NCVS Task 4 Report: Summary of Options Relating to Local Area Estimation

Submitted by:
Westat
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Submitted to:
Bureau of Justice Statistics
810 Seventh Street, NW
Washington, DC 20531

May 19, 2010

Draft 1

This project was supported by Grant No. 2007-DD-BX-K055 awarded by the Bureau of Justice Statistics. The opinions, findings, and conclusions or recommendations expressed in this document are those of the authors and do not necessarily reflect the views of the U.S. Department of Justice.

Authorship and Acknowledgments:

This report was written by a team of six individuals. David Cantor and Tom Krenzke jointly wrote the Summary and Recommendations for Next Steps and Chapter 1. They also contributed to Chapters 2, 3, and 4. Diana Stukel wrote Chapter 2 and Lou Rizzo wrote Chapters 3 and 4. Graham Kalton and Robert Fay reviewed and provided comments on the report.

Table of Contents

<u>**Appendices**</u> <u>**Page**</u>

Tables **Page**

Recommendations for Next Steps

This report reviews three related approaches for generating Local Area Estimates (LAEs) for the NCVS: 1) Direct estimates, 2) Small area estimates (SAEs) and 3) Adding sample using alternative methods. There are both short and long term activities that are needed to implement these different approaches. It isn't necessary to rule out any one of these approaches at this point. There are research and development activities that are common to all of these approaches. Prior to deciding on a particular approach, it is necessary to address basic questions about the level of geography and the critical measures to be estimated. Below, these recommendations for next steps are summarized. The reader is referred to the end of each major section in the body of the report for more detailed discussions of these recommendations.

Questions to be Answered

There are two important questions that need to be addressed for a successful local area program. One is determining the geographic unit(s) of interest. Finding the appropriate unit or sets of units is critical to maximizing the utility of the data. Users naturally want data for as small a geographic unit as possible. The discussions in Chapters 2 through 4 primarily concentrate on states, MSA's, or central cities. One could also generate estimates for groups of states, individual cities or counties. A third approach would form "crime statistics areas" that would be groups of small states and groups of counties in large states. For purposes of planning, it seems most appropriate to assume that states and MSAs would be the most realistic units to plan around, given the NCVS sample sizes.

A second important question is the type of estimate of interest. The discussion in this report concentrates on producing incidence rates (with the exception of Chapter 4). Feedback from data users suggest that crime characteristics are also an important element in a set of LAE estimates. In many ways crime characteristics are the strength of the NCVS, relative to other data sources, and any methodology that is eventually adopted for LAEs should seriously consider producing these as part of the program.

Generate LAEs with Current Samples

In the short term, using the NCVS sample to generate LAEs for areas that have sufficient sample size should be seriously considered. It is the least expensive alternative and it would provide estimates in the shortest period of time. Maximizing the number of areas for publication will involve aggregating data over multiple years. At least according to the feedback from potential state-level data users, publishing these data on a bi- or tri-annual basis would be very useful. Chapter 2 discusses the extent to which current datasets can be used for this purpose for producing state and/or MSA estimates.

If there are additional monies to add sample, Chapter 2 presents one scenario of adding sample to areas that cannot support a LAE by itself. Other methods might be used to allocate this additional sample, such as concentrating it in fewer areas or rotating the sample across different regions to provide estimates.

Assess the Feasibility of Using Small Area Estimation Techniques

Another short term approach for generating LAEs is Small Area Estimation (SAE). Chapter 3 describes the methodologies for estimating small area models using auxiliary data. The methodology can take several different forms, including a unit-level or area-level model. An area-level model was estimated for MSAs in Krenzke, Li, and Cantor (2009) using UCR and Census data. This analysis found the UCR rates were highly correlated with NCVS rates for property crimes and Census data were correlated to a lesser extent. While the small area model did not substantially increase the precision of the direct estimates, the model indicated some promise if effective predictors could be collected. This type of work should continue with more recent NCVS data, perhaps at other levels of geography. An important step would be to collect other covariates that could be used as predictors in a small area model. By collecting and testing predictors it would be possible to assess the feasibility of a small area model at either a unit or area level.

These activities can be completed without any field data collection and should be relatively inexpensive. This work could also be valuable if SAE methods are used in conjunction with supplementing the sample, as discussed in Chapter 4.

It is unclear how the NCVS constituencies would react to estimates based on a statistical model. If UCR data are used, this may also raise some questions. For example, reaction from some of the participants at the data-user workshop was somewhat mixed when this topic was discussed. Some jurisdictions were skeptical about the validity of the UCR data in their jurisdiction and would be suspicious of estimates based on these data. Others said they would consider the model-based estimates as official government statistics and they would use with confidence.

Supplement with Less Expensive Data Collection Methods

A third approach that addresses the concerns associated with a small area model is to administer a supplementary survey in the local areas for which estimates are to be derived. By basing the estimates on direct measures of victimization in the area, it would reflect idiosyncratic events or trends in the local jurisdiction. It would also address possible concerns by policymakers that the data are based entirely on a statistical model. The supplementary survey would utilize methods that were less expensive to administer than the main NCVS (e.g., mail, web or telephone). Any supplemental collection would have to develop methods to combine the information with the national NCVS. It will need to account for the measurement differences with the national NCVS when combining the two sources (e.g., time-in-sample bias, nonresponse bias, mode effects).

We divided the discussion of less expensive data collection methods into two types: a one-phase sample, and a two-phase sample.

A One-Phase Sample

This involves either a short mail questionnaire or a telephone survey with the full NCVS. The mail survey would resemble the current NCVS screener. A telephone interview would administer the entire NCVS interview, including the detailed incident form. In either case, the result would provide an additional covariate, or an additional set of estimates on crime statistics that could be combined with the main NCVS through small area modeling techniques discussed in Section 4.2. Questions remain as to the potential of this approach and therefore some further data driven research is necessary to gauge the correlation between the supplemental sample estimates, the NCVS estimates and other covariates. It would be of interest to compare the mail approach which received reports from a single household respondent, to a telephone survey that administered the entire NCVS interview to a randomly selected individual. This would be completed across a number of different areas (e.g, 30 – 50) to develop small area models that could be evaluated. Prior to conducting this research, it would be useful to conduct simulation studies that assessed the potential of this approach under different assumptions of the cost and empirical correlations. A discussion of this analysis is provided in Section 4.2.

A Two-Phase Sample with the Full NCVS

A two-phase survey is described in Section 4.1 as using a mail survey in the first phase and a full NCVS interview in the second phase.

Initial research would test the effectiveness of the phase-one survey to sample households for a follow-up, in-person interview. The phase-one interview would be evaluated by how accurately it could identify households and persons as it relates to particular types of victimizations. The two phase design makes strong assumptions about the ability of the screening interview to identify victims. The development of this method may be something BJS could consider as part of a longer term research program that could be incorporated into pilot tests related to the single phase design. Both require research into developing a screening interview that could be used to identify victims. If this proves successful, a two-phase method could move forward.

Regardless if a one-phase or two-phase method is used, the timing of the pilot testing should be done once the final NCVS design is in place. One would expect the measurement properties of the new design to be somewhat different from the current NCVS. If this is true, then it would be necessary that any methodology reflect the new design.

Introduction 1

The National Crime Victimization Survey (NCVS) has produced national-level estimates of crime and victimization since the early 1970s, and has provided important insights into national crime trends. Local policymakers, however, would find the survey data more useful if crime statistics could be produced at a local level. This was demonstrated in a 2009 meeting[1] on state crime statistics, where state government representatives expressed their needs for better crime data for their states. These representatives made little or no use of national crime estimates. Much of their decision making was driven by individual cases, on data from the Uniform Crime Reports, or on results from a local area survey. Although a small number of states ran their own crime surveys periodically; most found it difficult, if not impossible, to obtain the funding to do so.

As described in Krenzke and Cantor (2009) (referred to hereafter as the Task 2 Memo), there is a great deal of work being conducted on generating local area estimates (LAEs), as well as world-wide. There are two general approaches to producing LAEs. As discussed in Chapter 2, standard survey estimates, or 'direct estimates', are based on survey data derived directly from the sample units in the local area. The American Community Survey (ACS) is the gold-standard example, and it illustrates both the amount of resources and the extent of effort required to produce quality LAEs. Each year data are collected from about two million different households (sampling rate of about 1.5 percent after nonresponse), and through accumulation of data over a 5-year period (sampling rate of about 7.5 percent), the ACS produces direct estimates for local areas as small as block groups, based on a sample size of less than 50 housing units. While the NCVS sample is distributed proportionately across the Nation, the ACS distributes its sample households differentially, that is, smaller areas receive a larger proportion of the sample in order to improve the reliability of data produced for smaller areas.

