y_2y_3} \right) \qquad (2)\end{aligned}$$

where the terms $\pi^{R_1R_2R_3|GY_1Y_2Y_3}_{r_1r_2r_3|gy_1y_2y_3}$ determine the response mechanism assumed for the model. Under the assumption of ignorable nonresponse, the log likelihood function can be separated into two additive terms – one involving the parameters of the model in (1) and the other involving the nonresponse parameters. Thus, maximizing the likelihood associated with (1) will produce valid estimators of the MLCM.

In the case of MLCA, the default method for handling nonresponse in LatentGOLD (Vermunt and Magidson 2013) is Fuchs' approach for wave nonresponse and stochastic mean imputation for item nonresponse, making applying this technique straightforward and accessible to researchers. LatentGOLD 5.0 was used for all presented analyses.

2. Methods

2.1. Data: National Crime Victimization Survey

The NCVS is a nationally representative, probability-based household survey of the United States sponsored by the Bureau of Justice Statistics and conducted by the U.S. Census Bureau that gathers information on criminal victimization, reported and not reported to police (Truman and Morgan 2016). The NCVS incorporates a rotating panel design, which uses a stratified multistage cluster sample that includes roughly 50,000 households per sample group with each household interviewed every six months for a total of seven interviews. All households and people aged twelve or older in a rotation group are interviewed about the number and characteristics of victimizations experienced during the previous six months.

For this article, we focused on property crime and violent crime victimizations. For property crime, there is a single household respondent. For violent crime, each eligible person in the household responds. Because of the rareness of certain crimes and structure of the NCVS, multiple crime types were collapsed into a single violent or property victimization indicator at each wave. Violent crimes consisted of rape and sexual assault, aggravated assault, robbery, and simple assault; property crimes consisted of household burglary, motor vehicle theft, and theft. By collapsing, we gained model stability and avoided sparseness in the grouping classifications (Berzofsky and Biemer 2017).

The NCVS implements a two-phase approach to identify and enumerate victimizations. During the first phase of the interview, a screener is used to identify experiences with crime during the six-month reference period. The second phase of the interview is a detailed follow-up for each victimization identified during the screening phase. Indicators of specific types of victimization are created as a composite from various questions. Regarding household crimes, the respondent was asked about property break-ins or attempts and motor vehicle theft in the last six months in various scenarios (e.g., "did anyone steal gas from (it/them)"). At the person level, respondents are asked questions about victimization attacks and provided cues (e.g., location, weapon). For example, the question on theft with location cues was worded "since _____, were you attacked or threatened or did you have something stolen from you," and some of the cues provided were "at home including porch or yard" and "at work or school."

The amount of wave nonresponse observed in the crime victimization indicators at each wave of the study is more than 35% for violent crimes and less than 13% for household crimes. Wave nonresponse rates observed during the first four waves for the property and violent crime victimization indicators are presented in Table 1. Among typical reasons for nonresponse, the NCVS has two special considerations that may cause nonresponse during an individual wave. First, people may move out of a household. If an address is empty during the time of the interview, then the household and its members will have missing values for the wave. Second, new or newly eligible people may move into an existing household (e.g., a child turning twelve, a college graduate moving in with his or her parents at some point after the initial wave). In this case, the new or newly eligible person will have missing values for previous waves when they were either not in the household or ineligible. The NCVS does have unit-level response rates in the high 80% range at the person level (see, for example, Truman and Morgan 2016).

For our analysis, we limited the NCVS data to include panel and rotation groups for which all seven waves had occurred, resulting in data collected between 2007 and 2013. For these panel and rotation groups, all people and households in which at least one wave was completed were included in the analysis. Typically, multiple years of data would be pooled to reduce the standard errors of estimates, making the estimates more reliable. However, the NCVS public use files limit the number of years that can be pooled, because the household identifier was scrambled in 2006 when the new census primary sampling

Table 1. Crime victimization indicator wave nonresponse.

	Any	Wave 1	Wave 2	Wave 3	Wave 4
Violent					
Missing	110,236	58,935	58,960	58,520	56,955
Non-missing	51,635	102,936	102,911	103,351	104,916
Wave nonresponse rate (%)	68.10	36.41	36.42	36.15	35.19
Property					
Missing	47,713	8,037	7,929	8,560	8,779
Non-missing	34,678	56,423	57,339	58,360	59,434
Wave nonresponse rate (%)	57.91	12.47	12.15	12.79	12.87

units were integrated into the sample design. As MLCA requires linking households and people across time, the scrambling of the identifier limits the number of years that can be pooled. The issue of sparse cell sizes (i.e., model cells with zero or near-zero counts) can cause difficulties with model convergence (Biemer 2011; Bartolucci et al. 2013). Therefore, only the first four waves were used for the violent and property crime victimization analysis. This focus resulted in a total of 161,871 people and 68,213 households. Among these people, the number with an observed violent crime victimization was less than 1.5 percent, and among these households, the number with an observed property crime victimization was less than nine percent. Observed crime victimization prevalence is presented in Table 2.

