Predicting Violent Crime Rates for the 2010 Redesign of the National Crime Victimization Survey (NCVS)

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

Abstract

The National Crime Victimization Survey (NCVS) is a major crime survey for the United States. The survey collects data on several types of crimes, including the broad categories of violent crime and property crime. The 2010 redesign of the NCVS can potentially improve the efficiency of the survey if the level of crime can be predicted well by external data. Previously, we reported initial success in predicting the level of crime at the county level based on the Uniform Crime Reporting (UCR) System. A more fine-grained analysis shows, however, far greater success for property crime than for violent crime. This paper extends the previous results to examine the underlying associations more thoroughly. We find that the largest single component of violent crime in the UCR, aggravated assault, fails to add to the predictive accuracy of a regression equation including the other violent crime components of the UCR, rape and robbery. We believe this finding has implications for interpretation of the UCR. We extend the analysis to include demographic characteristics from the census and the ACS, and we examine the ability of tract-level characteristics to predict individual victimizations.

Key Words: Sample design, UCR

1. Introduction

The National Crime Victimization Survey (NCVS) collects data on the frequency and consequences of crime from its victims. Since 1972, the Census Bureau has conducted the survey for the Bureau of Justice Statistics. By surveying household members ages 12 and over in sampled housing units, the survey can include crimes not reported to the police.

The NCVS has a panel design, with households residing at the sampled housing units interviewed a total of 7 times at 6-month intervals. Since 1972, the design has specified a personal-visit interview on the first wave, but over the lifetime of the survey the emphasis on personal-visit interviews has varied; currently, continuing households may be interviewed by telephone on all subsequent waves, but incoming replacement households in the sampled housing units are to be interviewed through personal visit whenever possible. Until recently, the first wave was used solely as a bounding interview and excluded from the estimates. In the last few years, however, the first wave has been included by statistically adjusting the weights for reported incidents in the first wave.

As a result of both sample size cuts and decreasing victimization rates, the reliability of the NCVS had declined over time. A panel of the National Research Council (2008) recommended a thorough review of the design of the survey, including essentially all

aspects of the sample design. In 2008, BJS awarded a set of grants to outside organizations to study several aspects of the overall NCVS design, including a grant to Westat supporting our work on the sample design.

A classic strategy in survey sample design is to incorporate auxiliary information to improve efficiency, either through stratification, differential sampling, or estimation. In a multi-stage design, such as the design of NCVS, this strategy can be applied at each stage, quite possibly by using different sources of information at each stage. The search for suitable auxiliary information and development of methods to incorporate it into the NCVS design has been an important aspect of our overall effort. In the NCVS context, the two levels important for the sample design are (1) the county, because the primary sampling units at the first stage are composed on one or more counties; and (2) the household, because the ultimate clusters in the design are segments of nearby (although not adjacent) housing units.

The other major source of U.S. crime statistics, the Uniform Crime Report (UCR) Program of the FBI, relies on police reports to monitor crime nationally and at other geographic levels. Law enforcement agencies report counts of crimes by type, which are then published both in disaggregated form and summarized to higher geographic levels. Among the primary limitations of the system are incomplete coverage, irregular reporting by participating agencies, and possible differences in local interpretations of the UCR's categories of crimes. Although now approximately 17,000 law enforcement agencies participate in the system (Barnett-Ryan, 2007), covering approximately 96.3% of the population (FBI, http://www2.fbi.gov/ucr/cius2009/documents/aboutucrmain.pdf, downloaded 29 Jun 2011) in 2009, undercoverage has been a serious issue until recently. Additionally, some agencies, although participating, may submit only partial reports during the year, creating additional problems of missing data (Maltz, 2007). To prepare national and state estimates, the FBI imputes for missing data, but it does not release the imputations at the level of the individual law enforcement agency.