The British Crime Survey (BCS) is another survey that produces direct LAEs. The BCS is a victimization survey, where respondents report on property and violent victimizations. The BCS sample design allocates to each of their 42 Police Force Areas (PFA) in larger proportion than would

[1] "BJS Meeting on the Redesign of the NCVS", July 29, 2009. Presentations were given by representatives of the following states: Arizona, Colorado, Maine, Minnesota, Idaho, Illinois and Vermont, and by representatives of Council of State Governments, Virginia Criminal Sentencing Commission, Westat, and the University of Missouri-St.Louis. Topical discussions included the use of NCVS at the state or local level, ways to produce LAEs, and a discussion of direct versus model-based data for state-level estimates.

be the case if allocated to produce national estimates alone. The resulting sample yields approximately equal sample sizes of about 1000 to 1500 interviews in each PFA, with the exception of one larger PFA that requires over 3000 interviews. Disproportionate sampling (or oversampling), as in the ACS and BCS, is a means of ensuring enough sample is drawn in local areas of interest to facilitate direct estimation of designated survey characteristics. While attempting to achieve quality estimates at the local level, disproportionate sampling results in an increase in the variances associated with national estimates, since the national effective sample size is reduced with the use of disproportionate sampling.

The BCS was initially designed to produce estimates of police performance within PFAs.[2] Estimating victimization rates was a secondary purpose and one which the designers recognize cannot be done very precisely. The BCS performance indicators have not been universally accepted by local area government officials. Two important issues that are relevant to think about for the NCVS has been the reliability of the estimates and the geographic area covered. Even with the disproportionate sampling described above, there had to be about a seven percentage point difference for the performance indicators to be significant. A second issue is that the PFA is a relatively large area when discussing individual police performance. When police reacted to evaluations, they would comment that the data did not provide an idea of what part of the force the data pertained to.

This point emphasizes the need to carefully decide on the level of geography to be used. It affects the utility of the information, as well as the ability of the survey to derive statistically reliable estimates (e.g., see Chapter 2).

Model-based estimates are a second approach to producing LAEs. Sometimes referred to as 'indirect estimates', they are the result of using small area estimation (SAE) models. Model-based efforts are attractive because they produce estimates for local areas, even in cases where the sample size is too small to produce reliable direct estimates. The methodology used to produce such estimates typically combines survey data (direct estimates) for the LEA or from a larger area conglomeration with a type of regression model. The models use predictor variables from external sources (e.g., Census data). Sometimes, samples are combined across years to improve the precision of the estimates. A 95 percent prediction interval is typically associated with each estimate generated from the model, although smaller prediction intervals are frequently used. In Chapter 3, we present three relevant examples used in key government sponsored programs that could be applied to NCVS data.

---

[2] Information on the BCS was obtained through discussion with several senior staff working on the BCS.

Krenzke, Li, and Cantor (2009) (referred to hereafter as the Task 3 memo) provides an illustration of using SAE methodology to predict crime rates for 40 selected Metropolitan Statistical Areas (MSAs), and provides some indications of key relationships between NCVS direct estimates and Uniform Crime Reports data.
Another option to be considered is that of combining data from two surveys while also using a statistical model. An example of this approach is the effort by the National Cancer Institute to combine the national sample from the National Health Interview Survey (NHIS) with the independent state samples from the Behavioral Risk Factors Surveillance Study (BRFSS). Chapter 4 describes options for combining the NCVS with an independent supplemental victimization survey, following several different methodologies.

Prior to discussing the alternative designs, we provide a brief overview of the current NCVS design, and introduce the types of crime, geographic areas and cost assumptions that are discussed in this report.

1.1 Brief Overview of the Current Sample Design

The current NCVS is designed to produce direct estimates at the national level. The NCVS has a stratified, multi-stage cluster design, located in 203 Primary Sampling Units (PSUs). Of the 203 PSUs, 93 are large metro areas that are self-representing (selected with probability equal to one). The sample size in 2003 consisted of about 42,000 households, containing about 74,500 persons aged at least 12 years old, who were interviewed twice over a twelve month period. In 2005, 67,000 people participated and, in 2007, the sample was reduced to about 58,000 persons, because of a 14 percent reduction in sample size.

1.2 Types of Crime and the Characteristics

The NCVS produces national estimates for the population age 12 years old and older. At an aggregate level, the three major types of crime are (crime rate[3] from 2007 in parenthesis): violent (21, i.e., 21 victimizations per thousand population), personal theft (1 per thousand), and property (146 per thousand). The violent crimes are broken out into the general categories of rape/sexual assault (1 per thousand), robbery (2 per thousand), assault (17 per thousand), which is broken out further to

[3] Number of victimizations per 1,000 persons age 12 or older or per 1,000 households over one year

aggravated (3 per thousand) and simple assault (14 per thousand). Personal theft is not subdivided, however, property crimes are split into burglary (27 per thousand), motor vehicle theft (8 per thousand), and theft (111 per thousand). Due to the large sample size, the national design facilitates the production of estimates for several subgroups, including detailed crime types, demographics, geography (urbanicity), type of victim and offender, and time of crime event.

For local area designs, crimes rates can be produced for particular subgroups, such as gender, age and race categories. The discussions in this report will be focused on the person-level violent crimes, because of their high degree of interest, as well as household level property crimes, whose relatively high crimes rates will be more precise than the less prevalent violent crimes.

Also of interest are the characteristics of crime. By 'characteristics' we mean data on the use of force, extent of injury/property loss, relationship to offender, location of the event, reporting to the police and other consequences of the event. In general our focus is on crime rates, however, discussion on crime characteristics occurs in Chapter 4 in the context including sample that is supplemental to the national survey design.

1.3 Local Areas of Interest

Policy makers are likely to be interested in crime rate estimates for different levels of geography: states, crime statistic areas, MSAs, and counties. Crime statistic areas can be formed as a level of geography specifically for producing crime statistics, such as groups of counties in large states, or groups of small states. An example of areas created specifically for subject-specific programs is the set of Health Service Areas[4] originally defined by the National Center for Health Statistics based on hospital care areas and since modified by the National Cancer Institute. Another geographic area of interest is MSAs, which are large metro areas that consist of a single county or group of counties (Office of Management and Budget 2007). Using large MSAs as reporting areas is appealing since traditionally they are in the sample with probability equal to one, and therefore have sample continuously throughout the decade. Another possibility that may be of interest are central cities, defined by the Census Bureau. Another small geographic entity considered is the county. There are 3,141 counties in the U.S. Ultimately, crime rates for each police jurisdiction would be very useful for data users. However, there are over 17,000 such jurisdictions in the U.S. and estimation at that

[4] Health Service Areas are described in http://seer.cancer.gov/seerstat/variables/countyattribs/hsa.html (current as of March 20, 2010).

level is not feasible. Note that estimates were generated for Police Force Areas from the British Crime Survey, but there are only 42 of them (Bolling, Grant and Donovan, 2008).

1.4 Cost Assumptions

For purposes of illustration, this report considers situations where the survey is able to expend an additional 10 million dollars to add sample. The intent is to provide some idea of how a particular methodology may improve LAEs if additional resources were available. For example, in Chapter 2, on direct estimation, situations are simulated where this money is used to add additional NCVS interviews for a particular year.

In order to translate this money into actual interviews, it was assumed that it would cost approximately $800 to complete a full NCVS interview. This translates to roughly 13,000 completed interviews. This number is based on the assumption that the interview would be done on a one-time basis and includes the costs of putting the data collection systems in place, training the interviewers and collecting the information. It assumes a contractor, outside the Census Bureau, to conduct a national, in-person effort, like the NCVS. It applies a rough discount for economies related to interviewing more than one person in the household. It should be noted that these costs are not based on any set of detailed cost assumptions. They should only be used to provide a sense of how LAEs might be affected once adding sample under different assumptions.

These costs are not comparable to a large ongoing survey effort that has interviewing operations all year round and use staff on a continuous basis. For example, the Census Bureau cites a figure closer to $150 per sampled case for the ACS (Hughes and Griffin, 2010). It is not clear how these costs compare to the current NCVS costs or how they might compare for another organization to complete interviews under a similar model as the Census Bureau currently collects the NCVS.

Similarly, it was assumed that it would cost approximately $75 per completed mail survey. This cost is based on the assumption that it is being done as part of a one-time effort by an organization that would need some resources to develop systems to put the effort in place. As one point of comparison, Link et al (2008) cites $70/complete for the costs in reference to pilot work for the BRFSS. Another point of comparison is the ACS, which comes to around $25 per completed mail survey (Hughes and Griffin, 2010).

Direct Estimation 2

This chapter discusses possible ways of achieving LAEs of various crime rates from the NCVS using direct estimates based on survey data only from the sample units in the local area. Prior to discussing possible sample designs, section 2.1 provides a brief introduction to some general concepts relating to the measures of precision that will be used in the rest of the chapter, as well as an overview of the levels of precision of crime rates at the national level under the current NCVS design. This is followed by a short discussion on the geographic levels for which LAEs may be of interest and sample sizes that would be required for these various local areas to produce crime rate estimates of adequate precision. Section 2.2 investigates the extent to which LAEs are possible under a sample design that is proportionately allocated across the country, as is the case with the current design. Scenarios under both the current and an increased sample size are considered for estimates at the state and MSA levels. Multi-year estimates are also discussed as a possibility. Finally, in section 2.3, we consider an option that uses proportionate allocation for the core sample with a disproportionately distributed supplemental sample, to obtain estimates at the state and MSA levels. For all options considered, the impact on national estimates relative to the current design is explored.

