Rethinking the NCVS:

Small Area Estimation Approaches to Estimating Crime

Jianzhu Li, Mamadou S. Diallo, and Robert E. Fay[1]

Westat, Inc.
1600 Research Blvd, Rockville, MD 20850-3129, JaneLi@westat.com, MamadouDiallo@westat.com, BobFay@westat.com.

Abstract

The Bureau of Justice Statistics (BJS) is currently pursuing a set of strategies to expand the usefulness of the National Crime Victimization Survey (NCVS). BJS is simultaneously investigating designs for a core NCVS, which will continue to be conducted by the Census Bureau, and designs for a supplemental survey or surveys to be conducted by outside organizations. The overall goal will be to combine these sources into an integrated statistical program.

One facet of this overall research effort is to investigate how much can be accomplished from the core NCVS alone. In a companion paper, we report on the extent that selective boosting and reallocation of the sample, in combination with use of 3- and 5-year period averages, can support estimates for selected states or other subnational areas through direct estimation. In this paper we report on a parallel effort, namely, to investigate the extent to which small area estimation techniques can be applied to the core NCVS to produce useful subnational estimates.

The success of small area estimation projects often depends to a large degree on the availability of predictive auxiliary data. We previously analyzed NCVS data as a guide for the redesign of the core NCVS, but the results are also a starting point for small area estimation. We will report on our efforts to examine other auxiliary variables. The application includes a number of interesting technical challenges. The incidence of many crime variables, including the violent crime rate, is quite low, possibly requiring a more sophisticated technical approach than models based on the normal distribution. We will investigate how to make optimal use of the survey data across time. We will discuss the problem of communicating the interpretation of these estimates to the law enforcement agencies and other, primarily nonstatistical, audiences for the NCVS data.

1. Introduction

Since 1972, the National Crime Victimization Survey (NCVS) has provided national estimates of the incidence of crime by asking respondents about crimes committed against them in the previous 6 months, whether or not they had reported them to the police. The U.S. Census Bureau has collected the survey data for the Bureau of Justice Statistics (BJS). The primary focus of the survey has been on national estimates of crimes categorized by type of crime, on characteristics of the specific instances of crime, and on characteristics of the victims. In 1992, a major redesign of the questionnaire substantially increased the rates of reported crime for some components (see Rennison and Rand, 2007, for a summary), so that the data for years prior to 1992, when the survey was called the Nation Crime Survey (NCS), are often statistically adjusted to produce a continuous trend line.

An even older series, the Uniform Crime Reports (UCR) of the FBI, collects summaries of crimes by type as reported voluntarily by participating local law enforcement agencies, which numbered over 18,000 in 2010. The first year of data collection was 1930. For most of this period, the UCR series has been subject to significant problems of missing data from several sources, including non-participation and late or missing reports (Barnett-Ryan 2007; Maltz 2007). Although missing data in the UCR remains an issue, the level has diminished in recent years. The UCR has been the primary source for local-area crime statistics.

The NCVS complements the UCR by providing data that the UCR lacks. The NCVS provides characteristics of criminal incidents from the perspective of the victim, allowing more detailed characterizations than provided by the UCR. Because the NCVS asks whether the crime was reported to the police but collects the information on each crime regardless, the NCVS can measure crimes that are invisible to the UCR. In this paper we will describe preliminary regression results at the state level relating the crime rates in the UCR to NCVS rates.

As evident from the topic of this session, BJS is pursuing a research program to identify the best methods to extend the utility of the NCVS by producing subnational estimates. BJS is currently supporting research on different strategies. One strategy is to increase the sample size and expand the sample to produce direct estimates for some or all states (Fay and Li 2012). Another is to develop a lower-cost supplemental (or "companion") survey that can be carried out by an outside organization to produce data that can be combined with the core NCVS data from the Census Bureau (Brick, Edwards, and Lohr 2012). Because the NCVS and companion surveys may not cover identical universes and may differ systematically because of mode differences and other methodological factors, it will likely be necessary to combine the results through some form of modeling.

This paper describes work we have undertaken to develop methods to produce small area estimates of crime based only on the core NCVS and on currently available auxiliary data. We will describe our research at the state level, although our research will assess the possibility of substate estimates as well. A companion paper (Fay and Li 2012) reviews some previous efforts to produce direct estimates from NCVS for Metropolitan Statistical Areas (MSAs) and the amount of sample boosting required for direct state estimates.

Perhaps because publicly available national NCVS files lack detailed geographic information, there has been little work to model variation in crime rates geographically. But BJS released a file with the core areas of the 40 largest MSA's during 1979-2004, where both the specific MSA's and years were identified. Recently, Xi, Heimer, and Lauritsen (2011) reported on an analysis of violence against women using both the temporal and geographic information on this file. To our knowledge, however, there have been no published studies of NCVS estimates at the state level.

In a previous study, Krenzke, Li, and Cantor (2009) reported on an initial attempt to develop a cross-sectional small area model at the MSA level for the 2003 NCVS estimates extracted from the 1979-2004 file. A subsequent report (Cantor, et al. 2010) discussed different small area approaches that could be the pursued for NCVS, illustrating them by discussing three existing applications of small area methods to other surveys. Among its other recommendations, the report suggested the possible application of time series modeling to NCVS data. We will show why this suggestion seems particularly pertinent to the NCVS application and describe our work pursuing this line of research.

A typically important aspect of successful application of small area estimation methods is to identify auxiliary data useful to model the variables of interest. Cantor et al. (2010) identified three broad categories of candidate auxiliary variables for prediction: the American Community Survey, other BJS statistical programs, and the UCR. In this paper, we report on a preliminary analysis of what might be expected from some of these variables at the state level.

With access to confidential files at the Census Bureau, Fay and Li (2011) reported on a regression analysis to fit NCVS victimization rates at the county level for violent crime by component—aggravated assault; simple assault; robbery; rape and sexual assault; and property crime—as a function of UCR rates and census data. The work has not yet progressed to produce small area estimates for counties, but the early results were generally encouraging. The work also did not investigate state-level relationships. We will report here on preliminary results applying a similar regression analysis at the state level.

In summary, the paper reports on progress pursuing previously identified research directions, but it also includes some findings that were not previously anticipated. To organize the results, we report in Section 2 on several analyses of the publicly available auxiliary data as well as indicate some initial findings based on confidential NCVS data at the Census Bureau. We have placed this preliminary exploration of the data in Section 2 in an attempt to characterize the constraints and opportunities in the specific NCVS small area application. The section includes results we have already noted here in the introduction: (1) our preliminary analysis of the auxiliary data; and (2) the regression analysis results for states similar to those already reported for counties. The section also further describes the auxiliary data and features of the NCVS data sets available for our research.