In the introductory chapter of their co-edited book, *Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR*, Lynch and Addington (2007b) comment on the frequent apparent division of researchers into opposing camps debating the usefulness of one crime series to the virtual exclusion of the other. Lynch and Addington instead advocate the principle of *complementarity* introduced by Biderman and Lynch (1991). This principle asserts that each series measures crime differently and provides complementary sources of information when these differences are taken into account. Lynch and Addington (2007b, pp. 6-7) remark

> Much of the research community has ignored the concept of complementarity between the UCR and NCVS. This lack of attention is not surprising given two common misunderstandings concerning crime statistics that often inhibit researchers' ability to view the NCVS and the UCR as complementing each other. One of these misunderstandings concerns two related assumptions that there is an objective definition of the crime problem and that a statistical system can assess this problem without distortion. The second misunderstanding arises from ignorance regarding the social organization of the UCR and NCVS statistical systems.

Their edited volume (Lynch and Addington, 2007a) is devoted to analyzing the multiple sources of differences between the measurements achieved by the two systems. Rennison and Rand (2007) provided the past history and current status of the NCVS, and Barnett-

Ryan (2007) similarly summarized the UCR. McDowall and Loftin (2007) reviewed the considerable literature on the divergence of the UCR and the NCVS, that is, the degree to which the series differ in level or trend. Contributors (Catalano, 2007; Planty, 2007; Addington, 2007) examined methodological differences or changes in methodology as possible sources of divergence. Cohen and Lynch (2007), using alternative sample surveys of visits to emergency rooms, assembled a plausible case that components of aggravated assault in the NCVS appear considerably understated. Characterizing his evidence as circumstantial, Rosenfeld (2007) argued that reporting of aggravated assaults in the UCR may have gradually improved over time. His findings raise the possibility that some other components of the UCR may also have improved.

A complete analysis of NCVS/UCR differences is clearly beyond the scope of our research. In spite of this, the UCR remains a strong candidate as a source for auxiliary information possibly useful in designing the sample for the NCVS or other crime surveys. Last year, we reported an initial set of regression models to express the relationship between NCVS and the UCR at the county level (Li and Fay, 2010). We observed that the relationship was stronger for property crime than for violent crime, the major categories of crime covered by the two systems. In Section 2, we extend the previous analysis at the county level in an effort to clarify the sources of this variation in predictive accuracy.

The research community has extensively studied correlates of crime, including the commonplace variables of social science, namely, demographics and SES. We also examine the extent to which these variables might be usefully incorporated into the sample design. Section 3 examines these variables at the person and household level, where they potentially can guide decisions on sampling within PSU.

Because the 2010 Census will provide data for only a limited set of characteristics compared to the long-form in Census 2000 and previous censuses, the American Community Survey (ACS) has gained importance as the best available substitute for the census long-form information. But ACS estimates, including the ACS 5-year estimates, are subject to greater sampling variance than the previous long-form estimates. Section 4 presents an analysis, applicable to the NCVS design, to guide the choice between ACS sample information at the tract or block-group levels. This methodological approach would appear to extend to other sampling designs intending to incorporate ACS data.

2. Variation in Crime Rates at the County Level

2.1 The Types of Crime and Their Trends

The NCVS does not ask victims of crimes to classify them directly. Instead, the NCVS questionnaire uses the strategy of first asking the respondent to identify any incidents of crime that might be in scope within the 6-month recall period, using multiple cues to aid recall. The questionnaire then asks more detailed information about each incident, which is used later to classify each crime by type. Questions on crimes at the household level are directed to a single household respondent, who reports on property crimes for the household. Household crimes eventually become classified as *burglary*, *motor vehicle theft*, and all other types of *theft*, depending on the detailed information provided by the household respondent.

The NCVS uses self-response for persons age 12 and over. Eligible respondents report individually on pocket picking and purse snatching, which the NCVS classifies as *personal theft*. More importantly, self response is used for violent crimes, classified as *rape and sexual assault, robbery, aggravated assault,* and *simple assault.* The precise boundaries between aggravated and simple assault are not easily stated, but aggravated assault involves being threatened or harmed by a weapon or other serious bodily injury. These broad types—rape/sexual assault, robbery, aggravated assault, simple assault, personal theft, burglary, motor vehicle theft, and other theft—are used extensively in the Bureau of Justice's primary annual publication, the most recent of which is "Criminal Victimization, 2009" (Truman and Rand 2010). Finer detail is available from some publications and the public use files; for example, rape/sexual assault is categorized into 8 subtypes: completed rape, attempted rape, sexual attack with serious assault, sexual attack with minor assault, sexual assault without injury, unwanted sexual contact without force, verbal threat of rape, and verbal threat of sexual assault.