It is expected that the initial interview would have larger victimization rates compared with the later waves because it is unbounded and respondents may "telescope" by recalling incidents that occurred before the six-month reference period. Despite this consideration, Wave 1 data were included to be consistent with the NCVS, which, beginning in 2006, included Wave 1 responses in the published estimates (Rand and Catalano 2007). Data gathered in Wave 2 may be considered the most accurate because they are from the first bounded interview with the least amount of fatigue; however, being the first bounded interview does not imply a gold standard because the data can still suffer from other sources of measurement error (e.g., interviewer bias, questionnaire wording).

2.2. *Modeling Approach*

We followed the modeling strategy that worked best on most tested models as discussed by Berzofsky and Biemer (2017) (see also Biemer 2011), which consisted of two main steps. First, grouping variables were identified with a forward selection approach using the Bayesian Information Criterion (BIC) to identify when each grouping variable should

Table 2. Observed crime victimizations in the NCVS.

	Wave			
	1	2	3	4
Violent victimization				
Unweighted				
Victims	1,295	844	801	710
Non-victims	101,641	102,067	102,550	104,206
Weighted				
% Victimization	1.36	0.89	0.83	0.73
Standard error	0.05	0.04	0.03	0.03
Property victimization				
Unweighted				
Victims	5,199	3,524	3,299	3,173
Non-victims	59,261	61,744	63,621	65,040
Weighted				
% Victimization	8.19	5.48	4.99	4.68
Standard error	0.17	0.12	0.12	0.10

enter the model. Grouping variables create mutually exclusive groups whereby the classification error rates are homogenous within each group; grouping variables are further discussed in the following paragraph. These variables were added to the structural and measurement models. Likelihood ratio tests were used to determine the most parsimonious base model that removed group heterogeneity and met the MLCM assumptions: first-order Markov, ICE, time-invariant classification errors, and group-homogeneous error probabilities. Second, using the base model from step 1, all remaining MLCM assumptions were tested and relaxed according to the following procedure: (1) models with boundary or convergence issues that might make the model unstable were rejected, and (2) for models without estimation issues, results from likelihood ratio tests for nested models and BIC for non-nested models were used to select the final model.

The NCVS collects information on 14 grouping variables: twelve personal or household-level variables and two para-data variables (U.S. Department of Justice 2015). These 14 grouping variables formed the foundation of grouping variables considered for our models. Grouping variables were classified as time varying or time invariant (Bartolucci et al. 2013). Time-invariant grouping variables were those where fewer than five percent of respondents changed status from the first observed value to the last observed value; age category was an exception to this rule. Time-invariant grouping variables were defined by the first observed value. To reduce the complexity of the model and get parsimony without sacrificing fit, time-varying grouping variables were defined by the creation of an additional category to capture the "movers" who, regardless of movement direction, had similar classification error rates (Berzofsky and Biemer 2017). Because of low item nonresponse rates in all but one of the grouping variables (less than four percent), grouping variables were imputed before MAR analysis with a stochastic mean imputation technique, the default imputation method for covariates in LatentGOLD. Grouping variables considered for the violent and property victimization models with item nonresponse rates are detailed in Table 3.

One challenge of conducting MLCA with complex survey data is that one or more assumptions may be violated because of the sample design (Biemer 2011). For the structural component assumptions, first-order Markov models were tested against second-order Markov models, models where transition probabilities are assumed to depend on the previous two time points, and Mover-Stayer models, models with an additional latent construct to identify persons or households whose victimization status is constant (stayer) or changes (mover) over time (Goodman 1961). Time-invariant classification error rates were tested by relaxing assumptions on the coefficients for each time point. For the measurement component assumptions, group-homogeneous error probabilities were tested by relaxing assumptions on the coefficients for each indicator; ICE assumptions were tested using bivariate residual analysis to identify dependent indicators (Vermunt and Magidson 2013).

Table 4 highlights the various models that were compared using Goodman's notation for hierarchical models. In Table 4, X_1 to X_4 represent the latent construct of victimization (violent or property) at each wave; Y_1 to Y_4 represent indicator 1 through indicator 4, respectively; A represents marital status, B represents age category, C represents household ownership, D represents household size category, E represents age category of the oldest person in the household, F represents urbanity, and M is a latent construct to capture movement.

Table 3. Crime victimization grouping variable item nonresponse.

	Missing	Non-Missing	Item Nonresponse Rate (%)
Violent			
Age category[1,3]	0	161,871	0.00
Education[1]	3,756	158,115	2.32
Gender[1]	0	161,871	0.00
Household size category[2]	5,437	156,434	3.36
Household ownership[1,3]	0	161,871	0.00
Interview type (in person/phone)	0	161,871	0.00
Marital status[1,3]	1,506	160,365	0.93
Number of in person interviews	0	161,871	0.00
Proxy answered interview	0	161,871	0.00
Race category[1]	0	161,871	0.00
Urbanity[1]	0	161,871	0.00
Property			
Age category for oldest in household[1,3]	0	82,391	0.00
Household income[2]	18,768	63,623	22.78
Household size category[2,3]	0	82,391	0.00
Household ownership[1]	0	82,391	0.00
Interview type – all in person	0	82,391	0.00
Interview type – all/some/none in person	0	82,391	0.00
Number of in person interviews	0	82,391	0.00
Race category for oldest in household[1]	0	82,391	0.00
Urbanity[1,3]	0	82,391	0.00

[1] First observed value used for analysis.
[2] Time varying variable with single "mover" category.
[3] Grouping variable used in violent or property victimization model.