Section 3 follows by addressing the question of what available small area methods might best be adapted to the specific NCVS estimation goals. As already noted, the time-series aspect of this application figures prominently in our development, and we describe our candidate model.

In addition to commenting on the directions of our subsequent research, the discussion section remarks on our plans to work with the Census Bureau to establish the degree to which small area estimates based on the NCVS can be safely published without posing a threat to confidentiality. We also describe our plans to facilitate communication of model-based estimates to potential users of these results.

2. Data Sources and Preliminary Analysis

As we remarked in the introduction, the paper features our efforts to explore relationships in the NCVS data and the underlying phenomena that they measure before we propose specific statistical models and small area approaches. In short, data analysis precedes formal modeling.

Because statisticians often emphasize the technical aspects of their small area applications, we find some support for the arrangement of our paper from a remark of J.N.K. Rao (2003). For good reason, his book *Small Area Estimation* is highly regarded as an authoritative review of the technical aspects of the subject. To begin, he reviewed possible small area approaches including direct small area estimation through model-assisted estimation, traditional demographic methods, and three precursor methods (synthetic estimation, composite estimation, and the James-Stein estimator). In Chapter 5 he introduced and surveyed “Small Area Models” that represent the primary focus of his book. The second paragraph of the chapter observes (p. 75) the following:

> Although we present a variety of models for small area estimation in this chapter, it is important to note that the subject matter specialists or end users should have influence on the choice of models, particularly on the choice of auxiliary variables. Also, the success of any model-based method depends on the availability of good auxiliary data. More attention should therefore be given to the compilation of auxiliary variables that are good predictors of the study variables.

In the NCVS context, BJS has actively sought the advice of data users, including law enforcement officials, on future directions for the survey. These users supported a greater focus on subnational estimation, the goal of this project. We have consulted previous work by Westat colleagues and advice from BJS for possible directions to pursue. This section describes our search for appropriate auxiliary data and can serve as the basis for additional suggestions from BJS and other interested researchers.

Section 2.1 introduces the components of the UCR closely paralleling those of NCVS. Section 2.2 describes the NCVS and notes aspects of its design important in developing small area models. Some preliminary state-level regression results are summarized in Section 2.3.

2.1 Auxiliary Information from UCR

The UCR collects and reports data both on offenses known to the police and on arrests. Because they do not depend on the success in clearing cases, reports on offenses relate more directly to the measurement goals of the NCVS than statistics on arrests, although we comment on the arrest data in Section 4. Type 1 offenses in the UCR include the violent crimes of aggravated assault, robbery, forcible rape, and murder; and the property crimes of larceny, burglary, and motor vehicle theft. The original basis of the UCR was the aggregation of counts of incidents by type of crime compiled by law enforcement agencies, and this approach still accounts for the majority of the UCR totals. But a significant fraction of the UCR is now based on the NIBRS system, which collects detailed information at the level of the individual offense. In 2010, agencies participating in NIBRS accounted for about 28% of the U.S. population. Recently, Barnett-Ryan (2007) summarized the development and current state of the UCR, including the development of the NIBRS system.

The FBI publishes UCR statistics at the state and national level in *Crime in the United States*, which is now available over the Internet. National and state estimates appear first; statistics for 2010 were published in 2011. As previously noted, missing data affects the UCR statistics. For type 1 offense data, the FBI imputes for missing data in order to publish national and state estimates, although these methods have been the subject of study and criticism by outside researchers (Barnett-Ryan 2007; Maltz 2007). For example, the amount of imputation is not readily available from the website on a state-by-state basis.

The FBI also publishes the "raw" data from reporting jurisdictions on a somewhat more delayed schedule. At the end of 2011, jurisdictional data were available for 2009 but not 2010. On the web, the jurisdictional data are organized geographically but are not in a form convenient for geographic analysis. County-level jurisdictions are shown separately from city-level jurisdictions, and the FBI does not provide estimated disaggregations of cities that cross county boundaries. Upon request, the FBI provides researchers the highly detailed enforcement agency data (shown by month) that are the basis of the Internet release, but working with these data can be challenging. The FBI does not provide imputed data at this level.

The Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan aggregates the detailed FBI data into county estimates, including imputing some of the missing data for partial reports and analyzing the proportion of missing data for individual counties. At the end of 2011, the detailed county-level data on offenses were available only through 2008, however.

States exhibit considerable variation in their rates of reported crime in the UCR. Furthermore, the geographic patterns of this variation depend on the type of crime. Fig. 1 displays variation in the UCR rates of aggravated assault at the state level during 2008-2010 compared to the national average. Nationally, aggravated assault is generally the most frequently reported violent crime in the UCR. Eight states, including Florida, and also the District of Columbia are 50% or more above the national rate. Four states, Virginia and three in New England, are below by 50% or more. A group of northern states in the Midwest and Mountain states have rates below average, but Michigan is high relative to its neighbors.

In contrast, in Fig. 2 the UCR shows a somewhat different pattern for the second most frequent type of violent crime, robbery. Only three states—Delaware, Maryland, Nevada—plus the District of Columbia, are above 50% of the national average. These three also show relatively high rates in Fig. 1. Far more states are below 50% of the national rate, including several in the northern tier of the West North Central and Mountain divisions, as well as West Virginia and upper New England states. Overall, the graphs suggest a weak geographic correlation between robbery and aggravated assault. Many states drop one or two shades from Fig. 1 to Fig. 2, but California and Illinois, among other states, increase their relative ranking.

Fig. 3 compares UCR rates for forcible rape relative to the national average. Here, the state-level patterns are distinctly different from Fig. 1 and Fig. 2. Only two states, New York and New Jersey, have rates 50% below the national average. Six more states, Vermont, Connecticut, Maryland, Virginia, West Virginia and Michigan, have rates 20% or more below the national average. Most of the northern West North Central and Mountain and Pacific states, which were generally below the national average in Fig. 1 and Fig. 2, are at or above the national rate in Fig. 3. The rate for robbery in Colorado was 50% below the national average in Fig. 2 but 50% above the average for rape in Fig. 3.

For the sake of completeness, Fig. 4 shows the geographic distribution for murder by state, even though this characteristic is not measured by the NCVS. Although murder is also a violent crime, its geographic pattern is distinctly different from the other three.