2.1 Background

2.1.1 Measures of Precision

This section gives an overview of the measures of precision used for evaluating the utility of LAEs:

- *The standard error (SE).* The SE is a basic measure of the sampling error. Under simple random sampling (SRS), for an estimate of a crime rate p (measured per thousand) the SE is computed as $SE(p) = \sqrt{p(1,000 - p)/n}$, where n is the sample size. The SE formula given above is for SRS but in practice large-scale national surveys, such as NCVS, typically employ complex sample designs. The SE formula for SRS needs to be multiplied by the square root of the design effect ($DEFF$) to give the SE under the complex design. The $DEFF$ is a useful quantity to examine when comparing alternative designs. Since the NCVS design is comprised of a multi-stage cluster sample, the SE is larger than under a SRS . One typical interpretation of $DEFF$ says that if $DEFF = 2$, the sample size needs to be doubled in order to achieve the same precision as from a SRS. In general, the overall $DEFF$ is sometimes approximately expressed as the product of two components: $DEFF_{CLU}$, which is due to clustering and the $DEFF_{UEW}$, which is due to differential sampling rates (or unequal weighting). That is, $DEFF = DEFF_{CLU} * DEFF_{UEW}$.

In a four-stage design such as that of NCVS (see Chapter 1 for a description), the $DEFF$ due to clustering may be expressed approximately as:

$$DEFF_{CLU} = 1 + (\bar{b}_1 - 1)\rho_1 + (\bar{b}_2 - 1)\rho_2 + (\bar{b}_3 - 1)\rho_3$$

where $\bar{b}_1$ is the average number of sampled persons per PSU, and ρ_1 is the intracluster correlation that measures the homogeneity of the characteristic being measured for persons within the PSUs. Similarly, $\bar{b}_2$ is the average number of sampled persons per segment, and ρ_2 is the intracluster correlation for persons within segments. Finally, $\bar{b}_3$ is the average number of sampled persons per household, and ρ_3 is the intracluster correlation for persons within households. See Hanson, Hurwitz and Madow (1953, volume 1, section 17) for more details.

The $DEFF$ due to differential sampling rates by stratum, as given in Kish (1965), can be expressed as:

$$DEFF_{UEW} = (\Sigma W_B / k_B)(\Sigma W_B k_B)$$

where $W_B = N_B / N$, N is the total population size, N_B is population size for stratum B and k_B is the sampling rate within stratum B. The $DEFF_{UEW} = 1$ (approximately) for the current design since a constant sampling rate is used given the proportional allocation across the nation.

- *The margin of error (MOE).* The MOE is a multiple of the SE that represents the half-width of a confidence interval. For a 95 percent confidence interval, one can be 95 percent confident that the true crime rate is within the interval defined by $p \pm MOE$. The MOE is generally computed as $MOE = 1.96 \times SE(p)$, where 1.96 is the value taken from the normal distribution to give 95 percent coverage for the interval. However, when the SE estimate is based on few degrees of freedom, the normal distribution value should be replaced by a corresponding value for the t distribution with the given degrees of freedom. With a multistage sample design, the number of degrees of freedom for the estimate of the standard error of a rate depends on the number of sampled PSUs, which may be small for many local areas. For example, with eight degrees of freedom, the 95 percent t value is 2.31 in place of 1.96. It should also be noted that the confidence interval computed as $p \pm MOE$ is based on the approximation that the sampling distribution of p is a normal distribution. The validity of that approximation depends on the sample size n and the value of p. Cochran (1977, p.58) suggests that for an SRS a sample size of 30 is adequate if $p = 500$ per thousand but much larger sample sizes are needed if p is small, as is the case with many of the crime rates of interest. For example, the sample sizes needed for the approximation to hold reasonably well are 200 for $p = 200$, 600 for $p = 100$, and 1,400 for $p = 50$. When the approximation is inadequate, asymmetrical confidence intervals are needed.

- *The coefficient of variation* (CV). The CV is the SE relative to the estimated crime rate. The CV is commonly presented as a percentage, that is, $CV = 100 * SE(p) / p$. In general, a CV of 5 percent is considered to indicate a high level of precision for the estimate, with 10 percent being widely viewed as acceptable. In contrast, a CV of 30 percent is generally considered imprecise. However, for low rates, an assessment of the MOE may give the better guidance than the CV on what is an acceptable level of precision.

To illustrate the relationship between given values of estimated rates (p) and CVs, MOEs, and the corresponding 95 percent confidence intervals, Table 1 assumes a random sample of $n = 2{,}000$ cases with a complex design having an overall $DEFF = 2$. The CV of 22 percent associated with a rate of 20 per thousand may seem inadequate. However, in terms of the MOE of nine and the confidence interval ranging from 11 to 29 per thousand, the precision may be considered adequate.

Table 1. Rates and their *CV*s, *MOE*s and confidence intervals for a random sample of size 2,000 with *DEFF* = 2

Estimate of rate (per 1000)	*CV* (percent)	*MOE*	95 % Confidence interval	
			Lower bound	Upper bound
5	45	4	1	9
10	31	6	4	16
20	22	9	11	29
50	14	14	36	64
100	9	19	81	119
200	6	25	175	225

2.1.2 Precision of National Estimates

Before discussing the precision and required sample sizes for LAEs, it is instructive to look at crime estimates under the current NCVS design to see what levels of precision have been achieved at the national level. A brief overview of the current survey design was given in Chapter 1. An important feature of the design is that once a household is sampled, everyone within the household 12 years of age or older is interviewed. Information on crime is collected at both the person level and the household level. The sample size in 2003 consisted of about 74,500 persons aged at least 12 years old interviewed (approximately twice that many interviews were conducted since interviews occur every six months) from about 42,000 households. Results from the 2003 NCVS are displayed in Table 2A for person-based crimes and in Table 2B for household-based crimes. The NCVS crime rate estimates are those reported by the Bureau of Justice Statistics (2004) whereas the SEs were computed using published generalized variance function (GVF) parameters for major crime types. See Wolter (1985, chapter 5) for a discussion of GVFs. Table 2A shows that the CVs for person-based crimes ranged from 4 percent for violent crimes, assault and simple assault to 14 percent for rape/sexual assault and personal theft. While the CVs for rape/sexual assault and personal theft are relatively high, the MOE may be deemed acceptable. That is, to say with 95 percent confidence that the true rate is within the interval 0.8 ± 0.2, i.e., from 0.6 to 1.0 per thousand, may be acceptable. Note that, while in 2003, 74,500 persons participated in NCVS, that number dropped to 67,000 in 2005, and to 58,000 in 2007. Therefore, the precision of estimates in the years since 2003 is lower due to reduced sample sizes as well as lower response rates.

The final column of Tables 2A and 2B gives estimates of $DEFF$ for the various crime rates. These $DEFF$s were computed by dividing the GVF variance for each crime rate by the corresponding variance under a SRS. The $DEFF$s for national estimates computed in this manner range across the two tables from 1.2 for rape/sexual assault and personal theft to 3.0 for property crimes.

Table 2A. National crime rates and measures of precision by crime type for persons: 2003

Crime type	Estimate of crime rate (per thousand)	*SE*	*CV* (percent)	*MOE*	*DEFF*
Violent crimes	22.6	0.8	3.6	1.6	2.3
Rape/sexual assault	0.8	0.1	13.8	0.2	1.2
Robbery	2.5	0.2	8.8	0.4	1.4
Assault	19.3	0.7	3.8	1.5	2.2
Aggravated	4.6	0.3	6.7	0.6	1.6
Simple	14.6	0.6	4.3	1.2	2.0
Personal theft	0.8	0.1	13.8	0.2	1.2

Source: The crime rate estimates are taken from Bureau of Justice Statistics (2004). *SE*s were computed from the generalized variance function parameters.

Table 2B. National crime rates and measures of precision by crime type for households: 2003

Crime type	Estimate of crime rate (per thousand)	*SE*	*CV* (percent)	*MOE*	*DEFF*
Property crimes	163.2	3.1	1.9	6.1	3.0
Household burglary	29.8	1.1	3.8	2.2	1.8
Motor vehicle theft	9.0	0.5	5.9	1.0	1.3
Theft	124.4	2.7	2.1	5.2	2.8

Source: The crime rate estimates are taken from Bureau of Justice Statistics (2004). *SE*s were computed from the generalized variance function parameters.

2.1.3 Geographic Areas of Interest for LAEs

If a goal is to make minimal changes to the national NCVS design, then the most potential would be to construct LAEs comprised of either: large states and groups of smaller states, or large MSAs. This chapter will concentrate on investigating possibilities with regards to these two levels of LAEs. Most, but not all, states have at least some NCVS sampled cases, and large MSAs have NCVS sample on a continuous basis, under the current design.

2.1.4 General Sample Size Requirements for LAEs

Tables 3A and 3B provide approximate sample sizes that are needed to achieve specified level of precision by crime type for any subnational estimate. Note that in this context, "sample size" represents "number of persons" when estimating person-based crimes such as violent crimes and personal theft (Table 3A), but represents "number of households" when estimating household-based crimes such as property crimes (Table 3B). These tables provide the overall sample sizes (n) necessary to achieve CVs of 10, 15, and 20 percent and the corresponding MOEs for those sample sizes. The sample size calculations in the tables assume the $DEFF$s in Table 2. Note that the calculations in Tables 3A and 3B assume that the crime rates a local area are the same as those for the nation.