Three of the four categories of violent crime as defined by the NCVS—rape/sexual assault, robbery, and aggravated assault—are grouped together as *serious violent crime* and are conceptually close to violent crime in the UCR. Forcible rape (which includes attempted rape but excludes statutory rape) in the UCR excludes crimes NCVS categorizes as sexual assault.

The largest single contributor to the NCVS violent crime rate is simple assault. Simple assault lacks a UCR analogue, although the UCR collects and separately reports data on arrests for simple assault. Simple assault consequently creates an asymmetry between the NCVS and UCR violent crime rates, dominating the first while being omitted from the second. Criminal homicide, including murder, is included in UCR statistics but out of scope for the NCVS.

Table 1 illustrates the alignment of the two sources for two years: (1) 1996, the first full year of the post-1990 redesign and a few years after the 1993 introduction of major methodological changes that increased the reported crime rate by approximately 50%; and (2) 2009, the most recently published year. The overall pattern of change over this specific period is characteristic of the change over the last decade, with overall decreases in both series but generally larger ones in the NCVS.

Besides the use of different denominators and the other differences analyzed in Lynch and Addison (2007a), NCVS respondents state that they did not report a large proportion of the crimes committed against them to the police. Thus, it is difficult to compare the published rates from the two sources. Instead, Table 1 usefully shows that, according to both sources, aggravated assault occurs more frequently than robbery, and that robbery occurs more frequently than rape. Again, simple assault is the dominant category of violent crime in the NCVS.

Although the primary focus here will be on violent crime, the NCVS and UCR similarly agree on the ranking of theft as the most common form of property crime, followed by burglary and motor vehicle theft.

2.2 Analyzing the Usefulness of the UCR in the Design

The majority of studies comparing the UCR and the NCVS focus on national data, because little geographic detail is publicly available for the NCVS. Like a few other researchers (e.g., Zhang, Woodburn, and Scheuren, 2009), we have instead worked with

the Census Bureau's internal files to analyze the relationship between the UCR and the NCVS geographically. The UCR data are potentially useful in the redesign of the NCVS to the extent that it can be used to predict geographic differences in the underlying crime rate as measured by the NCVS.

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).

In Li and Fay (2010), we presented a set of regression relationships for county-level crime rates over the 5-year period 2001-2005. We selected this 5-year period to reflect recent crime but avoided 2006, a year exhibiting unusual estimates nationally and particularly for rural areas (Rand and Catalano, 2007; Rand 2008). Regressions were fitted separately for property crimes, violent crimes, violent crimes excluding simple assault (that is, serious violent crime), and simple assault. Counties were partitioned into self-representing and non-self-representing. The analysis was restricted to counties with complete or almost complete UCR coverage and reporting; these counties represented approximately 50% of the total population. For convenience, the results from this earlier work are provided here as Table 2.

For simplicity this paper will focus on self-representing counties only, primarily because the results that emerged from these generally large counties were more stable. The SR results for property crimes and for violent crimes are specifically related to our new results. Table 3 presents the regression analysis for property crime for three different time periods: (1) 1996-2005, spanning a 10-year period where NCVS used essentially the same first-stage sample of counties; (2) 2001-2005, from Table 2; and (3) 2006-2009, a 4-year period taking advantage of recent NCVS data, but using as an independent variable the UCR rates for 2006-2007 because more recent county-level UCR data were unavailable from the University of Michigan. The new results are based on the methods we used previously (Li and Fay, 2010).

The dependent means in Table 3 exhibit the expected downward trend, but the regression coefficient for the matching UCR grows slightly, from 2.26 for 1996-2005 to 2.98 in 2006-2009. For both 2001-2005 and 2006-2009, the intercept term of the model, which predicts the NCVS property crime rate when the UCR rate is 0, is less than 50 percent of the dependent mean, suggesting a reasonably successful level of prediction. In Table 2, the analogous comparison for violent crime in SR areas, a dependent mean of 21.6 and an intercept of 16.0, suggests a considerably lower relative success in prediction.