We have similarly evaluated the geographic pattern for the property crimes of burglary, larceny, and motor vehicle theft. Although there a few patterns emerge, such as a North to South gradient on murder and burglary, each geographic pattern appears distinctively different. Of all of the major categories of crime, the rates for larceny appear the most similar from state to state, while most other crimes exhibit considerable geographic variation.

In short, the UCR statistics present a complex picture of the geographic distribution of crime, suggesting that crime is a multi-dimensional phenomenon. Thus, modeling the geographic distribution for each type separately would seem to be more informative, and possibly more productive, than modeling violent crime or property crime as a whole.

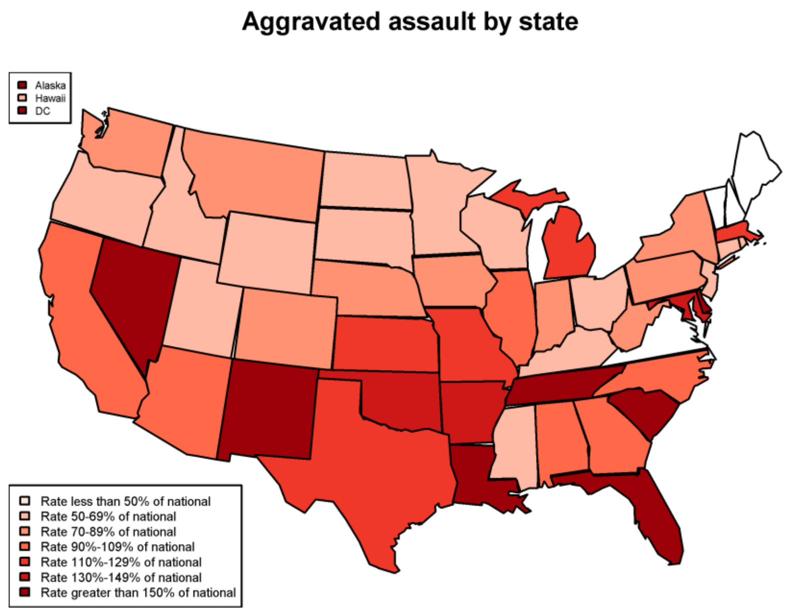

Fig. 1. UCR rates for aggravated assault, 2008-2010, compared to the national rate.

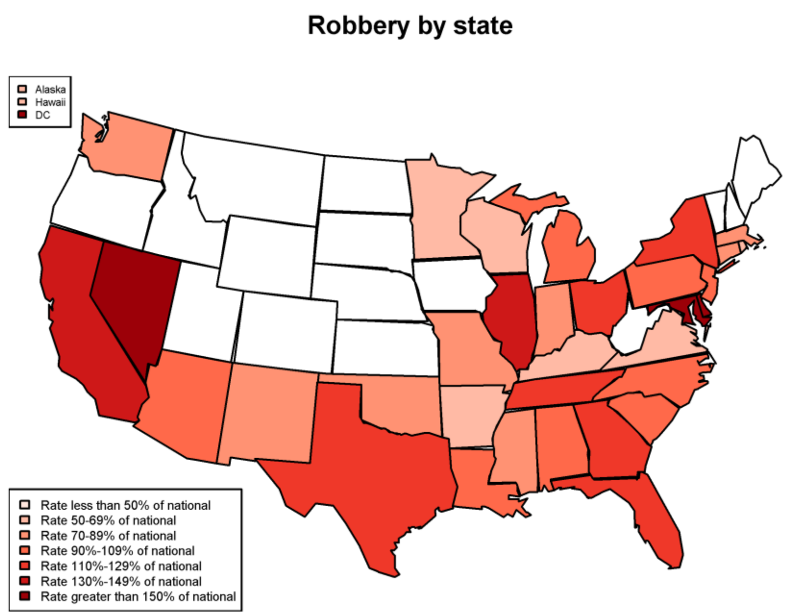

Fig. 2. UCR rates for robbery, 2008-2010, compared to the national rate.

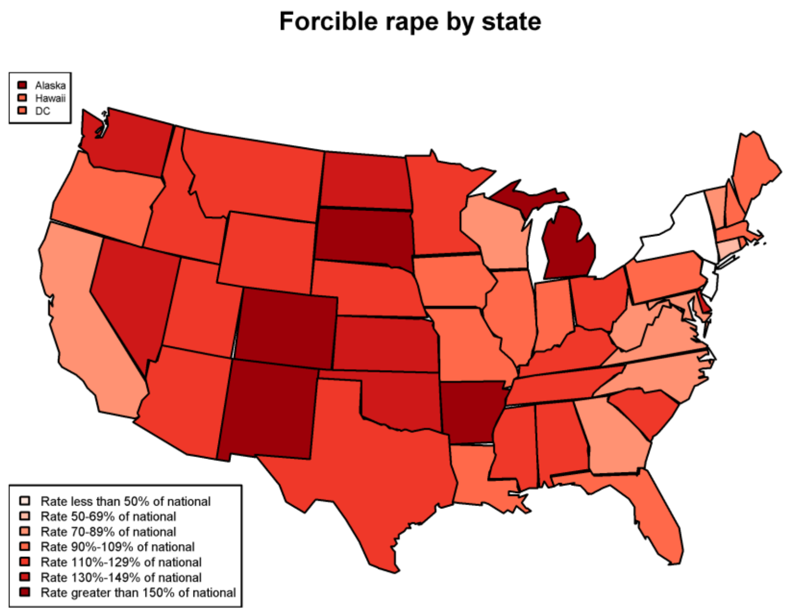

Fig. 3. UCR rates for forcible rape, 2008-2010, compared to the national rate.

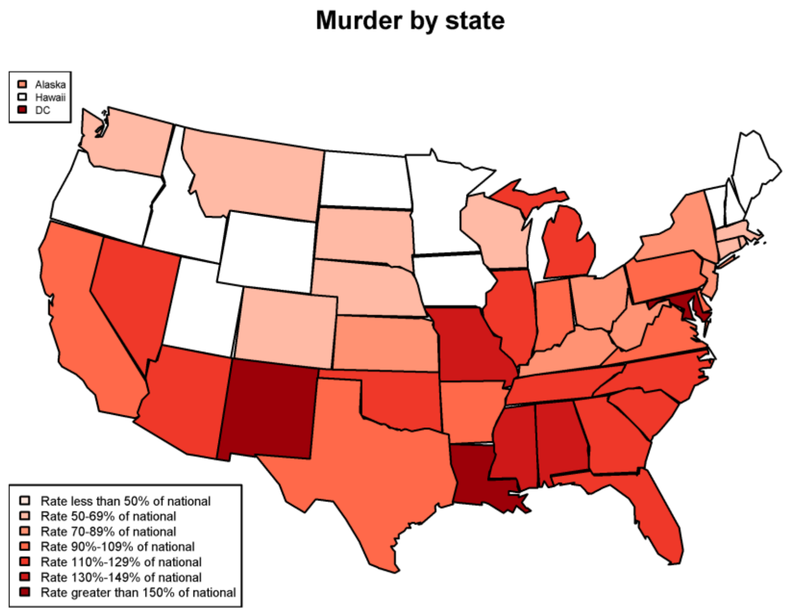

Fig. 4. UCR rates for murder, 2008-2010, compared to the national rate.