Table 3A. Sample sizes and *MOE* for selected *CV*s for person-based crimes

Crime Type	Crime rate estimate (per thousand)	*CV* = 0.1		*CV* = 0.15		*CV* = 0.2	
		n	*MOE*	*n*	*MOE*	*n*	*MOE*
Violent crimes	22.6	9,774	2.9	4,344	4.4	2,444	5.9
Rape/sexual assault	0.8	151,129	0.1	67,168	0.2	37,782	0.3
Robbery	2.5	56,658	0.4	25,181	0.6	14,165	0.8
Assault	19.3	11,027	2.6	4,901	3.9	2,757	5.1
Aggravated	4.6	34,190	0.7	15,195	1.1	8,547	1.4
Simple	14.6	13,701	2.0	6,089	3.0	3,425	4.0
Personal theft	0.8	151,129	0.1	67,168	0.2	37,782	0.3

Table 3B. Sample sizes and *MOE* for selected *CV*s for household-based crimes

Crime type	Crime rate estimate (per thousand)	*CV* = 0.1		*CV* = 0.15		*CV* = 0.2	
		n	*MOE*	*n*	*MOE*	*n*	*MOE*
Property crimes	163.2	1,528	18.5	679	27.8	382	37.1
Household burglary	29.8	5,925	4.3	2,634	6.5	1,481	8.7
Motor vehicle theft	9.0	14,535	1.5	6,460	2.3	3,634	3.1
Theft	124.4	1,943	14.7	863	22.0	486	29.4

The focus of the following discussion will also be limited to estimates of the two major categories of crimes: violent crimes, and property crimes. These two major categories serve as an illustration of the kind of estimates of crime rates that are possible at local area levels for both person-based and household-based crimes. These two broad categories also illustrate the two ends of the spectrum in terms of low and high rates for such crimes. The third major category, personal theft, has sample size requirements (see Table 3A) which make it untenable for LAEs with reasonable precision. The estimation of household burglary has more or less the same stringency of requirements as that of violent crimes. The Task 3 Memo discusses burglary if the reader is interested.

Table 3A shows a violent crime rate in 2003 of 22.6 per thousand, which is fairly small. Under the assumptions provided above, a sample size of about 2,444 participating persons is needed to achieve an MOE of about 5.9 (corresponding to a CV of 20 percent). Thus, a 95 percent confidence interval would range from 16.7 to 28.5, which may be acceptable. If a tighter confidence interval is desired, a sample size of about 4,344 participating persons is needed to achieve an MOE of about 4.4 with a corresponding 95 percent confidence interval ranging from 18.2 to 27.0 (corresponding to a CV of 15 percent). Finally, even greater precision can be attained by using a sample size of about 9,774 participating persons to achieve an MOE of about 2.9 with a corresponding 95 percent confidence interval ranging from 19.7 to 25.5 (corresponding to a CV of 10 percent). The ultimate choice of sample size that is utilized for cross-sectional LAEs should be a balanced consideration between available budget and level of precision that is deemed acceptable to the readers. In the tables that display the various allocation options throughout this chapter, all three sample sizes will be presented to illustrate the range of options available to the users. The focus of the discussion will be limited to MOEs rather than CVs due to the small rates involved in the analysis.

In addition to cross-sectional estimates of crime rates, estimates of change in crime rates over time are also of interest to policy makers. Thus, another consideration that can drive the choice of precision level deemed acceptable is the ability of the sample size and MOE to be able to detect significant differences in change in crime rates over time. To investigate this issue for violent crimes, we assume independence between violent crime rates p_1 and p_2 associated with any two time points, although this is not likely the case for time points spaced by three years or less given the overlap in the sample. Assuming the power to be 80 percent and the type one error to be 0.05, a two-sided test of hypothesis for the difference between the true violent crime rates at the two time points being zero can detect a difference of 13 per thousand based on a sample size of 2,444 persons, 9 per thousand based on a sample size of 4,344 persons, or six per thousand based on a sample size of 9,774 persons.

The Crime Victimization Bulletins (Bureau of Justice Statistics, 2002 and 2009) show that the violent crime rate has gone down from around 50 per thousand in 1994 to around 19 per thousand in 2008, a difference of 31 per thousand. Furthermore, the largest year-to-year change during this period was around 6 per thousand. Therefore, a test of hypothesis for the difference between year-to-year violent crime rates can only detect the largest year-to-year violent crime rate difference of 6 as being significant based on a sample size of 9,774, and cannot detect the differences for any of the other smaller year-to-year changes based on any of the sample sizes given above. However, a test of hypothesis can detect the 14-year difference in violent crime of 31 per thousand between the years 1994 and 2008, based on any of the three sample sizes (2,444 persons, 4,344 persons or 9,774 persons). Therefore, any of these sample sizes with their corresponding MOEs can be used in general to detect significant long term trends in violent crime rates at the local level.

Turning to property crimes, according to Table 3B, a sample size of about 382 participating households could achieve an MOE of about 37.1 on a property crime rate of 163.2 (corresponding to a CV of 20 percent). Thus, a 95 percent confidence interval would range from 126.1 to 200.3. If a tighter confidence interval is desired, a sample size of about 679 participating households is needed to achieve an MOE of about 27.8 with a corresponding 95 percent confidence interval ranging from 135.4 to 191.0 (corresponding to a CV of 15 percent). Finally, even greater precision can be attained by using a sample size of about 1,528 participating households to achieve an MOE of about 18.5 with a corresponding 95 percent confidence interval ranging from 144.7 to 181.7 (corresponding to a CV of 10 percent).

We use the same assumptions as before for testing differences in change in rates for property crimes over time. Assuming the power to be 80 percent and the type one error to be 0.05, a two-sided test of hypothesis for the difference between the true property crime rates at the two time points being zero can detect a difference of 82 per thousand based on a sample size of 382 households, 60 per thousand based on a sample size of 679 households, or 39 per thousand based on a sample size of 1,528 households.

The Crime Victimization Bulletins (Bureau of Justice Statistics, 2002 and 2009) show that the property crime rate has dropped from 318 per thousand in 1993 to 135 per thousand in 2008, a difference of 183. The largest year-to-year change during this period was roughly 20 per thousand. Thus, a test of hypothesis based on any of the three sample sizes above (382 households, 679 households or 1,528 households) cannot detect the largest single year difference of 20, but can detect the long term trend difference of 183. Therefore, any of these sample sizes with their corresponding MOEs can be used to detect significant long term trends in property crime rates at the local level.

2.2 LAEs Under Designs with Proportionate Allocation

In the discussions that follow (and throughout the chapter), we illustrate the options by considering estimates at the level of large states and large MSAs. In later sections of the chapter, we also consider estimates for a combination of large states and remainders of Census Divisions within which the large states are located. Publishable direct estimation for hundreds of substate regions or for each of the 3,141 counties would require such a substantial redesign and an increase in sample size and cost that it is not explicitly treated here. However, the sample sizes needed to produce estimates at the levels of substates and counties are identical to those discussed below for states and MSAs.

2.2.1 Estimates for States Assuming Proportionate Allocation and the Current Sample Size

In 2005, 67,000 persons from 38,600 households participated in the NCVS, with roughly 1.74 participants aged 12 or more per household. The 2007 design of the NCVS resulted in about 58,000 persons participating, and we assume the same ratio of 1.74 to arrive at an estimate of roughly 33,415 participating households nationally. The sample was distributed proportionately across the country to the U.S. population. As a result, larger states had larger sample sizes compared to smaller states. Some states may not have had any sampled PSUs and therefore no sample whatsoever. A problem for the estimation of state crime rates is that strata under the current NCVS design cut across state boundaries in some cases. Nevertheless, as a preliminary exercise, it is useful to consider options involving minimal change to the current design (and at minimal additional cost), to see which, if any, estimates have acceptable levels of precision. Note that even under this minimum change approach, stratification would need to be reworked to respect state boundaries (for large states at least).

An important related issue concerns the stability of variance estimates (square of the SEs) corresponding to crime rate estimates. Standard variance formulae for estimates of crime rates under complex designs are related to the number of PSUs in the local area for which estimates are desired. As the number of PSUs approaches zero, the variance becomes extremely large (approaches infinity). Therefore, a reasonable number of PSUs should be allocated to the level at which estimates and their variances are to be computed to avoid such large variances. Other surveys (such as National Adult Assessment of Literacy) that have supplemental sample at the state level have typically used eight to 12 PSUs per state as a rough guideline to ensure stability.

For the discussion below, some assumptions are made regarding how the 203 PSUs are distributed under the current NCVS design. Since the sample is distributed roughly proportionately across the country, the assumption is made that the PSUs are as well. However, some states likely have no PSUs whatsoever, and states containing certainty PSUs (often highly populated MSAs) will have proportionately more PSUs (where disproportionately more PSUs are allocated to those certainty PSUs). Regardless, an approximation that distributes the 203 PSUs across states shows that, on average, each state likely has roughly four PSUs and that only eight to 10 of the largest states will have a sufficient number of PSUs for variance estimation under the current design (i.e., eight or more PSUs per state). Therefore, under a design of proportionate allocation, some of the mid-sized states for which LAEs may be possible would need to be allocated a reasonable number of PSUs in order to ensure stable variances – particularly those states lacking certainty PSUs. Thus, even under a scenario of no alteration to the proportionate allocation, the resultant design would not be cost neutral since an increase in the number of PSUs would also imply increased costs due to the necessity of hiring additional interviewers.