Table 2: Weighted County-Level Regressions Based on 5-Year Averages in Counties with the Highest Rates of Complete UCR Reporting, 2001-2005.

Counties	*Dependent Mean*	*Intercept*	*SE*	*Beta*	*SE*
			Property Crimes		
NSR	174.6	55.0	21.2	3.29	0.60
SR	167.0	72.2	9.2	2.60	0.23
ALL	169.0	68.9	8.9	2.75	0.23
			Violent Crimes		
NSR	22.0	21.1	2.8	0.24	0.63
SR	21.6	16.0	1.4	0.97	0.21
ALL	21.7	17.5	1.2	0.80	0.20
			Violent Crimes Excluding Simple Assault		
NSR	6.9	5.3	1.0	0.42	0.23
SR	7.8	3.8	0.6	0.70	0.09
ALL	7.5	4.2	0.5	0.65	0.08
			Simple Assault		
NSR	15.1	15.8	2.3	-0.18	0.51
SR	13.8	12.2	1.0	0.27	0.15
ALL	14.1	13.3	0.9	0.15	0.15

Source: Table 1, Li and Fay (2010).

Table 3: Regression Prediction of County-Level NCVS Property Crime Rates from the UCR Rates in Self-Representing Counties with the Highest Rates of Complete UCR Reporting, for 1996-2005, 2001-2005, and 2005-2009 (Standard errors in parentheses).

	NCVS property crime, 1996-2005	*NCVS property crime, 2001-2005*	*NCVS property crime, 2006-2009*
Dependent mean	201.2	167.0	146.5
Intercept	113.6	72.2	46.2
	(9.3)	(9.2)	(9.8)
UCR property crime, 1996-2005	2.26		
	(0.22)		
UCR property crime, 2001-2005		2.60	
		(0.23)	
UCR property crime, 2006-2007			2.98
			(0.27)

Because effective sample design for violent crime was our primary goal, we subsequently studied the interrelationships between the components of violent crime in the NCVS and the UCR. Table 4 presents the results for 1996-2005. When the UCR violent crime rate is used to predict the NCVS violent crime rate, the intercept is 23.14 relative to a dependent mean of 30.28, similar to the relationship between the dependent mean and intercept for violent crime in Table 2. When the UCR violent crime rate is disaggregated into its three component parts, however, the corresponding coefficients are strikingly different from each other. The coefficient for UCR forcible rape is 31.04, suggesting that this variable is not only useful as a predictor for NCVS rape/sexual assault but is an effective symptomatic indicator for other components of NCVS violent crime. The coefficient for robbery, 2.76, is quite strong, whereas the coefficient for aggravated assault is negative.

Table 4: 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)

Table 4 also includes the results from separately analyzing three of the four components of the NCVS violent crime rate. Two sets of results are straightforward: UCR robbery is the best single predictor of NCVS robbery, and similarly UCR forcible rape is the best single predictor of NCVS rape. The coefficients, 1.83 and 3.36, respectively, can exceed 1.0 for several reasons, including differences in the denominator of the rates and the possibility that the UCR crime rate might be partially successful in predicting unreported crimes as well as reported ones. In the regression for NCVS aggravated assault, the coefficient for UCR aggravated assault is estimated to be 0.00. Instead, it is UCR forcible rape which stands out as a strong predictor.

NCVS simple assault is the remaining component of NCVS violent crime. By comparing the coefficients for UCR aggravated assault across the four models, the geographic relationship between UCR aggravated assault and NCVS simple assault is apparently negative. It also appears that UCR robbery is primarily predicting NCVS robbery and not other components of violent crime, but that UCR forcible rape is a symptomatic predictor of both simple assault and aggravated assault, as well as a direct predictor of NCVS rape and sexual assault.

Although not shown here, the regression findings are broadly similar for the periods 2001-2005 and 2006-2009.