The UCR data also suggest stability in the relative ranking of states over time. Fig. 5 compares the average rates for the three most recent available years, 2008-2010, with rates for 1998-2000 ten years earlier. The correlations are quite high even over a decade. In Fig. 6, similar high correlations are obtained for the three components of property crime, larceny, burglary, and motor vehicle theft. This evidence favors selecting small area models attuned to the relative stability of the geographic variation in the crime rates over time.

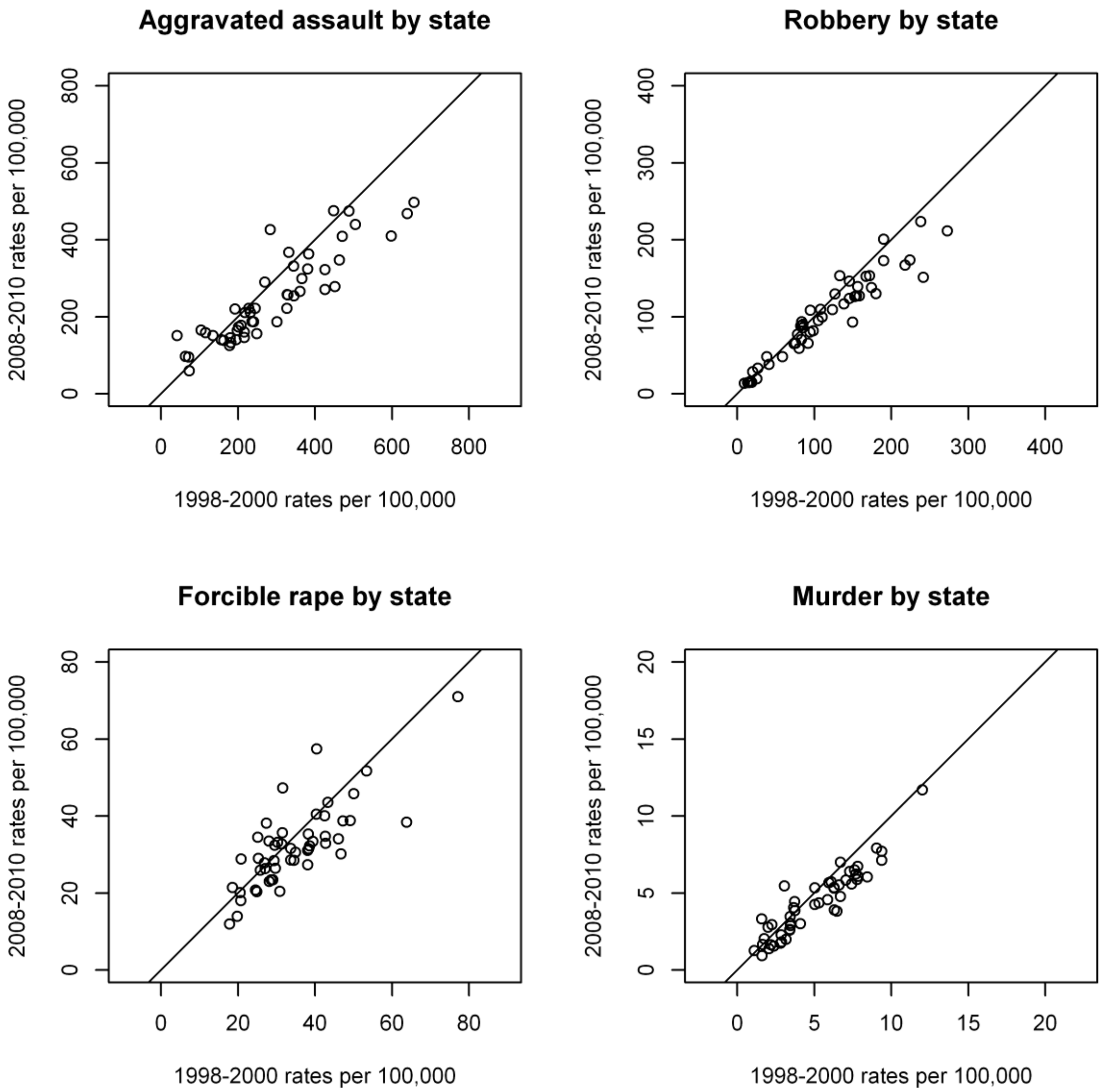

Fig. 5. Comparison of UCR state-level rates for components of violent crime for 1998-2000 and for 2008-2010. The District of Columbia is omitted from the comparisons because of high rates for robbery and murder. A line with slope 1 through the origin is shown, and the extent to which states fall generally below the line is consistent with the national drop in crime during this period.