The sample sizes in Tables 3A and 3B will be used to determine the minimum sample sizes required to ensure the desired levels of precision as reflected through the MOEs . To be more precise there would need to be an adjustment to these numbers to reflect the $DEFF$ due to clustering at the state level: the contribution from PSUs, segments and households. Since the number and types of PSUs (e.g., certainty vs. non-certainty) across states will naturally differ, and there is no ready source that can be used to get this information, there is no simple way to make this adjustment. For purposes of discussion, this approximation should be adequate to provide an idea of the samples sizes required to achieve particular levels of precision. Specifically the three benchmark minimum sample sizes to be used are 2,444 persons (corresponding to an MOE of 5.9), 4,344 persons (corresponding to an MOE of 4.4) and 9,774 persons (corresponding to an MOE of 2.9) as gauges for determining adequate levels of precision for estimating violent crimes at the state level. Similarly, the three benchmark minimum sample sizes to be used are 382 households (corresponding to an MOE of 37.1), 679 households (corresponding to an MOE of 27.8) and 1,528 households (corresponding to an MOE of 18.5) as gauges for determining adequate levels of precision for estimating property crimes at the state level.

Based on the 2007 NCVS sample size of 58,000 persons and 33,415 households, Table 4A columns B and C give the expected sample sizes both for persons and households per state, with the sample allocated proportionately.

The relevant states are highlighted with grey shading – the light shading corresponds to the states that meet the most stringent sample size requirements (9,774 persons or 1,528 households), the medium shading corresponds to the additional states that meet the minimum sample size requirements of 4,344 persons or 679 households, and the dark shading corresponds to the additional states that meet the least stringent sample size requirements of 2,444 persons or 382 households.

From Table 4A column B, it is apparent that the five largest states have a sufficient number of persons sampled for estimating violent crimes with an MOE of 5.9 and the two largest states have a sufficient number of persons sampled for estimating violent crimes with an MOE of 4.4. Note that no states have a sufficient number of persons sampled for estimating violent crimes with an MOE of 2.9. Similarly, from Table 4A column C, it is apparent that the 29 largest states have a sufficient number of households sampled for estimating property crimes with an MOE of 37.1, the 16 largest states have a sufficient number of households sampled for estimating property crimes with an MOE of 27.8, and the four largest states have a sufficient number of households sampled for estimating property crimes with an MOE of 18.5.

2.2.2 Estimates for States Assuming Proportionate Allocation and an Increased Sample Size

It is clear from the previous section that the current sample sizes are adequate for estimates of violent crimes for only a handful of states. It is apparent that a design which allocates sample proportionately to population size across the country to enable optimal national estimates is not optimal for producing LAEs of adequate precision. A preferable design would be one that allocates sample to local areas disproportionately to population size. Designs with disproportionate allocations have an impact on national estimates, although minimally given the large sample size for estimates at that level. However, even under the NCVS design that uses proportionate allocation, given the steadily declining sample size from 74,500 persons in 2003 to 67,000 persons in 2007 and 58,000 persons in 2007, some estimates at the national level have been eroded with regards to precision - for example, the ability to detect significant differences in estimates of year-to-year change of crime rates has deteriorated to a level where it has become necessary to combine multiple years of data in order to be able to make inference on change over time (See National Research Council of the National Academies, 2008). Therefore, it is of interest to consider an option of increasing the sample size under a design with proportionate allocation with the main aim of improving national estimates and to see whether LAEs might be facilitated to some degree under this scenario.

Assuming an increased budget of 10 million dollars (see Chapter 1 for a discussion on cost assumptions), it is estimated that an additional 13,000 persons (7,490 households) could be interviewed, for a total sample size of 71,000 persons and 40,905 households.

Table 4A, columns D and E present the proportionate distribution of the additional persons and households across states assuming an increased sample size. In a comparison analogous to the one above for columns B, from column D, it is apparent that the seven largest states have a sufficient number of persons sampled for estimating violent crimes with an MOE of 5.9, the three largest states have a sufficient number of persons sampled for estimating violent crimes with an MOE of 4.4, and no states have a sufficient number of persons sampled for estimating violent crimes with an MOE of 2.9. Similarly, from Table 4A column E, it is apparent that the 31 largest states have a sufficient number of households sampled for estimating property crimes with an MOE of 37.1, the 21 largest states have a sufficient number of households sampled for estimating property crimes with an MOE of 27.8, and the seven largest states have a sufficient number of households sampled for estimating property crimes with an MOE of 18.5. In most cases, columns D and E show an increase of only two to five states over the scenario prior to the increase in sample size represented by columns B and C. Therefore, allocating supplemental sample in this manner is not the most effective way of ensuring the greatest number of states with publishable estimates for either violent crimes or property crimes.

Note that a sample increase of 13,000 persons (7,490 households) would necessitate an additional 45 PSUs (for a total of roughly 250 PSUs), assuming a proportionate increase from the existing 203. It is assumed that these additional PSUs would be proportionately allocated to the supplemental sample.

Table 4A. Large states: Proportionate allocation with and without increased sample

State	A: Percentage of the U.S. population *	B:Estimated sample size of individuals - Current sample size	C: Estimated sample size of households - Current sample size	D:Estimated sample size of individuals - Increased sample	E: Estimated sample size of households - Increased sample
California	12.1	7,018	4,043	8,591	4,949
Texas	7.9	4,582	2,640	5,609	3,231
New York	6.4	3,712	2,139	4,544	2,618
Florida	6.1	3,538	2,038	4,331	2,495
Illinois	4.3	2,494	1,437	3,053	1,759
Pennsylvania	4.1	2,378	1,370	2,911	1,677
Ohio	3.8	2,204	1,270	2,698	1,554
Michigan	3.3	1,914	1,103	2,343	1,350
Georgia	3.2	1,856	1,069	2,272	1,309
North Carolina	3.0	1,740	1,002	2,130	1,227
New Jersey	2.9	1,682	969	2,059	1,186
Virginia	2.6	1,508	869	1,846	1,064
Washington	2.1	1,218	702	1,491	859
Massachusetts	2.1	1,218	702	1,491	859
Indiana	2.1	1,218	702	1,491	859
Arizona	2.1	1,218	702	1,491	859
Tennessee	2.0	1,160	668	1,420	818
Missouri	1.9	1,102	635	1,349	777
Maryland	1.9	1,102	635	1,349	777
Wisconsin	1.9	1,102	635	1,349	777
Minnesota	1.7	986	568	1,207	695
Colorado	1.6	928	535	1,136	654
Alabama	1.5	870	501	1,065	614
South Carolina	1.5	870	501	1,065	614
Louisiana	1.4	812	468	994	573
Kentucky	1.4	812	468	994	573
Oregon	1.2	696	401	852	491
Oklahoma	1.2	696	401	852	491
Connecticut	1.2	696	401	852	491
Iowa	1.0	580	334	710	409
Mississippi	1.0	580	334	710	409
Remainder	9.5	5,510	3,174	6,745	3,886
Total	100.0	58000	33415	71000	40905

* Source: American Community Survey Public Use Microdata Sample 2007.

2.2.3 Estimates for MSAs Assuming Proportionate Allocation and the Current Sample Size

An analysis similar to that of Table 4A for states is presented in Table 4B for the large MSAs. Using NCVS files with data for the 40 largest MSAs for 2003 (Bureau of Justice Statistics 2007), an average $DEFF$ due to clustering at the MSAs level was calculated as 2.11 for violent crime and 1.98 for property crime. We assume roughly the same averages in 2007 for MSAs. This compares with $DEFF$s at the national level of 2.26 for violent crimes and 2.98 for property crimes (see Tables 2A and 2B). The national $DEFF$s are larger because of the additional clustering due to PSUs at the national level, not present at the MSA level. Thus, there is a 3 percent decrease in the SE, and therefore the MOE, for violent crime at the MSA level as compared to the national level. Similarly, there is roughly an 18 percent decrease in the SE, and therefore the MOE, for property crime at the MSA level as compared to the national level.

To achieve the same level of precision for violent crime rates as that for the nation, the sample size requirements of 2,444 persons from Table 3A (corresponding to an MOE of 5.9), 4,344 persons (corresponding to an MOE of 4.4) and 9,774 persons (corresponding to an MOE of 2.9) diminish by a factor of $2.11/2.26$ to 2,282 persons, 4,055 persons and 9125 persons, respectively. The relevant MSAs are highlighted with grey shading – the light shading corresponds to the MSAs that meet the most stringent sample size requirements (9,125 persons), the medium shading corresponds to the additional states that meet the minimum sample size requirements of 4,055 persons, and the dark shading corresponds to the additional states that meet the least stringent sample size requirements of 2,282 persons. From Table 4B, column B, it is apparent that only the two largest MSAs (New York and Los Angeles) have a sufficient number of persons sampled for estimating violent crimes with an MOE of 5.9. No states have a sufficient number of persons sampled for estimating violent crimes with an MOE of 4.4 or 2.9. Table 4B provides the sample sizes for MSAs.

Similarly, to achieve the same level of precision for property crime rates as that for states, the sample size requirements of 382 households from Table 3B (corresponding to an MOE of 37.1), 679 households (corresponding to an MOE of 27.8) and 1,528 households (corresponding to an MOE of 18.5) diminish by a factor of $1.98/2.98$ to 254 households, 451 households, and 1,015 households, respectively.