Once the final model was determined from MCAR and MAR analysis according to the approach detailed previously, models were fit to each category of victimization – violent and property. Each model was applied to two data sets:

(1) MCAR analysis using complete case data (e.g., listwise deletion) and
(2) MAR analysis using the Fuchs FIML approach on the outcome (victimization) and mean imputation on the grouping variables.

Thus, a total of four models were used to address the aims of this article:

(1) violent victimization MCAR model applied to the person-level MCAR data set,
(2) violent victimization MAR model applied to the person-level MAR data set,
(3) property victimization MCAR model applied to the household-level MCAR data set, and
(4) property victimization MAR model applied to the household-level MAR data set.

LatentGOLD software was used for all analyses in this report; LatentGOLD addresses the issue of clustering and weighting through a pseudo-maximum likelihood technique and

Table 4. Models considered.

	Violent models	Property models
Base grouping variable model with all MLC assumptions	$\{X_1A\ X_1B\ X_1C\ X_tX_{t-1}A\ X_tX_{t-1}B\ X_tX_{t-1}C\}$ (for t = 2, 3, 4) $\{Y_tX_tA\ Y_tX_tB\ Y_tX_tC\}$ (for t = 1, 2, 3, 4)	$\{X_1D\ X_1E\ X_1F\ X_tX_{t-1}D\ X_t\ X_{t-1}E\ X_tX_{t-1}F\}$ (for t = 2, 3, 4) $\{Y_tX_tD\ Y_tX_tE\ Y_tX_tF\}$ (for t = 1, 2, 3, 4)
Time-invariant classification error assumption relaxed	$\{X_1A\ X_1B\ X_1C\ X_tX_{t-1}A\ X_tX_{t-1}B\ X_tX_{t-1}C\}$ (for t = 2, 3, 4) $\{Y_tX_tA\ Y_tX_tB\ Y_tX_tC\}$ (for t = 1, 2, 3, 4)	$\{X_1D\ X_1E\ X_1F\ X_tX_{t-1}\ X_tX_{t-1}D\ X_tX_{t-1}E\ X_tX_{t-1}F\}$ (for t = 2, 3, 4) $\{Y_tX_tD\ Y_tX_tE\ Y_tX_tF\}$ (for t = 1, 2, 3, 4)
First-order markov property assumption relaxed		
Second order markov	$\{X_1A\ X_1B\ X_1C\ X_2X_1A\ X_2X_1B\ X_2X_1C\ X_tX_{t-1}X_{t-2}A\ X_tX_{t-1}X_{t-2}B\ X_tX_{t-1}X_{t-2}C\}$ (for t = 3, 4) $\{Y_tX_t\ Y_tX_tA\ Y_tX_tB\ Y_tX_tC\}$ (for t = 1, 2, 3, 4)	$\{X_1D\ X_1E\ X_1F\ X_2X_1D\ X_2X_1E\ X_2X_1F\ X_tX_{t-1}X_{t-2}D\ X_tX_{t-1}X_{t-2}E\ X_tX_{t-1}X_{t-2}F\}$ (for t = 3, 4) $\{Y_tX_tD\ Y_tX_tE\ Y_tX_tF\}$ (for t = 1, 2, 3, 4)
Mover stayer	$\{X_1A\ X_1B\ X_1C\ X_1M\ X_tX_{t-1}A\ X_tX_{t-1}B\ X_tX_{t-1}C\ X_tX_{t-1}M\}$ (for t = 2, 3, 4) $\{Y_tX_tA\ Y_tX_tB\ Y_tX_tC\}$ (for t = 1, 2, 3, 4)	$\{X_1D\ X_1E\ X_1F\ X_1M\ X_tX_{t-1}D\ X_tX_{t-1}E\ X_tX_{t-1}F\ X_tX_{t-1}M$ (for t = 2, 3, 4) $\{Y_tX_tD\ Y_tX_tE\ Y_tX_tF\}$ (for t = 1, 2, 3, 4)
Group-homogeneous error probabilities assumption relaxed		
Wave 1 different	$\{X_1A\ X_1B\ X_1C\ X_2X_1A\ X_2X_1B\ X_2X_1C\ X_tX_{t-1}X_{t-2}A\ X_tX_{t-1}X_{t-2}B\ X_tX_{t-1}X_{t-2}C\}$ (for t = 3, 4) $\{Y_tX_t\ Y_tX_tA\ Y_tX_tB\ Y_tX_tC\}$ (for t = 1, 2, 3, 4)	$\{X_1D\ X_1E\ X_1F\ X_1M\ X_tX_{t-1}D\ X_tX_{t-1}E\ X_tX_{t-1}F\ X_tX_{t-1}M\}$ (for t = 2, 3, 4) $\{Y_tX_tD\ Y_tX_tE\ Y_tX_tF\}$ (for t = 1, 2, 3, 4)
All waves different	$\{X_1A\ X_1B\ X_1C\ X_2X_1A\ X_2X_1B\ X_2X_1C\ X_tX_{t-1}X_{t-2}A\ X_tX_{t-1}X_{t-2}B\ X_tX_{t-1}X_{t-2}C\}$ (for t = 3, 4) $\{Y_tX_t\ Y_tX_tA\ Y_tX_tB\ Y_tX_tC\}$ (for t = 1, 2, 3, 4)	$\{X_1D\ X_1E\ X_1F\ X_1M\ X_tX_{t-1}D\ X_tX_{t-1}E\ X,X_{t-1}F\ X_tX_{t-1}M\}$ (for t = 2, 3, 4) $\{Y_tX_tD\ Y_tX_tE\ Y_tX_tF\}$ (for t = 1, 2, 3, 4)