Regional tabulations of the NCVS are published through 2005 but not thereafter. Table 5 compares aggravated assault at the regional level between the NCVS and UCR. Although standard errors are not readily available, the Northeast appears to have lower rates of aggravated assault in both the NCVS and the UCR. In the NCVS, the Midwest and South have estimates that are quite close to the national average, but the West stands out for its high rate. In the UCR, however, the Midwest exhibits a rate close to the Northeast, while the South exhibits the highest rates and the West is barely above the national average. This partial disagreement between the UCR and NCVS at the regional level possibly is symptomatic of the poor performance of the UCR aggravated assault as a predictor of NCVS aggravated assault.

Table 5: Comparison of Aggravated Assault Rates from the NCVS and UCR by Region, 1996-2005

	U.S rate	*Northeast*	*Midwest*	*South*	*West*
			NCVS		
1996-2000	7.5	5.5	7.6	7.4	9.2
2001-2005	4.6	3.5	4.8	4.5	5.4
1996-2005	6.0	4.5	6.2	5.9	7.3
			UCR		
1996-2000	3.6	2.9	3.0	4.2	3.9
2001-2005	3.0	2.3	2.5	3.6	3.1
1996-2005	3.3	2.6	2.7	3.9	3.5

Sources: Table 57 in the series "Criminal Victimization in the United States, 1996 Statistical Tables," and subsequent years, downloaded 13 July 2011. The UCR counts were provided by the FBI at the state level through www.ucrdatatool.gov/search/crime/statebystate.cfm.

2.3 County-Level Conclusions

The lack of any consistent relationship between UCR and NCVS aggravated assault rates is both notable and puzzling. The UCR is arguably a statistical system with many of the characteristics of an administrative data system. In general, the statistical usefulness of administrative data can vary considerably. Typically, the relationship between an administrative record system and a statistical objective is imperfect, but in this case it is virtually non-existent for aggravated assault if rape and robbery are controlled. At the same time, the observed relationships between the UCR and NCVS for rape and for robbery are reasonable.

Returning to the sample design questions at hand, the evidence favors omitting UCR aggravated assault as a stratification variable while including UCR rape and robbery. This approach in fact was apparently previously taken by the Census Bureau, on the basis of an analysis approximately 20 years ago (Barbara Blass, U.S. Census Bureau, personal communication). Thus, we may have reconfirmed, using appropriately recent data, a statistical relationship that had previously been observed but largely forgotten.

3. Person-Level Analysis

3.1 Methods

In other research related to the sample design, we have detected relatively small interclass correlations for the reporting of violent crime for households within segments, for persons within households, and for 6-month recall periods across time (Fay, 2010). Thus, an analysis that can predict which individuals are at increased risk of victimization relative to others is likely to provide sensible guidance on how to draw the NCVS sample within sampled counties, particularly if the predictor variables are measured at the block or housing unit level and are relatively stable across time.

Starting in 2005, approximately 85-90% of the housing unit records on the Census Bureau's internal files contain Census 2000 geographic information down to the block level. (The exceptions are the units sampled through the address permit frame, which represents units not in the census.) We matched the NCVS sample cases for 2005-2009 to selected block and tract-level characteristics from Census 2000. Specifically, we examined the proportion renter at the block and tract level. Tenure is a short-form item in both Census 2000 and the 2010 Census. As an alternative, we used the NCVS reported tenure as a predictor in some models.

We also studied the proportion of the households with income less than $30,000 in 1999. This characteristic was measured on the long form in Census 2000 and is available at tract but not the block level. Finally, we included median age at the tract level as a predictor. As the dependent variable, we examined whether an individual had been a victim of one or more violent crimes during the 6-month reporting period. We used proc surveylogistic in SAS to account for the complex sample design.

3.2 Results

Table 6 presents the coefficients of logistic regressions for the complete matched data set and for SR and NSR separately. In the first three models, the proportion renter at the block level is used, whereas the last three the reported NCVS tenure status is substituted.

The estimated coefficients are highly significant and typically substantively important in the national and SR models, and generally so in the NSR models. The coefficients of the models are roughly similar but far from identical in SR and NSR areas, however, so that it is unlikely that a single model completely captures the relationships among these variables in predicting violent crime.

When other models were fitted with both block-level proportion renter and NCVS reported tenure, the block-level results did not add significantly to the fit. Similarly, when both tract-level and block-level proportion renters were added to the model (without including NCVS reported tenure), the block-level proportion was the dominant predictor. Interestingly, at least with respect to tenure, there appears to be little or ecological effect, that is, information on tenure is simply more predictive the closer it is to the individual.