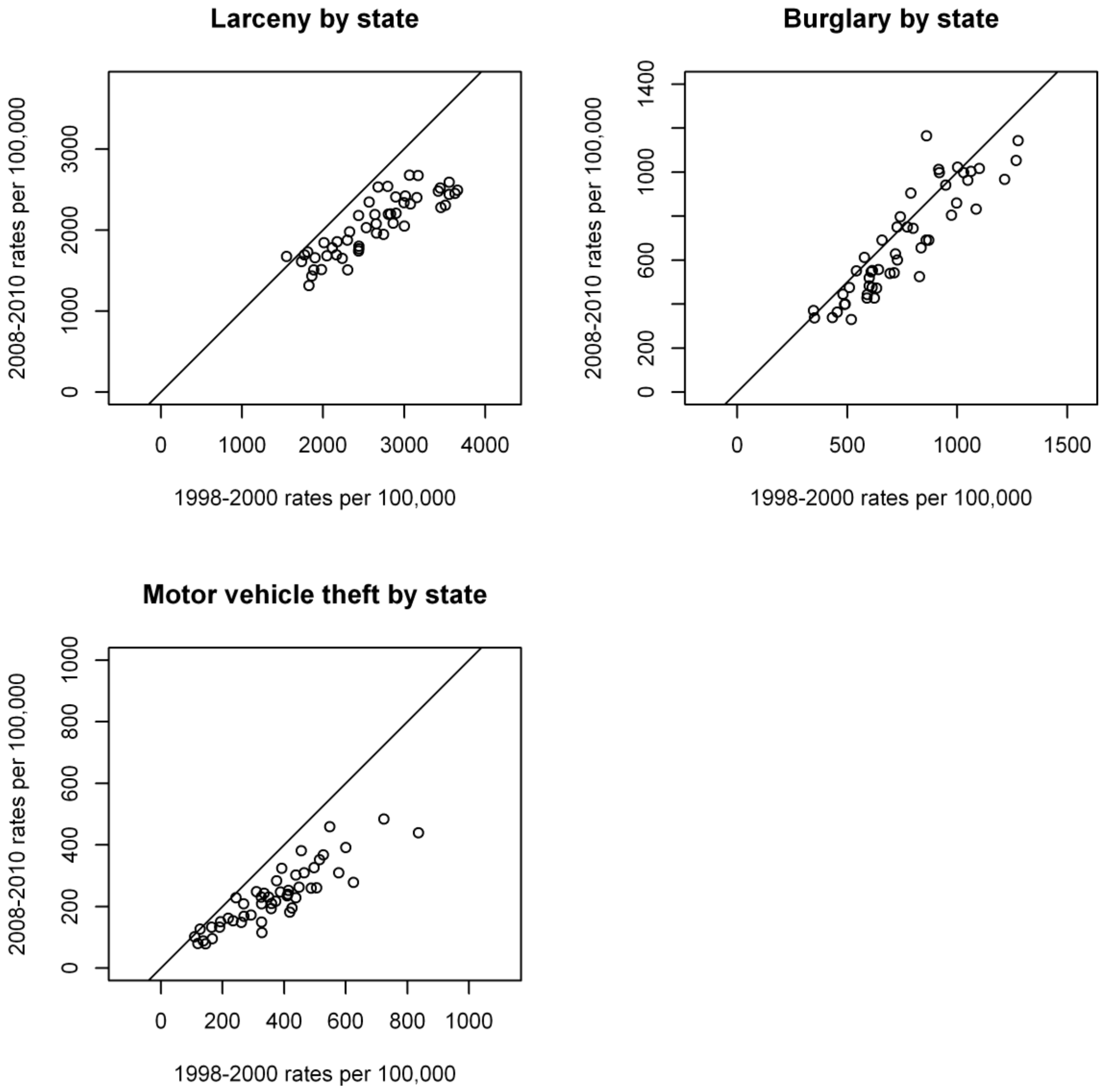

Fig. 6. Comparison of UCR state-level rates for components of violent crime for 1998-2000 and for 2008-2010. As in Fig. 5, the District of Columbia is omitted from the comparisons.

As a further illustration of the strong correlation across time exhibited in the UCR data, Fig. 7 compares the UCR crimes rates in the four largest states with the national rates across time. For each type of crime, the trends in each state follow the national trends to a large extent. In California, UCR rates are relatively high for robbery and murder but low for forcible rape. Starting from a comparatively high level, the rate for aggravated assault appears to converge to the national average by 2010. In Texas, rates are near the national average for robbery, a bit above the national average for aggravated assault, somewhat high for murder and high for forcible rape. New York reports consistently low rates for forcible rape and murder but high rates for robbery. UCR rates in Florida for robbery, aggravated assault, and forcible rape are all higher than the national average, although the rate for forcible rape converges to the national mean by 2010. Overall, rates appear somewhat less dispersed by 2010 than in 1996.