MLC = Markov latent class.

addresses nonresponse through applying FIML and stochastic mean imputation to categorical data analysis. The ability to apply Fuchs' FIML approach is built into the software as the default method for addressing nonresponse on the dependent variables. For independent variables, LatentGOLD applies stochastic mean imputation by default. In regard to MLCA, FIML is used to address wave nonresponse in the indicators, and stochastic mean imputation is used to address item nonresponse in the grouping variables.

3. Results

The aims of this article are (1) to demonstrate the importance of using full information in an MLCM and (2) to evaluate the effect MCAR and MAR missing data assumptions have on MLCA model estimates of misclassification and prevalence. Subsection 3.1 provides details on the model fitting process and final models used in our analysis for both victimizations (violent and property). Subsection 3.2 compares estimates of misclassification from MCAR and MAR MLCMs for each type of victimization. Subsection 3.3 compares prevalence estimates from the structural component of MCAR and MAR MLCMs for each type of victimization.

3.1. Modeling Results

With respect to victimization type, models with and without missing data identified the same grouping variables and relaxed the same MLCA assumptions, resulting in identical final models. Table 5 presents victimization model diagnostics for violent and property crime victimization. The dissimilarity index indicates the percentage of data that would need to change cells for the model to fit perfectly; it is an alternative way to assess the fit of the model. Full measurement models for violent and property crime victimization are given in Supplemental data, Appendix A (available online at: http://dx.doi.org/10.1515/JOS-2017-0026). Complete LatentGOLD syntax for model estimation of violent and property crime victimizations is given in Supplemental data, Appendix B (available online at: http://dx.doi.org/10.1515/JOS-2017-0026). Subsections 3.1.1 and 3.1.2 provide specific details on the base and final models for violent and property crime victimization, respectively.

3.1.1. Violent Crime Victimization Modeling Results

The violent crime victimization final model without missing data (i.e., after listwise deletion, respondents without missing indicators or grouping variables) included 51,528 cases, 31.8% of all respondents. The base model found three grouping variables to be significant in the measurement component of the MLCM – first observed value of marital status, household ownership, and first observed value of categorized age. When missing data were included, the same grouping variables were found to be significant. The identified grouping variables are listed in Table 3.

For violent crime victimizations, a full model with interaction terms between the grouping variables and the latent wave indicator of victimization status was deemed appropriate, and several MLCA assumptions were relaxed, regardless of the missing data assumption. Our models were able to relax model assumptions because four time points were used in the models. Based on the bivariate residual test, the ICE assumption was not

Table 5. Model fitting statistics.

Attribute of model	Missing data approach	# of parameters	Degrees of freedom	Log-likelihood	BIC	Dissimilarity index[1]
VIOLENT VICTIMIZATION						
Base model	MCAR	35	325	−7161	14703	0.0029
	MAR	35	1812	−20439	41299	0.0075
Final model	MCAR	93	267	−7099	15207	0.0019
	MAR	93	1754	−20363	41841	0.0071
PROPERTY VICTIMIZATION						
Base model	MCAR	45	630	−38196	76878	0.0166
	MAR	45	2594	−56442	113393	0.0392
Final model	MCAR	101	574	−38147	77384	0.0146
	MAR	101	2538	−56367	113876	0.0370

[1]Formula for dissimilarity index. $D = \sum_i |n_i - \hat{m}_i|/(2N)$, where n_i = observed cell count, $\hat{m}_i$ = estimated expected cell count, N = # of observations, i = cell identifier.

violated. The final MCAR and MAR violent victimization models consisted of a second-order Markov model with varying covariates for the observed victimizations and varying classification errors between the first wave and all following waves.

3.1.2. Property Crime Victimization Modeling Results

The property crime victimization final model without missing data included 48,590 cases, 59.0% of all responding households. The base model found three grouping variables to be significant in the measurement component of the MLCM – categorized household size, first observed value of categorized age of oldest household member, and urbanity. As with violent crime victimization when missing data were included, the same grouping variables found to be significant in the MCAR models were also identified in the MAR models.