Although median age contributes significantly to most of the models, we confirmed that reduced models omitting this variable retained most of the predictive value. Our primary recommendation was to use tenure and household income to stratify the sample within PSUs.

Table 6: Logistic Regression Prediction of Individual Victimization by Violent Crime in the Previous 6 Months, NCVS, 2005-2009

	All	*SR*	*NSR*	*All*	*SR*	*NSR*
Intercept	-4.681	-4.470	-4.855	-4.759	-4.603	-4.795
	(0.153)	(0.160)	(0.267)	(0.148)	(0.148)	(0.274)
Median age (tract)	-0.016	-0.022	-0.009	-0.014	-0.019	-0.011
	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.006)
Proportion low income (tract)	0.948	1.142	0.543	0.880	0.972	0.590
	(0.121)	(0.168)	(0.224)	(0.115)	(0.150)	(0.206)
Proportion renter (block)	0.663	0.439	1.022			
	(0.064)	(0.094)	(0.104)			
Renter (NCVS)				0.722	0.627	0.847
				(0.039)	(0.047)	(0.066)

We wish to note that while we were developing these logistic regression models, staff at the Census Bureau conducted a search for variables useful in stratification, arriving at similar conclusions on the importance of income and tenure.

4. ACS Estimates as Within-PSU Stratification Variables

4.1 The Issue

In the previous section, we examined two variables, tenure and age, available from Census 2000 and soon to be available from the 2010 Census, asked of all persons. Income was asked on a sample basis in Census 2000, but the ACS now takes the place of a long-form census to provide sample data on income for geographic units as small as tracts or block groups. (Each tract is divided into block groups, with three block groups in a tract being typical.) In the past, the apparent practice was to give no special attention to the presence of sampling error in auxiliary data used for stratification of other sample surveys. In early 2011, however, it became apparent that a more formal analysis of the effect of sampling error was desirable. We had assumed that income data would be used at the tract level to mitigate the effect of sampling error while staff at the Census Bureau were considering using the ACS at a finer geographic level, equivalent to block groups or smaller. The issue of trading off greater tract-level reliability against presumably greater block group detail formed the impetus for the research investigation to be reported here.

4.2 Method

We developed an analysis based on publicly available data. ACS estimates with accompanying standard errors (expressed as MOE, measures of error, equal to 1.65 times an estimated standard error) were not available for low income households, so instead we analyzed the percent of families in poverty at the tract and block group levels. The MOE for the estimated proportion, $\hat{p}$, of families in poverty was not given directly, but we estimated it by linearization

$$MOE^2(\hat{p}) = \frac{(MOE^2_{poorf} + \hat{p}^2 MOE^2_{totf} - \hat{p}(MOE^2_{totf} + MOE^2_{poorf} - MOE^2_{nonpoorf}))}{totf^2}$$
$$= \frac{((1-\hat{p})MOE^2_{poorf} - \hat{p}(1-\hat{p})MOE^2_{totf} + \hat{p}MOE^2_{nonpoorf})}{totf^2}$$

where MOE_{poorf} , MOE_{totf} , and $MOE_{nonpoorf}$ are the measures of error for the number of poor families, the number of families, and the number of nonpoor families, respectively, and $totf$ is the estimated number of families. In releasing ACS sample estimates of 0, the Census Bureau frequently supplies a positive MOE for them, but for purposes of this analysis we changed these estimates to 0.