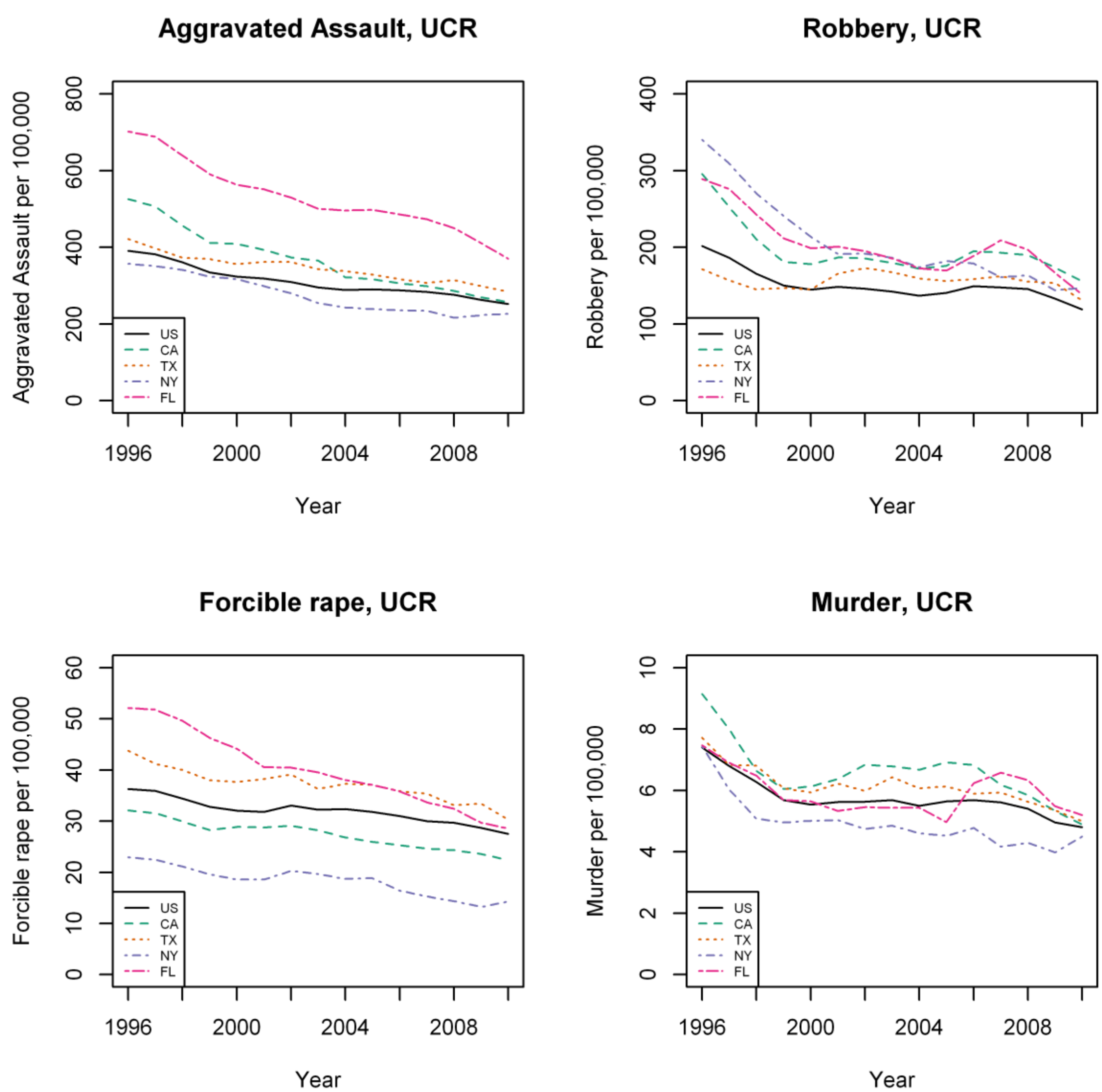

Fig. 7. Trends in the UCR rates of four types of violent crime, 1996-2010, in California, Texas, New York, and Florida.

2.2 The NCVS

The introduction reviewed some of the rationale for a national household sample survey to collect information on crime from victims. Depending on the type of crime, surveys including the NCVS have found that victims fail to report a significant proportion of crimes to the police. The NCVS collects detailed information on the occurrence and subsequent consequences of a criminal victimization. The survey also collects information on the victims, such as their age, sex, relationship to the offender, including whether the offender was a stranger, acquaintance, relative, or intimate.

The NCVS uses definitions of violent and property crime that closely parallel, but do not completely overlap, the definitions of the corresponding crimes as defined by the UCR. Differences are noted by several contributors to Lynch and Addington's edited volume (2007a). Nonetheless, we might expect each UCR component to be a useful predictor of its corresponding NCVS analogue. But the largest single component of violent crime in the NCVS, simple assault, is not reported as a Type 1 offense in the UCR.

Table 1 compares NCVS and UCR rates for 1996 and 2009 available from publications. Direct comparisons between the UCR and NCVS rates are difficult because of differing denominators: The UCR uses total population for all rates but the NCVS uses the estimated population age 12 and over for violent crime and the estimated number of households for property crime. In spite of this difficulty, the comparison shows a steeper drop during this period for the NCVS relative to the UCR.

Table 1: Comparison of UCR and NCVS Published Estimates of Crime by Type in 1996 and 2009 (UCR rates per 1,000 persons; NCVS rates per 1,000 persons age 12+, except for NCVS household burglary, motor vehicle theft, and theft, which are shown per 1,000 households. NCVS estimates are on a collection year basis, and include some crimes committed in the previous year.)

UCR classification	*1996*	*2009*	*% change*	*NCVS classification*	*1996*	*2009*	*% change*
Forcible rape	0.36	0.29	-21%	Rape/sexual assault	1.4	0.5	-64%
Robbery	2.02	1.33	-34%	Robbery	5.2	2.1	-60%
Aggravated assault	3.91	2.63	-33%	Aggravated assault	8.8	3.2	-64%
				Simple assault	26.6	11.3	-58%
				Personal theft	1.5	0.5	-67%
Burglary	9.45	7.16	-24%	Household burglary	47.2	25.6	-46%
Motor vehicle theft	5.26	2.59	-51%	Motor vehicle theft	13.5	6.0	-56%
Larceny	29.80	20.61	-31%	Theft	205.7	95.7	-53%

Sources: For UCR, http://www2.fbi.gov/ucr/cius2009/documents/table_01.html, referenced 29 Jun 2011; for NCVS, Ringle (1997) and Truman and Rand (2010). From Table 1, Fay and Li (2011).

Fig. 8 further illustrates differences in the national trend between the UCR and NCVS.

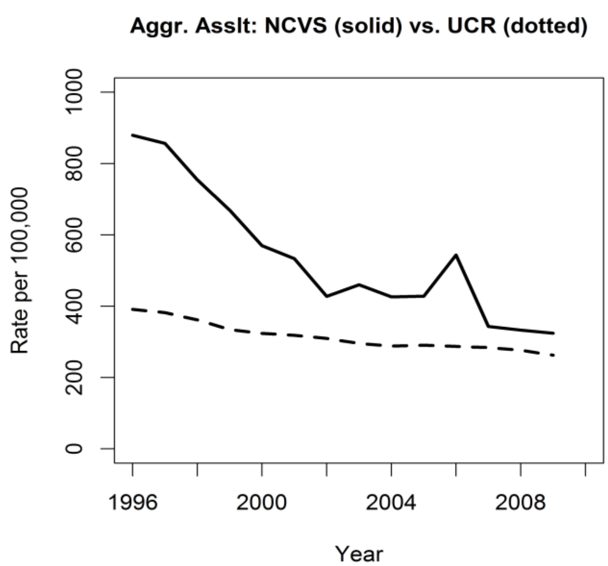

Fig. 8. Comparison of NCVS and UCR national rates for aggravated assault by year. The spike in 2006 was particularly evident in rural areas and may have been an effect of a large number of new interviewers.

Leaving aside both personal theft (purse snatching and pocket picking) and simple assault reported in the NCVS, the UCR and NCVS similarly rank aggravated assault as the most frequent violent crime, followed by robbery, and finally by rape. Similarly, larceny is the dominant form of property crime, followed by burglary and motor vehicle theft.

The NCVS survey design and sample design are both important considerations in formulating a small area model at the state level. The NCVS has a rotating panel design, where segments of 4 housing units are the ultimate sample unit. Households in sampled housing units are interviewed every 6 months, for a total of 7 times over a period of 3 years. Until 2006, data from the first interview were excluded from the analysis, treating the first interview as a bounding interview, but now these data are statistically adjusted and included in the published estimates.

In addition to the clustering of persons within households and households within segments, the sample design for the NCVS is similar to other large national samples incorporating personal visits, such as the Current Population Survey. (The American Community Survey is an exception to the following because of its enormous size and goal of subnational statistics at detailed levels of geography.) In designing the sample, counties are first grouped into PSUs. The most populous PSUs, referred to as *self-representing*, are included in the sample with certainty, and a probability sample of segments is drawn from each of them. The remaining *non-self-representing* PSUs were stratified and a single PSU selected from each. Generally this first-stage of the design, that is, the selection of PSUs, is revised only once per decade. Thus the complex design of the NCVS includes features potentially affecting the variance of the estimated crime rates: the sampling of non-self-representing PSUs at the first stage and the clustered sampling of segments of households. These features induce interclass correlations between segments in non-self-representing PSUs, between households within segments, between persons in sampled households, and between the 6-month reports for the same persons over time.