The property crime victimization model experienced similar MLCM assumption violations as the violent crime victimization model. The base model for property crime victimization consisted of a full model with main effects and interaction terms between the grouping variables and the latent wave indicator of victimization status. The time homogeneous classification error, first-order Markov assumptions, and group homogeneous classification error assumptions were relaxed. The final MCAR and MAR property crime models consisted of a mover-stayer full model with varying covariates for the observed victimizations and varying classification errors between the first wave and all following waves.

3.2. *Estimated Misclassifications*

Now we use the second-order MLCM and the mover-stayer MLCM to create estimates of misclassification and prevalence by fitting the measurement and structural components with each missing data assumption (MCAR, MAR). The measurement component provides estimates of false positive and false negative rates at each time point. False positive rates measure the probability of respondents identifying as victims when in truth they are nonvictims (i.e., $P(Y_t = 1|X_t = 2)$). False negative rates result from respondents identifying as nonvictims when in truth they are victims (i.e., $P\left(Y_t = 2|X_t = 1\right)$). Trends of estimated false positive and false negative rates for violent and property crime victimization at each wave of the NCVS are presented in Figures 2 and 3, respectively, with 95% confidence intervals represented by error bars.

From Figure 2, it is clear that regardless of model type, the false positive rates for violent victimizations are larger for the first interview. These larger rates are probably the

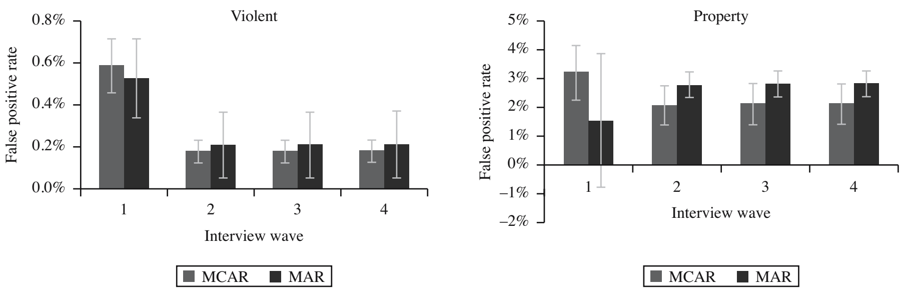

Fig. 2. False positive rates for violent and property crime victimization.

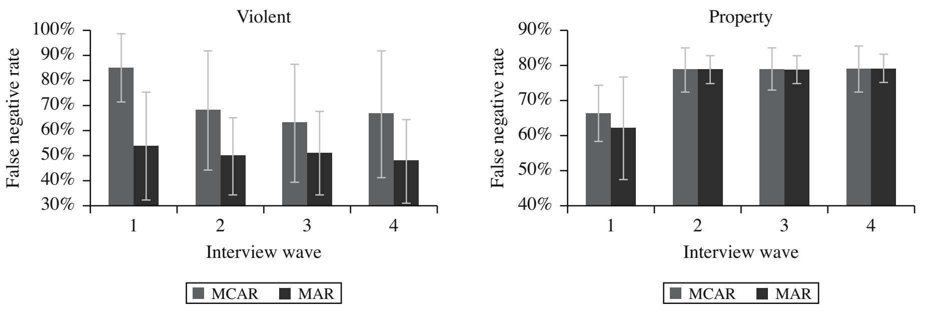

Fig. 3. False negative rates for violent and property crime victimization.

result of telescoping (as noted earlier, the first interview used for estimation is unbounded, whereas all follow-up interviews are bounded). This finding also held for the MCAR property victimization model, but not the MAR property victimization model. False positive rates are low regardless of model and victimization type: less than one percent for violent and less than five percent for property victimizations.

Based on false positive rates, victimization type appears to affect the results from the MCAR and MAR models after the first wave in different manners. Both MAR models yielded higher estimates of false positive rates compared to the corresponding MCAR estimates. For violent crimes, the MCAR and MAR models yield similar estimates. For property crimes, the MAR model yields estimates near the upper end of the 95% confidence interval of the MCAR model estimates; MAR estimates for false positive rates are larger than the MCAR estimates at every wave, except Wave 1, by roughly 0.7%.

From Figure 3, the manner in which false negative rates change over time for violent victimizations differs depending on the mechanism for missing response. Under the MCAR model, the false negative rate is significantly higher in the first interview (85%) compared with the later interviews (≈65%); however, for the MAR model, the false negative rate is statistically unchanged across the four periods (≈51% in all waves). This result is an indication that the inclusion of those who do not respond helps control for differences in the false negative rate over time. For property victimization, although there appears to be an increase in the false negative rate from interview Wave 1 (66% for MCAR and 62% for MAR) compared with the later waves (≈80% for all waves for MCAR and MAR) regardless of the missing data mechanism, these differences are not statistically significant.

Interestingly, our results do not detect an increase in the false negative rate in interview wave 4. Some research (see, for example, Hart et al. 2005) has shown that respondent fatigue occurs in later waves of the NCVS. Respondent fatigue is likely to increase the classification error rates over time. One possible reason that our models do not demonstrate this pattern is because we limited our analysis to the first four interview waves. Hart and colleagues (2005) looked at all seven waves, finding respondent fatigue to have its greatest effects in Waves 6 and 7, which are not included in our current analysis.