In analyzing the relative value of the information provided by the ACS, we adopted the ANOVA-like decomposition:

$$p_{(bg)} = p_c + (p_t - p_c) + (p_{(bg)} - p_t)$$

We will treat the three terms of this decomposition as (approximately) orthogonal. Because our interest is in stratification of tracts and block groups within counties, using 5-year ACS data, we will treat p_c as an unknown fixed effect. We will treat the tract proportions and block group proportions as random effects, with

$$Var(p_t) = \sigma^2_t$$
$$Var(\hat{p}_t) = \sigma^2_t + \sigma^2_{e(t)}$$
$$Var(p_{(bg)}) = \sigma^2_t + \sigma^2_{(bg)}$$
$$Var(\hat{p}_{(bg)}) = \sigma^2_t + \sigma^2_{(bg)} + \sigma^2_{e(t)} + \sigma^2_{e(bg)}$$

We can use the sampling variances derived from the Census Bureau's calculations of tract and block group level MOE's to estimate $\sigma^2_{e(t)}$ and $\sigma^2_{e(t)} + \sigma^2_{e(bg)}$, respectively. By computing the population variance of $\hat{p}_t$ and $\hat{p}_{(bg)}$ over a large number of tracts or block groups, we can estimate the right-hand sides of the second or fourth equation, and, by subtraction, estimate $\hat{\sigma}^2_t$ and $\hat{\sigma}^2_{(bg)}$. By working with the orthogonal components separately, the empirical best linear unbiased prediction (EBLUP) under this model is

$$\hat{p}_{(bg)} = \hat{p}_c + \frac{\hat{\sigma}^2_t}{\hat{\sigma}^2_t + \hat{\sigma}^2_{e(t)}}(\hat{p}_t - \hat{p}_c) + \frac{\hat{\sigma}^2_{(bg)}}{\hat{\sigma}^2_{(bg)} + \hat{\sigma}^2_{e(bg)}}(\hat{p}_{(bg)} - \hat{p}_t)$$

For example, when this calculation for is carried out for 1720 block groups in Los Angeles County in the upper third of percent renter in the 2000 Census, the coefficients are .82 and .33, respectively. Thus, the EBLUP would pull this group of tracts somewhat

toward a common mean, but shrink the block group variation within tract considerably. Regrouping the terms of the EBLUP gives

$$\hat{p}_{(bg)} = \frac{\hat{\sigma}^2_{e(t)}}{\hat{\sigma}^2_t + \hat{\sigma}^2_{e(t)}}\,\hat{p}_c + \left(\frac{\hat{\sigma}^2_t}{\hat{\sigma}^2_t + \hat{\sigma}^2_{e(t)}} - \frac{\hat{\sigma}^2_{(bg)}}{\hat{\sigma}^2_{(bg)} + \hat{\sigma}^2_{e(bg)}} \right) \hat{p}_t + \frac{\hat{\sigma}^2_{(bg)}}{\hat{\sigma}^2_{(bg)} + \hat{\sigma}^2_{e(bg)}}\,\hat{p}_{(bg)}$$

With this regrouping, the coefficient of the middle term is .82 - .33 = .49. The EBLUP for this set of block groups would put about 50% more weight on the tract level estimate than the block level estimate.

Because the interest is in stratification within the county, the two coefficients of the EBLUP can be normalized to sum to 1. In this example, the normalized weights become .60 on the tract estimate and .40 on the block group estimate. The normalized form would produce the same stratification decisions as the EBLUP.

Table 7 is based on an analysis of the District of Columbia and counties in California, Florida, Georgia, Illinois, Maryland, Michigan, New York, North Carolina, Ohio, Pennsylvania, and Texas. In each case, the analysis was restricted to block groups in the upper third of the distribution of percent renters. For simplicity, Table 7 shows only one of two analyses that were performed. The other analysis, which substituted an estimate of variance for the block group by adjusting the tract-level variance for the differences in sample size, gave results placing somewhat more evidence on the block group. On the basis of this evidence, we recommended to the Bureau of Justice Statistics that the Census Bureau use a weighted estimate for the income groupings, giving weight .4 to the tract estimate and .6 to the block group estimate.

Table 7: Estimation of Optimum Normalized Weights for Combining the ACS Tract and Block-Group Estimates

	Number of block groups in each county				
	500+	*200-499*	*100-199*	*50-99*	*25-49*
Number of counties	5	22	37	57	109
Number of block groups analyzed	4291	6556	5142	3859	3795
Tract weight	.52	.40	.47	.45	.42
Block group weight	.48	.60	.53	.55	.58

5. Discussion

Over the last few years, BJS has adopted a multi-faceted research strategy to improve the NCVS. Examination of sampling design options and refinements for the current redesign of the survey is part of the overall effort. Our intention in this paper is to document the analyses shaping our own recommendations and to possibly provide an additional perspective on the NCVS and its relationship to the UCR.