To enable the calculation of sampling variances for this complex design, the Census Bureau assigns a pseudo-stratum code to each individual record. The code may be treated for purposes of variance estimation as if it were the actual sampling stratum. The Census Bureau also assigns a code ("SECU code") of 1 or 2 to each record to be treated as a half-sample assignment within pseudo-stratum. All households in the same segment receive identical codes, as do all the segments in each non-self-representing PSU. Except for changes due to sample redesigns each decade, the assignments remain stable over time, thus supporting the estimation of variances and covariances over time reflecting the complex design. A companion paper (Fay and Li 2012) reports our finding that between-PSU variance in non-self-representing areas was an important contributor to the overall variance. Section 3 describes further the role of the variance calculations in small area estimators we will consider.

2.3 Predicting State-Level NCVS Rates from the UCR

Fay and Li (2011) previously reported the results of attempting to fit variation in NCVS crime rates at the county level on the basis of the UCR. The analysis was restricted to self-representing counties with relatively complete UCR reporting. In most cases, the UCR rate predicted the corresponding NCVS rate, when both were averaged over the same set of years. Specifically, NCVS property crime was predicted by UCR property crime, NCVS rape and sexual assault was predicted by UCR forcible rape, and NCVS robbery was predicted by UCR robbery. But UCR aggravated assault was minimally successful in predicting either NCVS aggravated assault or simple assault. Instead, UCR forcible rape displaced UCR aggravated assault as a predictor of either NCVS aggravated or simple assault. Table 2 shows the regression results for violent crime as previously reported. The large coefficients for UCR forcible rape as a predictor of NCVS aggravated assault or total violent crime is effectively using UCR forcible rape variable as a symptomatic indicator of the level of violent crime generally.

With the shift in focus to state-level small area modeling, we recently fitted a series of similar regression models at the state level. It would be quite easy to posit a number of reasons that the regression relationships observed for large counties might not replicate at the state level. But they do. Table 3 summarizes the findings qualitatively, including a disaggregation of the components of property crime not previously investigated for counties.

Table 2: Regression Prediction of County-Level NCVS Violent Crime Rates from the UCR Rates in Self-Representing Counties with the Highest Rates of Complete UCR Reporting, for 1996-2005

	NCVS violent crime	*NCVS violent crime*	*NCVS Aggrav. assault*	*NCVS Rape/sexual assault*	*NCVS Robbery*
Dependent mean	30.28	30.28	6.05	1.25	4.29
Intercept	23.14	18.89	3.20	0.59	0.98
	(1.36)	(1.60)	(0.47)	(0.19)	(0.31)
UCR violent crime	1.15				
	(0.19)				
UCR aggravated assault		-1.14	0.00	-0.15	-0.19
		(0.56)	(0.16)	(0.07)	(0.11)
UCR forcible rape		31.04	8.49	3.36	0.42
		(5.68)	(1.66)	(0.66)	(1.12)
UCR robbery		2.76	0.08	0.08	1.83
		(0.67)	(0.20)	(0.08)	(0.13)

Source: Table 4, Fay and Li (2011).

Table 3: Best UCR Predictor for Separate Types of NCVS Crime at the State Level

NCVS Rate	*Best UCR Predictor*
Rape/sexual assault	Forcible rape
Robbery	Robbery
Aggravated assault	*Forcible rape*
Simple assault	*Forcible rape*
Household burglary	Burglary
Motor Vehicle Theft	Motor Vehicle Theft
Theft	Larceny

Although each type of crime measured by NCVS is correlated with some component of the UCR, the results for aggravated and simple assault are disappointing. Relying on the comparatively low rates for forcible rape to predict the dominant components of the violent crime rate is far less than ideal. We will consider whether other sources of auxiliary data might improve the prediction. Alternatively, at least in most of the larger states, Fig. 7 suggests that small area estimates for states should attempt to take advantage of the strong stability across time in the underlying crime rates.

3. Developing a Small Area Strategy

Rao (2003, Sec. 5.4.3) introduced area-level models combining time series and cross-sectional models by discussing a number of related strategies. In one of these, Rao and Yu (1992, 1994) proposed an extension to the Fay-Herriot model. The extension proposed the standard relationship between a sample estimate, y_{it} , for area *i* and time *t*, and its expected population value, θ_{it} ,

$$y_{it} = \theta_{it} + e_{it} \qquad (3.1)$$

for i=1,…,m, t=1,…,T, where the normally distributed error terms $e_i = (e_{i1},..., e_{iT})^T$ are assumed to have mean zero and covariance $\sum_i$, which is assumed known. The error terms are assumed independent between areas. The approach proposed the following model for the underlying population:

$$\theta_{it} = \mathbf{x}_{it}^T\boldsymbol{\beta} + \nu_i + u_{it} \qquad (3.2)$$

with

$$u_{it} = \rho u_{i,t-1} + \varepsilon_{it} \qquad (3.3)$$

where

$\mathbf{x}_{it} = (x_{it1},\ldots,x_{itp})^T$ is the vector of auxiliary variables for area *i* at time *t*,
$\boldsymbol{\beta}$ is a vector of regression coefficients,
$\nu_i \sim N(0,\sigma_\nu^2)$ for *i*=1,…,*m* are iid random effects, representing time-independent differences among areas, and
$\varepsilon_{it} \sim N(0,\sigma_u^2)$ are iid random variables, which induce variability in the series u_{it}, *t*=1,…*T*.

For $\rho = 1$, $\sigma_u^2 > 0$, (3.3) results in a random walk, where the variance of u_{it} increases with time *t*. Allowing $|\rho| < 1$ permits (3.3) to be stationary for specific combinations of ρ and σ_u^2.