Perhaps due to the less sensitive nature of property crimes and a more engaged respondent, the estimated false negatives for property crime victimization at each wave of the NCVS by model type show different trends than those for violent crimes. Estimates

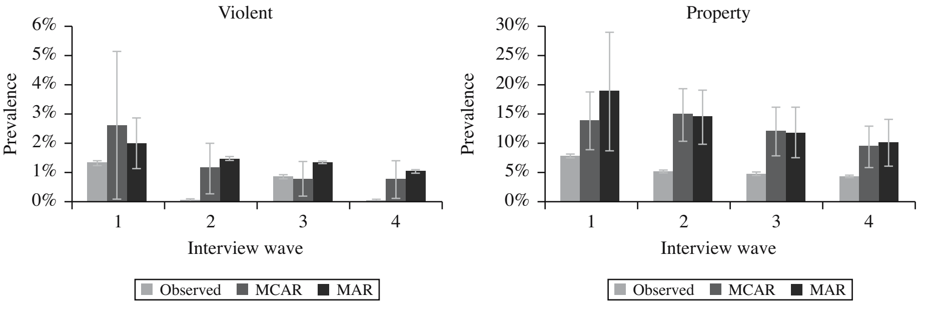

Fig. 4. Prevalence rates for violent and property crime victimization.

differed by 4.1% at the first wave but were similar during following waves, with MAR estimates being slightly higher by at most 0.2%. The false positive and false negative rate estimates are somewhat consistent across waves with respect to model type.

3.3. Estimated Prevalence

Victimization prevalence is measured in the structural component of the MLCM. Estimates of violent and property victimization were computed three ways:

(1) based on observed responses (i.e., direct estimates from the data set),
(2) based on the MCAR model, and
(3) based on the MAR model are presented in Figure 4, with 95% confidence intervals represented by error bars.

As with the classification error rates, standard errors are larger for the FIML methods compared with the complete case analysis. If the missingness is MCAR, then we would expect the standard errors from the MAR model to be smaller, because the MAR model uses more information than the MCAR model. If the missingness is MAR, increased standard errors are to be expected because missing values contribute more variance to the final model. Estimates differ between model types for violent crime victimizations, suggesting that missing data do affect estimates of prevalence. FIML models estimate violent crime prevalence to be higher than the observed and complete case analysis. Prevalence rates for either victimization are highest during the first wave; this finding may be attributed to telescoping because the initial wave is unbounded, which inflates the number of reported victimizations (U.S. Census Bureau 2014).

Overall estimates of property crime victimization are similar across model types, with the MAR model differing the most from the MCAR model during the first wave by 5.1%. For violent crime victimization, for all but the first wave, the MAR estimates are higher than the MCAR estimates. The FIML MAR model appears to be correcting for respondent fatigue by keeping the violent crime prevalence consistent across waves.

4. Discussion

In this article, we fit two different types of models for the response mechanism in NCVS data. One model (MCAR) was fit in a complete case analysis that included only records with no missing values on all the victimization indicators or grouping variables across all four waves. This model excluded about 70% of the cases for violent victimization and about 40% of the cases for property victimization. The other model (MAR) used FIML techniques to account for missing data in the indicators and mean imputation to account for missing data in the grouping variables. This model included all cases that responded in one or more waves. Estimates of classification error rates and prevalence rates were produced from both models. The MAR model attempts to compensate for any bias that could be introduced into the analysis by excluding the missing cases. MAR models assume that the missing data mechanism does not depend on the variable that is missing but may depend on other influencing factors that can be modeled using additionally observed variables.

A third type of missing data mechanism, MNAR, can also be modeled using FIML techniques. This type of model assumes that the missing data mechanism associated with the outcome variable (i.e., victimization indicator) depends on that same outcome variable. However, MNAR FIML models are difficult to apply with existing software, and there are trade-offs in doing so. MNAR model estimates will often have larger variances, which may offset any gains in reducing nonresponse bias. The software we used experienced issues with EM convergence and local minima, leading to model instability. Besides being more difficult to program, MNAR models that may be specified can be limited by the computer's memory capabilities. In our case, 16 gigabytes of RAM were not sufficient to run some models. As a result, the MNAR models we fit resulted in implausible estimates, which were most likely due to weak identifiability and local minima (Bartolucci et al. 2013; Biemer 2011). Given the poor performance of the MNAR models, those results were not included in this article.

MAR estimates that differ considerably from MCAR estimates usually indicate that the MCAR assumption is untenable; thus, excluding the cases with missing data from the analysis will yield biased estimates. For violent and property crime, MAR models produced substantially different estimates from the MCAR models. For violent crime victimization, FIML MAR estimates of prevalence were higher than MCAR estimates at all but the first wave. For property crime victimization, MAR and MCAR estimates of prevalence were similar in all but the first wave. This result suggests that nonrespondents are more likely to be victims of violent crime but perhaps not property crime.