From Table 4B, column C, it is apparent that the three largest MSAs (New York, Los Angeles and Chicago) have a sufficient number of households sampled for estimating property crimes with an

MOE of 18.5, the 14 largest MSAs have a sufficient number of households sampled for estimating property crimes with an MOE of 27.8, and the 22 largest MSAs have a sufficient number of households sampled for estimating property crimes with an MOE of 37.1.

2.2.4 Estimates for MSAs Assuming Proportionate Allocation and an Increased Sample Size

An alternative to the allocation of supplemental sample to states is the allocation to MSAs instead, to facilitate the production of LAEs at that level. Table 4B columns D and E present such an allocation to persons and households. Column D shows that there is almost no improvement compared to the situation with no sample increase except that now one MSA (New York) can produce estimates of violent crime with an MOE of 4.4, whereas prior to the sample increase, none could. In column E, 29 MSAs have sufficient households sampled to estimate property crimes with an MOE of 37.1, an increase of seven MSAs over the scenario with no increased sample. Apart from this, the results are identical to those in columns B and C. Therefore, as was the case with states, allocating supplemental sample proportionately to MSAs is not the most effective way of ensuring the greatest number of MSAs with publishable estimates for either violent crimes or property crimes.

Table 4B. Large MSAs: Proportionate allocation with and without increased sample

MSA	A: Percentage of the U.S. population **	B: Estimated sample size of individuals - Current sample size	C: Estimated sample size of households - Current sample size	D: Estimated sample size of individuals - Increased sample	E: Estimated sample size of households - Increased sample
New York-Northern New Jersey-Long Island, NY-NJ-PA	6.2	3,622	2,087	4,434	2,554
Los Angeles-Long Beach-Santa Ana, CA	4.2	2,453	1,413	3,003	1,730
Chicago-Naperville-Joliet, IL-IN-WI	3.1	1,824	1,051	2,232	1,286
Dallas-Fort Worth-Arlington, TX	2.1	1,200	692	1,470	847

Table 4B. Large MSAs: Proportionate allocation with and without increased sample (Continued)

MSA	A: Percentage of the U.S. population **	B: Estimated sample size of individuals - Current sample size	C: Estimated sample size of households - Current sample size	D: Estimated sample size of individuals - Increased sample	E: Estimated sample size of households - Increased sample
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	1.9	1,113	641	1,362	785
Houston-Sugar Land-Baytown, TX	1.9	1,092	629	1,336	770
Miami-Fort Lauderdale-Pompano Beach, FL	1.8	1,032	594	1,263	728
Atlanta-Sandy Springs-Marietta, GA	1.8	1,024	590	1,254	723
Washington-Arlington-Alexandria, DC-VA-MD-WV	1.8	1,021	588	1,250	720
Boston-Cambridge-Quincy, MA-NH	1.5	862	497	1,055	608
Detroit-Warren-Livonia, MI	1.5	843	486	1,032	595
Phoenix-Mesa-Scottsdale, AZ	1.4	816	470	999	575
San Francisco-Oakland-Fremont, CA	1.4	815	469	997	574
Riverside-San Bernardino-Ontario, CA	1.4	784	452	960	553
Seattle-Tacoma-Bellevue, WA	1.1	637	367	780	450
Minneapolis-St. Paul-Bloomington, MN-WI	1.1	615	355	753	434
San Diego-Carlsbad-San Marcos, CA	1.0	572	329	700	403
St. Louis, MO-IL	0.9	537	309	657	379
Tampa-St. Petersburg-Clearwater, FL	0.9	521	300	638	367
Baltimore-Towson, MD	0.9	508	293	622	358
Denver-Aurora, CO	0.8	478	275	585	337
Pittsburgh, PA	0.8	448	258	548	316
Portland-Vancouver-Beaverton, OR-WA	0.7	421	242	515	297

Table 4B. Large MSAs: Proportionate allocation with and without increased sample (Continued)

MSA	A: Percentage of the U.S. population **	B: Estimated sample size of individuals - Current sample size	C: Estimated sample size of households - Current sample size	D: Estimated sample size of individuals - Increased sample	E: Estimated sample size of households - Increased sample
Cincinnati-Middletown, OH-KY-IN	0.7	411	237	503	290
Sacramento–Arden-Arcade–Roseville, CA	0.7	402	232	492	284
Cleveland-Elyria-Mentor, OH	0.7	398	229	487	281
Orlando-Kissimmee, FL	0.7	392	226	479	276
San Antonio, TX	0.7	387	223	474	273
Kansas City, MO-KS	0.7	381	220	467	269
Remainder of MSAs	39.4	22,829	13,152	27,946	16,100
nonMSAs	16.5	9,563	5,509	11,706	6,744
TOTAL	100.0	58,000	33,415	71,000	40,905

** Source: U.S. Census Bureau Population Estimates, 2008

2.2.5 Multi-year Estimates for States and MSAs under a Proportionate Design

Another way of increasing the sample size is to combine sample across years. The Bureau of Justice Statistics (BJS) has recently moved to combining two years of data at the national level in order to increase the number of detectable differences in crime rates over time. Doing so has resulted in the need to measure change between successive pairs of years instead of successive single years.

To increase the precision of the estimates at the local area level, instead of increasing the sample size directly as in the previous section, one option would be to combine three years of NCVS sample. When combining three years of independent samples, we would expect the SEs to be roughly 57 percent of the standard errors for a single year. However, since persons within households are in sample for seven times in a three and a half year period, the samples for three consecutive years are not independent. Since there is a 'clustering' of responses for each person, we would expect the gain in precision to be a little lower than 57 percent, depending upon the magnitude of the correlation between responses for an individual over time. The correlation between responses by an individual will likely also vary by crime type. Using NCVS data for MSAs from 2003, it is estimated that the SEs of three-year estimates to be about 63 percent of that of a single year SE for violent crimes and property crimes. The same results are assumed to hold for states in 2007. Thus, the correlated responses have a minimal impact on the precision of three-year estimates and are ignored for the purposes of this exercise. Therefore, it is assumed that the same sample size requirements as in earlier sections hold for LAEs.

Note, however, in considering the possibility of three-year estimates, estimates of change should be constructed for non-overlapping periods of time in order to minimize the correlated response, such as those given by three-year-to-three-year estimates of change.

Estimates for States

Table 5A, columns D and E present a scenario where the estimates for states are combined over a three-year period, effectively tripling the one-year sample size. Comparing to the one-year estimates given in columns B and C (the same as those in Table 4A), one can see that estimates at the state level have been greatly facilitated, using the benchmarks of acceptable precision discussed earlier (9,774 persons for an MOE of 2.9, 4,344 persons for an MOE of 4.4 and 2,444 persons for an MOE of 5.9). From column D, four states, 12 states and 24 states have a sufficient number of persons sampled for estimating violent crimes with MOE s of 2.9, 4.4 and 5.9 respectively. This is a five to six-fold increase in the number of states that have adequate precision under the proportionate allocation design in column B. Using the benchmarks of acceptable precision discussed earlier (1,528 households for an MOE of 18.5, 679 households for an MOE of 27.8 and 382 households for an MOE of 37.1), we can see from column E that 22 states, 36 states and 43 states have a sufficient number of households sampled for estimating property crimes with MOEs of 18.5, 27.8 and 37.1, respectively. Once again, this is a substantial increase in the number of states that have adequate precision in comparison with the proportionate allocation design in column C.

Table 5A. Large states: Proportionate allocation with three-year estimates

State	A: Percentage of the U.S. population	B: Estimated sample size of individuals - Current sample size	C: Estimated sample size of households - Current sample size	D: Estimated sample size of individuals - Three year combined	E: Estimated sample size of households - Three year combined
California	12.1	7,018	4,043	21,054	12,130
Texas	7.9	4,582	2,640	13,746	7,919
New York	6.4	3,712	2,139	11,136	6,416
Florida	6.1	3,538	2,038	10,614	6,115

Table 5A. Large states: Proportionate allocation with three-year estimates (Continued)

State	A: Percentage of the U.S. population	B: Estimated sample size of individuals - Current sample size	C: Estimated sample size of households - Current sample size	D: Estimated sample size of individuals - Three year combined	E: Estimated sample size of households - Three year combined
Illinois	4.3	2,494	1,437	7,482	4,311
Pennsylvania	4.1	2,378	1,370	7,134	4,110
Ohio	3.8	2,204	1,270	6,612	3,809
Michigan	3.3	1,914	1,103	5,742	3,308
Georgia	3.2	1,856	1,069	5,568	3,208
North Carolina	3.0	1,740	1,002	5,220	3,007
New Jersey	2.9	1,682	969	5,046	2,907
Virginia	2.6	1,508	869	4,524	2,606
Washington	2.1	1,218	702	3,654	2,105
Massachusetts	2.1	1,218	702	3,654	2,105
Indiana	2.1	1,218	702	3,654	2,105
Arizona	2.1	1,218	702	3,654	2,105
Tennessee	2.0	1,160	668	3,480	2,005
Missouri	1.9	1,102	635	3,306	1,905
Maryland	1.9	1,102	635	3,306	1,905
Wisconsin	1.9	1,102	635	3,306	1,905
Minnesota	1.7	986	568	2,958	1,704
Colorado	1.6	928	535	2,784	1,604
Alabama	1.5	870	501	2,610	1,504
South Carolina	1.5	870	501	2,610	1,504
Louisiana	1.4	812	468	2,436	1,403
Kentucky	1.4	812	468	2,436	1,403
Oregon	1.2	696	401	2,088	1,203
Oklahoma	1.2	696	401	2,088	1,203
Connecticut	1.2	696	401	2,088	1,203
Iowa	1.0	580	334	1,740	1,002
Mississippi	1.0	580	334	1,740	1,002
Arkansas	0.9	522	301	1,566	902
Kansas	0.9	522	301	1,566	902
Utah	0.9	522	301	1,566	902
Nevada	0.9	522	301	1,566	902
New Mexico	0.7	406	234	1,218	702
West Virginia	0.6	348	200	1,044	601
Nebraska	0.6	348	200	1,044	601
Idaho	0.5	290	167	870	501
Maine	0.4	232	134	696	401
New Hampshire	0.4	232	134	696	401
Hawaii	0.4	232	134	696	401