Our county-level analysis can contribute to plans to stratify the PSUs in non-self-representing areas. The yield from the UCR, by itself, is useful but not strong as might be hoped; but the results show that attempts to improve them should focus on improving the prediction for both aggravated and simple assault as measured by NCVS. Our person-

level analysis confirms the importance of variables that, to our knowledge, the Census Bureau plans to incorporate into the stratification for within-PSU selection.

The composite of tract and block group estimates that we propose was constrained by resources to use publicly available ACS summaries. A more detailed analysis could be conducted using the Census Bureau's internal ACS files could be used to produce a more fine-grained analysis. We estimated key variance components by subtraction, which required averaging over a large sample. More precise estimates could be developed by direct estimation of the variances of the differences between block group and tract rates, but these are not publicly available. The Census Bureau may wish to pursue this idea further, because many of their household surveys are likely to incorporate ACS estimates into the stratification.

Acknowledgements

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

We want to thank Mike Planty for his consultation and contributions at different stages of this research. We thank the Bureau of Justice Statistics for the original support for this work.

References

Addington, L.A. (2007), "Using NIBRS to Study Methodological Sources of Divergence Between the UCR and NCVS," in Lynch, J.P. and Addington, L.A. (eds.) (2007), *Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR*, pp. 225-250.

Barnett-Ryan, C. (2007), "Introduction to the Uniform Crime Reporting Program," in Lynch, J.P. and Addington, L.A. (eds.) (2007), *Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR*, pp. 55-89.

Biderman A.D. and Lynch, J.P. (1991), *Understanding Crime Incidence Statistics: Why the UCR Diverges from the NCS*, Springer-Verlag, New York, NY.

Catalano, S.M. (2007), "Methodological Change in the NCVS and the Effect on Convergence," in Lynch, J.P. and Addington, L.A. (eds.) (2007), *Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR*, pp. 125-155.

Cohen, J. and Lynch, J.P. (2007), "Exploring Differences in Estimates of Visits to Emergence Rooms for Injuries from Assaults Using the NCVS and NHAMCS," in Lynch, J.P. and Addington, L.A. (eds.) (2007), *Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR*, pp. 183-222.

Fay, R.E. (2010), "Components of Sampling Variance in the NCVS Design," unpublished draft document dated 15 Dec 2010.

Li, J. and Fay, R.E. (2010), "Assessing the Association Between the Crime Rates from the National Crime Victimization Survey and the Uniform Crime Reporting Program," presented at the Joint Statistical Meetings, Vancouver, British Columbia, Canada, Jul 31 - Aug. 5, 2010.

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.

Lynch, J.P. and Addington, L.A. (2007b), "Introduction," in Lynch, J.P. and Addington, L.A. (eds.) *Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR*, pp. 3-13.

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

McDowall, D. and Loftin, C. (2007), "What Is Convergence, and What Do We Know About It?" in Lynch, J.P. and Addington, L.A. (eds.) (2007), *Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR*, pp. 93-121.

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.

Planty, M. (2007), "Series Victimizations and Divergence" in Lynch, J.P. and Addington, L.A. (eds.) (2007), *Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR*, pp. 156-182.

Rand, M.R. (2008), "Criminal Victimization, 2007," NCJ 224390, issued by the Bureau of Justice Statistics, Dec. 2008.

Rand, M. and Catalano, S. (2007), "Criminal Victimization, 2006," NCJ 219413, issued by the Bureau of Justice Statistics, Dec. 2007.

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

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

Rosenfeld, R. (2007), "Explaining the Divergence Between UCR and NCVS Aggravated Assault Trends," in Lynch, J.P. and Addington, L.A. (eds.) (2007), *Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR*, pp. 251-268.

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

Zhang, Z., Woodburn, L., and Scheuren, F. (2009), "Analysis of Possible Nonresponse Bias in the National Crime and Victimization Survey," presented at the 2009 Federal Committee on Statistical Methodology Conference, Washington, DC, Nov. 2-4, 2009, available at http://www.fcsm.gov/09papers/Zhang_II-A.pdf.