As previously noted, the purpose of this article was to demonstrate that MLCMs can be used to account for measurement error and nonresponse and to evaluate the differences between MLCMs with and without missing data. However, further research is needed. For example, the nonresponse bias implied by the MAR models for violent crimes presents an intriguing finding, namely, omitting respondents with wave and/or item nonresponse from the analysis of violent crime could substantially bias the results. These models would benefit from further refinement and verification. Although item nonresponse was minimal in our final models, future research could treat the two types of nonresponse (wave and item) differently because the mechanism that causes an individual to opt out at a wave may be different than that which causes an individual to not respond to an item in the interview.

One opportunity to develop qualitative research to support our findings would be through cognitive interviewing. We hypothesize that much of the measurement error from wave to wave is due to comprehension error, recall error, or respondent fatigue (also known as “satisficing”). Perhaps evidence of these errors can be found using cognitive interviewing techniques where respondent conditions that give rise to these error sources could also be explored. Another method for verifying our findings would be via a simulation study. Using Monte Carlo simulation, multiple data sets, each with a unique and known nonresponse mechanism, could be generated from the current NCVS data to determine how various types of nonresponse errors manifest themselves as biases in victimization estimates. In addition, the simulations could also investigate the extent to which measurement errors affect the application of MAR and MNAR nonresponse models. A variation on the simulation study could explore the validity of the models by generating data sets with varying levels of nonresponse and measurement errors. Then the results from the models could be compared with the known model-generating parameters.

The data themselves presented a few unique challenges. Because of the design of the NCVS, it is difficult to pool data from a larger time period. We could pool only panels that started data collection in 2007 because of issues of household and person ID linkage related to the scrambled household identifiers introduced in 2006. This small sample could be contributing to the large standard errors observed for the model-based estimates. Pooling data from a larger time period would increase the sample size, which could then result in more stable models.

The problems with small samples were compounded by the fact that crime victimization is a rare event, which resulted in few positives in the sample on which to build a model for misclassification of positives. We addressed this issue by combining crime types into two distinct categories – property crimes and personal crimes – to build up their prevalence. In 2014, the overall rate of violent crimes was 20.1 per 1,000 people aged twelve or older; rape and sexual assault crimes accounted for just 1.1 per 1,000 people aged twelve or older of the overall rate (Langton and Truman 2015). For this reason even with 15 years of data, standard errors could still be large, particularly for false negative estimates. In addition, analyzing 15 years of data may expose other issues such as temporal changes in definitions of certain types of crime or crime reporting over time. Our analysis excluded data collected from the later waves (i.e., Waves 5, 6, and 7). This exclusion was done primarily to reduce the number of sparse cells due to cross-classifying responses from seven waves, which could be compounded because of potentially greater respondent fatigue in later waves.

The goal of this analysis is not to quantify all of the errors present in the NCVS, but to show a model-based way of addressing two types of errors and the effect nonresponse can have on model estimates. Despite the limitations of the data, our findings demonstrated that excluding respondents with missing data may bias estimates of prevalence.

5. References

Allison, P.D. 2001. “Missing Data.” In *Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-136*. Thousand Oaks, CA: Sage.

Allison, P.D. 2012. “Handling Missing Data by Maximum Likelihood.” In Proceedings of SAS Global Forum 2012, Statistics and Data Analysis, April 22–25, 2012. 312. Haverford, PA: SAS Institute. Available at: http://www.statisticalhorizons.com/wp-content/uploads/MissingDataByML.pdf (accessed August 2016).

Bartolucci, F., A. Farcomeni, and F. Pennoni. 2013. *Latent Markov Models for Longitudinal Data*. Boca Raton, FL: CRC Press.

Berzofsky, M.E., P.P. Biemer, and S.L. Edwards. 2015. "Latent Class Analysis with Missing Data under Complex Sampling: Results of a Simulation Study." Presented at 60th World Statistics Conference, July 26–31, 2015. Rio de Janeiro, Brazil: World Statistics Conference.

Berzofsky, M. and P.B. Biemer. 2017. "Classification Error in Crime Victimization Surveys: A Markov Latent Class Analysis." In *Total Survey Error in Practice*, edited by P.P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L.E. Lyberg, N.C. Tucker, and B.T. West, 387–412. Hoboken, NJ: Wiley.

Biemer, P.P. 2004. "An Analysis of Classification Error for the Revised Current Population Survey Employment Questions." *Survey Methodology* 30(2): 127–140.

Biemer, P.P. 2011. *Latent Class Analysis of Survey Error*. Hoboken, NJ: Wiley.

Di Mari, R., D.L. Oberski, and J.K. Vermunt. 2016. "Bias-Adjusted Three-Step Latent Markov Modeling with Covariates, Structural Equation Modeling." *Structural Equation Modeling* 23(5): 649–660. Doi: http://dx.doi.org/10.1080/10705511.2016.1191015.

Dias, J.G., J.K. Vermunt, and S. Ramos. 2008. "Heterogeneous Hidden Markov Models." In *Compstat 2008 Proceedings*, August, 2008. City, State: Compstat. Available at: http://members.home.nl/jeroenvermunt/dias2008.pdf (accessed March 2015).

Enders, C.K. 2010. *Applied Missing Data Analysis*. New York: Guilford Press.