Table 5A. Large states: Proportionate allocation with three-year estimates (Continued)

State	A: Percentage of the U.S. population	B: Estimated sample size of individuals - Current sample size	C: Estimated sample size of households - Current sample size	D: Estimated sample size of individuals - Three year combined	E: Estimated sample size of households - Three year combined
Rhode Island	0.4	232	134	696	401
Remainder of States	1.9	1,102	635	3,306	1,905
Total	100.0	58,000	33,415	174,000	100,245

Estimates for MSAs

Table 5B, columns D and E presents an analogue to Table 5A using three-year estimates for large MSAs. Once again, comparing to the one-year estimates given in columns B and C, one can see that estimates at the MSA level have been greatly facilitated, using the benchmarks of acceptable precision discussed earlier (9,125 persons for an MOE of 2.9, 4,055 persons for an MOE of 4.4 and 2,282 persons for an MOE of 5.9). From column D, 1 MSA, 3 MSAs and 14 MSAs have a sufficient number of persons sampled for estimating violent crimes with MOEs of 2.9, 4.4, and 5.9 respectively. This is an increase of 12 in the number of MSAs that can produce estimates with an MOE of 5.9 in comparison to the proportionate allocation design both with and without sample increase (see Table 4B). Using the benchmarks of acceptable precision discussed earlier (1,015 households for an MOE of 18.5, 451 households for an MOE of 27.8 and 254 households for an MOE of 37.1), we can see from column E that 16 MSAs, 39 MSAs and 67 MSAs have a sufficient number of households sampled for estimating property crimes with MOEs of 18.5, 27.8, and 37.1, respectively. Once again, this is a substantial increase in the number of MSAs that have adequate precision to estimate property crime rates in comparison to the proportionate allocation design both with and without sample increase (see Table 4B).

Table 5B. Large MSAs: Proportionate allocation with three-year estimates

MSA	A: Percentage of the U.S. population	B: Estimated Sample size of individuals - Current sample size	C: Estimated sample size of households - Current sample size	D: Estimated sample size of individuals - Three year combined	E: Estimated sample size of households - Three year combined
New York-Northern New Jersey-Long Island, NY-NJ-PA	6.2	3,622	2,087	10,865	6,260

Table 5B. Large MSAs: Proportionate allocation with three-year estimates (Continued)

MSA	A: Percentage of the U.S. population	B: Estimated Sample size of individuals - Current sample size	C: Estimated sample size of households - Current sample size	D: Estimated sample size of individuals - Three year combined	E: Estimated sample size of households - Three year combined
Los Angeles-Long Beach-Santa Ana, CA	4.2	2,453	1,413	7,359	4,240
Chicago-Naperville-Joliet, IL-IN-WI	3.1	1,824	1,051	5,471	3,152
Dallas-Fort Worth-Arlington, TX	2.1	1,200	692	3,601	2,075
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	1.9	1,113	641	3,338	1,923
Houston-Sugar Land-Baytown, TX	1.9	1,092	629	3,275	1,887
Miami-Fort Lauderdale-Pompano Beach, FL	1.8	1,032	594	3,095	1,783
Atlanta-Sandy Springs-Marietta, GA	1.8	1,024	590	3,073	1,771
Washington-Arlington-Alexandria, DC-VA-MD-WV	1.8	1,021	588	3,063	1,765
Boston-Cambridge-Quincy, MA-NH	1.5	862	497	2,586	1,490
Detroit-Warren-Livonia, MI	1.5	843	486	2,530	1,457
Phoenix-Mesa-Scottsdale, AZ	1.4	816	470	2,448	1,410
San Francisco-Oakland-Fremont, CA	1.4	815	469	2,444	1,408
Riverside-San Bernardino-Ontario, CA	1.4	784	452	2,353	1,356
Seattle-Tacoma-Bellevue, WA	1.1	637	367	1,912	1,102
Minneapolis-St. Paul-Bloomington, MN-WI	1.1	615	355	1,846	1,064
San Diego-Carlsbad-San Marcos, CA	1.0	572	329	1,716	988
St. Louis, MO-IL	0.9	537	309	1,610	928

Table 5B. Large MSAs: Proportionate allocation with three-year estimates (Continued)

MSA	A: Percentage of the U.S. population	B: Estimated Sample size of individuals - Current sample size	C: Estimated sample size of households - Current sample size	D: Estimated sample size of individuals - Three year combined	E: Estimated sample size of households - Three year combined
Tampa-St. Petersburg-Clearwater, FL	0.9	521	300	1,563	900
Baltimore-Towson, MD	0.9	508	293	1,525	878
Denver-Aurora, CO	0.8	478	275	1,433	826
Pittsburgh, PA	0.8	448	258	1,344	774
Portland-Vancouver-Beaverton, OR-WA	0.7	421	242	1,262	727
Cincinnati-Middletown, OH-KY-IN	0.7	411	237	1,232	710
Sacramento–Arden-Arcade–Roseville, CA	0.7	402	232	1,206	695
Cleveland-Elyria-Mentor, OH	0.7	398	229	1,194	688
Orlando-Kissimmee, FL	0.7	392	226	1,175	677
San Antonio, TX	0.7	387	223	1,161	669
Kansas City, MO-KS	0.7	381	220	1,144	659
Las Vegas-Paradise, NV	0.6	356	205	1,067	614
San Jose-Sunnyvale-Santa Clara, CA	0.6	347	200	1,040	599
Columbus, OH	0.6	338	195	1,014	584
Indianapolis-Carmel, IN	0.6	327	188	981	565
Charlotte-Gastonia-Concord, NC-SC	0.6	324	187	973	560
Virginia Beach-Norfolk-Newport News, VA-NC	0.5	316	182	948	546
Austin-Round Rock, TX	0.5	315	181	945	544
Providence-New Bedford-Fall River, RI-MA	0.5	304	175	913	526
Nashville-Davidson–Murfreesboro–Franklin, TN	0.5	295	170	886	511

Table 5B. Large MSAs: Proportionate allocation with three-year estimates (Continued)

MSA	A: Percentage of the U.S. population	B: Estimated Sample size of individuals - Current sample size	C: Estimated sample size of households - Current sample size	D: Estimated sample size of individuals - Three year combined	E: Estimated sample size of households - Three year combined
Milwaukee-Waukesha-West Allis, WI	0.5	295	170	886	510
Jacksonville, FL	0.4	250	144	751	433
Memphis, TN-MS-AR	0.4	245	141	735	423
Louisville/Jefferson County, KY-IN	0.4	237	137	712	410
Richmond, VA	0.4	234	135	701	404
Oklahoma City, OK	0.4	230	132	690	397
Hartford-West Hartford-East Hartford, CT	0.4	227	131	681	392
New Orleans-Metairie-Kenner, LA	0.4	216	124	648	373
Buffalo-Niagara Falls, NY	0.4	214	123	643	370
Birmingham-Hoover, AL	0.4	213	123	639	368
Salt Lake City, UT	0.4	213	122	638	367
Raleigh-Cary, NC	0.4	207	120	622	359
Rochester, NY	0.3	197	114	591	341
Tucson, AZ	0.3	193	111	579	333
Tulsa, OK	0.3	175	101	524	302
Fresno, CA	0.3	173	100	520	299
Honolulu, HI	0.3	172	99	517	298
Bridgeport-Stamford-Norwalk, CT	0.3	171	98	512	295
Albany-Schenectady-Troy, NY	0.3	163	94	488	281
New Haven-Milford, CT	0.3	161	93	484	279
Albuquerque, NM	0.3	161	93	484	279
Omaha-Council Bluffs, NE-IA	0.3	160	92	479	276
Dayton, OH	0.3	159	92	478	276
Allentown-Bethlehem-Easton, PA-NJ	0.3	154	89	462	266
Bakersfield, CA	0.3	153	88	458	264
Oxnard-Thousand Oaks-Ventura, CA	0.3	152	88	456	263

Table 5B. Large MSAs: Proportionate allocation with three-year estimates (Continued)

MSA	A: Percentage of the U.S. population	B: Estimated Sample size of individuals - Current sample size	C: Estimated sample size of households - Current sample size	D: Estimated sample size of individuals - Three year combined	E: Estimated sample size of households - Three year combined
Worcester, MA	0.3	149	86	448	258
Grand Rapids-Wyoming, MI	0.3	148	85	444	256
Baton Rouge, LA	0.3	148	85	443	255
Remainder of MSAs	24.7	14,338	8,260	43,013	24,781
nonMSAs	16.5	9,563	5,509	28,689	16,528
TOTAL	100.0	58,000	33,415	174,000	100,245

In conclusion,

- Under the proportionate allocation design with no sample increase, no more than five states and two MSAs have sufficient sample to produce estimates with any level of precision for violent crimes (less than a 5.9 MOE). However, as many as 29 states and 22 MSAs have sufficient sample to produce estimates of adequate precision for property crimes (less than a 27.8 MOE).