Assuming first that σ_u^2, σ_ν^2, and ρ are known, the best linear unbiased predictor (BLUP) for area *i* at time *T* is

$$\tilde{\theta}_{iT} = \mathbf{x}_{iT}^T\tilde{\boldsymbol{\beta}} + (\sigma_\nu^2\mathbf{1}_T + \sigma_u^2\boldsymbol{\gamma}_T)^T(\boldsymbol{\Sigma}_i + \sigma_u^2\boldsymbol{\Gamma} + \sigma_\nu^2\mathbf{J}_T)^{-1}(\mathbf{y}_i - \mathbf{X}_i\tilde{\boldsymbol{\beta}})$$

where

$\boldsymbol{\Gamma}$ is a $T \times T$ matrix with elements $\rho^{|i-j|}/(1-\rho^2)$,
$\mathbf{J}_T$ is a $T \times T$ matrix with elements = 1,
$\mathbf{V}_i = \boldsymbol{\Sigma}_i + \sigma_u^2\boldsymbol{\Gamma} + \sigma_\nu^2\mathbf{J}_T = Cov(\mathbf{y}_i)$,
$\mathbf{V} = diag_i(\mathbf{V}_i) = Cov(\mathbf{y})$,
$\tilde{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$ is the generalized least-squares estimator of $\boldsymbol{\beta}$, and
$\boldsymbol{\gamma}_T$ is the *T*th row of $\boldsymbol{\Gamma}$.

The form of the model corresponds well to the NCVS application. It allows for sampling covariances across time, but the assumption that the NCVS rates are independent across states is reasonable. The high correlation across time shown by the UCR data suggests that the underlying ρ is large, although less than 1.

Rao and Yu (1994) addressed the problem of estimating the variance components by first treating ρ as known, and estimating $\tilde{\sigma}_u^2$ and $\tilde{\sigma}_\nu^2$ by consistent estimates based on a transformation of the data. The authors provide the details in their paper. They also discuss the problem of estimating ρ from the data, but report practical difficulty when the sampling errors are large relative to the other variance components.

We are adapting the strategy of first studying the performance of model (3.2) – (3.3) applied to the UCR data before extending the analysis to the NCVS. We recognize that σ_u^2, σ_ν^2, and ρ depend on the specific choice of the auxiliary data, but nonetheless will examine the implications of the UCR analysis for interpreting the NCVS. We will investigate restricted maximum likelihood (REML) as an alternative to their original approach, especially because the NCVS variances are quite varied among the states.

4. Discussion

At this point, we have identified a candidate small area approach and moved towards implementation. In spite of the many contrasts that have been made between the NCVS and the UCR, we have taken the strategy of basing our initial models on the UCR. In this respect we are attempting to follow a general suggestion from Lynch and Addington (2007b), who advocated finding most productive uses of the NCVS and UCR by considering their relative strengths and limitations.

We have taken some initial steps to examine other variables, including the FBI's series on arrests and other BJS series, including the annual Probation and Parole Survey, the Census of State and Local Law Enforcement Agencies, and the Justice Expenditure and Employment Extract Series. Some of these series provide annual data, and thus may help us to estimate both the u_{it} and v_i components of (3.2), whereas those with significant time lags or without annual estimates may be suited specifically to refining v_i.

As the modeling develops, we expect to address the disclosure issues that a public release of state estimates might entail. The complexity of the EBLUP form of the BLUP estimator, which will combine observations from several years, may provide significantly more protection from disclosure than the release of geographic identifiers on a microdata file.

Acknowledgements

We wish to thank the Bureau of Justice Statistics for their support of this research and Dr. Michael Planty for his help and guidance. We also acknowledge considerable assistance from members of the Census Bureau staff, including Stephen Ash, Barbara Blass, and Meagan Wilson. The views expressed here are ours and do not reflect those of the Bureau of Justice Statistics, the Census Bureau, or Westat.

References

Barnett-Ryan, C. (2007), "Introduction to the Uniform Crime Reporting Program," in M.A. Lynch and L.A. Addington (eds.) *Understanding Crime Statistics* (2007a), pp. 55-89.

Brick, J.M., Edwards, W.S., and Lohr, S. (2012), "Designing a Low(er)-Cost Complement to the National Crime Victimization Survey," to be presented at the Federal Committee on Statistical Methodology Research Conference, Washington, DC, Jan. 10-12, 2012.

Cantor, D., Krentzke, T., Stukel, D., and Rizzo, L. (2010), "NCVS Task 4 Report: Summary of Options Relating to Local Area Estimation," issued by Westat to the Bureau of Justice Statistics, May 19, 2010.

Fay, R.E. and Li, J. (2011), "Predicting Violent Crime Rates for the 2010 Redesign of the National Crime Victimization Survey (NCVS)," presented at the Joint Statistical Meetings, Miami Beach, FL, Jul 31 - Aug. 5, 2011.

______ (2012), "Rethinking the NCVS: Subnational Goals through Direct Estimation," presented at the Federal Committee on Statistical Methodology Research Conference, Washington, DC, Jan. 10-12, 2012.

Krenzke, T., Li, J., and Cantor, D. (2009). NCVS Task 3: Alternative Designs for Local Area Estimates (draft memorandum from Tom Krenzke, Jane Li, and David Cantor of Westat to Michael Rand, Allen Beck, and Erika Harrell of the Bureau of Justice Statistics). Rockville, MD: Westat, June 2, 2009.

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

______ (2007b), "Introduction," in M.A. Lynch and L.A. Addington (eds.) *Understanding Crime Statistics* (2007), pp. 3-13.

Maltz, M.D. (2007), "Missing UCR Data and Divergence of the NCVS and UCR Trends," in M.A. Lynch and L.A. Addington (eds.) *Understanding Crime Statistics* (2007a), pp. 269-294.

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

Rao, J.N.K. (2003), *Small Area Estimation,* John Wiley & Sons, Hoboken, NJ.

Rao, J.N.K. and Yu, M. (1992), "Small Area Estimation Combining Time Series and Cross-Sectional Data**,"** *Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 1-9.

______ (1994), "Small Area Estimation by Combining Time Series and Cross-Sectional Data," *Canadian Journal of Statistics,* 22, 511-528.

Rennison, C.M. and Rand, M. (2007), "Introduction to the National Crime Victimization Survey," in M.A. Lynch and L.A. Addington (eds.) *Understanding Crime Statistics* (2007a), pp. 17-54.

Ringle, C. (1997), "Criminal Victimization, 1996," NCJ 165812, issued by the Bureau of Justice Statistics, Nov. 1997.

Truman, J.L. and Rand, M. (2010), "Criminal Victimization, 2009," NCJ 231327, issued by the Bureau of Justice Statistics, Oct. 2010.

Xi, M., Heimer, K., and Lauritsen, J.L. (2011), "Violence Against Women in U.S. Metropolitan Areas: Changes in Women's Status and Risk, 1980–2004," *Criminology,* in press, doi: 10.1111/j.1745-9125.2011.00245.x.