Fay, R.E. 1986. "Causal Models for Patterns of Nonresponse." *Journal of the American Statistical Association* 81(394): 354–365. Doi: http://dx.doi.org/10.1080/01621459.1986.10478279.

Fuchs, C. 1982. "Maximum Likelihood Estimation and Model Selection in Contingency Tables with Missing Data." *Journal of the American Statistical Association* 77(378): 270–278. Doi: http://dx.doi.org/10.2307/2287230.

Goodman, L.A. 1961. "Statistical Methods for the Mover-Stayer Model." *Journal of the American Statistical Association* 56(296): 841–868. Doi: http://dx.doi.org/10.2307/2281999.

Goodman, L.A. 1973. "The Analysis of Multidimensional Contingency Tables when Some Variables are Posterior to Others: A Modified Path Analysis Approach." *Biometrika* 60(1): 179–192. Doi: http://dx.doi.org/10.2307/2334920.

Graham, J.W. 2009. "Missing Data Analysis: Making It Work in the Real World." *Annual Review of Psychology* 60: 549–576. Doi: http://dx.doi.org/10.1146/annurev.psych.58.110405.085530.

Hart, T.C., C.M. Rennison, and C. Gibson. 2005. "Revisiting Respondent 'Fatigue Bias' in the National Crime Victimization Survey." *Journal of Quantitative Criminology* 21(3): 345–363. Doi: http://dx.doi.org/10.1007/s10940-005-4275-4.

Hess, S., N. Sanko, J. Dumont, and A. Daly. 2013. "A Latent Variable Approach to Dealing with Missing or Inaccurately Measured Variables: The Case of Income." In Proceedings of the Third International Choice Modelling Conference, July 3–5, 2013. Sydney, Australia: ICM Conference. Available at: http://www.icmconference.org.uk/index.php/icmc/ICMC2013/paper/viewFile/744/233 (accessed August 2015).

Iannacchione, V. 1982. "Weighted Sequential Hot Deck Imputation Macros." In Proceedings of the SAS Users Group International Conference, February 14–17, 1982.

759–763. San Francisco, CA. Available at: http://www.sascommunity.org/sugi/SUGI82/Sugi-82-139%20Iannacchione.pdf (accessed March 2015).

Langton, L. and J. Truman. 2015. *Criminal Victimization, 2014*. Washington, DC: Bureau of Justice Statistics. (NCJ 248973).

Lazarsfeld, P.F. 1950. "The Logical and Mathematical Foundation of Latent Structure Analysis." In *Studies on Social Psychology in World War II, Vol. 4, Measurement and Prediction*, edited by S. Stauffer, E.A. Suchman, P.F. Lazarsfeld, S.A. Starr, and J. Clausen. Princeton, NJ: Princeton University Press.

Little, R.J. and D.B. Rubin. 2002. *Wiley Series in Probability and Statistics: Statistical Analysis with Missing Data*. 2nd ed. Somerset, NJ: Wiley.

Poulsen, C.A. 1982. *Latent Structures Analysis with Choice Modeling Applications*. Aarhus, Denmark: Aarhus School of Business Administration and Economics.

Rand, M. and S. Catalano. 2007. *Criminal Victimization, 2006*. Washington, DC: U.S. Department of Justice, Office of Justice Programs. (NCJ 219413).

Rubin, D.B. 1976. "Inference and Missing Data." *Biometrika* 63(3): 581–592. Doi: http://dx.doi.org/10.1093/biomet/63.3.581.

Schafer, J.L. and J.W. Graham. 2002. "Missing Data: Our View of the State of the Art." *Psychological Methods* 7(2): 147–177. Doi: http://dx.doi.org/10.1037//1082-989x.7.2.147.

Truman, J.L. and R.E. Morgan. 2016. *Criminal Victimization, 2015*. Washington, DC: Bureau of Justice Statistics. (NCJ 250180).

U.S. Census Bureau. 2014. *National Crime Victimization Survey: Technical Documentation*. Washington, DC: U.S. Census Bureau. (NCJ 247252).

U.S. Department of Justice. 2015. *Bureau of Justice Statistics. National Crime Victimization Survey, 2014*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

Van de Pol, F. and J. de Leeuw. 1986. "A Latent Markov Model to Correct for Measurement Error." *Sociological Methods & Research* 15: 118–141. Doi: http://dx.doi.org/10.1177/0049124186015001009.

Van de Pol, F. and R. Langeheine. 1990. "Mixed Markov Latent Class Models." In *Sociological Methodology*, edited by C.C. Clogg, 213–247. Oxford: Blackwell.

Vermunt, J.K. 1997. *Log-Linear Models for Event Histories*. London: Sage.

Vermunt, J.K. and J. Magidson. 2013. *Technical Guide to Latent Gold 5.0: Basic, Advanced, and Syntax*. Belmont, MA: Statistical Innovations.

Wiggins, L.M. 1973. *Panel Analysis, Latent Probability Models For Attitude And Behavior Processing*. Amsterdam: Elsevier SPC.

Received January 2016
Revised April 2017
Accepted April 2017