- Even with increased sample of 13,000 supplemental persons allocated proportionately, an additional one to five states and one to seven MSAs (over and above what is stated above) have sufficient sample to produce estimates for violent crimes or property crimes, depending on the level of precision desired. Balancing the substantial cost involved in a supplemental sample against the modest gains in the increased number of states and MSAs for which LAEs can be produced, the extra use of resources is not warranted under the proportionate allocation design scenario. Disproportionate allocation of the supplemental sample will yield better results, as will be seen in a subsequent section.

- However, under proportionate allocation with the utilization of three-year estimates, depending on the level of precision desired, between four to 24 states and one to 14 MSAs have sufficient sample to produce estimates for violent crimes, whereas between 22 to 43 states and 16 to 67 MSAs can do so for property crimes. Therefore, this last scenario offers the most promise under a situation of minimal design change.

- Even under this minimum change approach, the current stratification needs to be reworked to respect state boundaries, and a sufficient number of PSUs need to be allocated to states and/or MSAs in order to ensure stable variance estimates at the appropriate level.

2.3 LAEs Under Designs with Disproportionate Allocation

To increase the numbers of states and MSAs that have reliable estimates of crime rates from those achieved in section 2.2, the current design that uses proportionate allocation needs to be altered. As mentioned earlier, disproportionate allocation schemes, while not optimal for national estimates, are more appropriate to facilitate the production of these LAEs. In this section, a number of variants under scenarios of disproportionate allocation are considered, including the cases of both current and increased sample sizes.

2.3.1 Estimates for States with Disproportionate Allocation

Variant 1 Equal Allocation

For estimates at the state level, one possible variant is equal allocation to each state. The term "disproportionate" is used here to reflect the fact that sample is being distributed unevenly with respect to population size across the country, although it is distributed evenly to each state. In the case of equal allocation, if larger and smaller states alike are allotted identical sample, this is clearly advantageous for producing estimates at the state level, but will cause a loss of precision for estimates at the national level. Equal allocation is an extreme form of disproportionate allocation that focuses on the production of state level estimates.

For estimation at the state level, fixing the sample size at the current level (58,000 in 2007) results in about 1,160 participating persons and about 668 participating households per state. If we also assume that the current 203 PSUs are also allocated evenly across states, this results in roughly four PSUs per state and therefore, the $DEFF$ due to clustering at the state level under equal allocation remains the same as that at the national level under proportionate allocation for the various crime rates (see Tables 2A and 2B). The same sample size requirements can be applied as in section 2.2 to achieve estimates for violent crimes and property crimes to ensure the corresponding MOEs discussed earlier. Clearly under this scenario, there will be sufficient sampled households to achieve only the lowest level of precision for property crimes for all states, and insufficient sampled persons to produce estimates for violent crimes having any of the given levels of precision discussed earlier. Increasing the sample size as before (to 71,000 sampled persons) will result in a sample size of roughly 1,420 persons (818 households) per state – again seriously short of the required minimum for violent crimes for any level of precision, but sufficient for all states for property crimes for modest levels of precision. As an alternative to directly increasing the sample size, the use of three-year estimates will triple the current sample size to 3,480 persons (2,004 households) per state, which is adequate for all states for the production of both violent crime estimates (at the lowest level of precision) and property crimes (at any level of precision).

Regardless, a major disadvantage of the equal allocation design is that it would have an impact on national estimates. From the formula given in section 2.1.1, it can be shown that differential sampling rates by state under equal allocation would result in a $DEFF$ for unequal weighting of $DEFF_{UEW} = 2.25$. Therefore at the national level, assuming an average value of 2.0 for the $DEFF$ due to clustering, the overall $DEFF$ under equal allocation would be $DEFF = DEFF_{CLU} * DEFF_{UEW} = 2.0*2.25 = 4.5$, a serious increase from $DEFF = 2.0*1.0 = 2.0$ under the design of proportionate allocation. Thus, we would expect SEs and CVs for national estimates to be over 50 percent higher than under the proportionate allocation design. The increase of 50 percent is a result of taking the ratio $DEFF = 4.5$ (under equal allocation) to $DEFF = 2.0$ (under proportionate allocation), and then taking the square root of the ratio.

Variant 2 Proportional to Square Root Allocation (PSR Allocation).

The equal sample size allocation to states is an extreme situation that is best for state estimates, whereas a proportionate allocation to states is best for national estimates. Another option is to use a compromise between the two allocations, such as power allocations discussed in Bankier (1988). A version of power allocation distributes sample sizes to states proportional to the square root (PSR) of the population. In effect, this will give a larger sample size to large states than with equal allocation, but not as large as with proportionate allocation. This allocation will also distribute a smaller sample size to smaller states than with equal allocation, but not as small as with proportionate allocation. Thus, the PSR allocation has the effect of dampening the sample size somewhat for larger states, and augmenting the sample size for smaller states, relative to proportionate allocation. The utility of this option depends on whether the relative increase in sample size given to smaller states is sufficient to provide estimates of reasonable quality for these states. However, a quick calculation (not shown) under this allocation shows that fewer states have sufficient sample for estimates of violent crimes and property crimes than under the design with proportionate allocation shown in Table 4A. This is because the allocation to the large states that could produce reliable estimates under the design with proportionate allocation has been dampened by reallocating some of the sample to the smaller states. Thus, this option will not be given further consideration since clearly it will not render results that are any better than under a design with proportionate allocation.

Variant 3 Proportionate Allocation for the Core Sample with a Supplemental Sample Disproportionately Distributed

Another alternative allocation uses proportionate allocation for the core sample of 58,000 persons, and a supplemental sample allocated to the large states that are just short of having adequate sample in relation to the precision levels discussed earlier. For example, one could allocate the 13,000 additional cases to the large states, in decreasing order of size, so that each has a minimum sample of 2,444 persons (for example) to estimate violent crimes (corresponding to a MOE of 5.9 and a CV of 20 percent). Thus, from Table 4A column B, the candidates for additional sample are: Pennsylvania, Ohio, Michigan, Georgia, North Carolina, etc. Note that estimates for violent crimes have more stringent sample size requirements than those for property crimes for the same level of CV (i.e., 382 households for an MOE of 37.1 and a CV of 20 percent). Thus, all states that will be able to produce reliable estimates for violent crimes also will be capable of producing reliable estimates for property crimes for MOEs corresponding to the same CV level.

Table 6A shows such an allocation in columns B and C (labeled Scheme 1). In column B, roughly 13,000 additional persons have been allocated to 14 states (Pennsylvania through Maryland) in such a way that the core sample from proportionate allocation is brought up to the level of exactly 2,444. (Compare with the values in Table 4A, column B to see the original core sample allocation). Note that the intention of this allocation is to bring as many states as possible to the level where they can produce estimates of violent crimes with an MOE of 5.9 or less. Column B shows that 19 large states have the capacity to do so. This compares with seven states under the increased sample scenario of Table 4A column D. Table 6A column C gives the corresponding number of households that results from the allocation of supplemental persons in column B. Here four states, 19 states, and 29 states have a sufficient number of households sampled for estimating property crimes with MOEs of 18.5, 27.8 and 37.1 respectively. This is two to three states less than under the scenario of increased sample given in Table 4A column E, for any level of precision.

An alternative example to the above scenario is one that would allocate 7,490 additional households (instead of allocating 13,000 persons) to the large states so that each has a minimum sample of 1,528 households (for example) to estimate property crimes (corresponding to an MOE of 18.5 and a

CV of 10 percent). Table 6A shows such an allocation in columns D and E (labeled Scheme 2). In column E, roughly 7,490 additional households have been allocated to 13 states (Illinois through Tennessee) in such a way that the core sample from proportionate allocation is "topped up" to exactly 1,528. (Compare with the values in Table 4A, column C giving the original core sample allocation). The intention of this allocation is that as many states as possible should be able to produce estimates of property crimes with an MOE of 18.5. Column E shows that 17 large states have the capacity to do so. This compares with seven states under the increased sample scenario of Table 4A column E. Table 6A column D gives the corresponding number of persons that results from the allocation of supplemental households in column E. Here zero states, two states, and 17 states have a sufficient number of persons sampled for estimating violent crimes with MOEs of 2.9, 4.4, and 5.9 respectively. This is ten states more than under the scenario of increased sample given in Table 4A column D for an MOE of 5.9.

Thus, this approach shows the greatest promise in terms of providing estimates for crime rates for the largest number of states, utilizing a supplemental sample of 13,000 persons or 7,490 households. Note that the thresholds of 2,444 persons or 1,528 households that were used in the above schemes were chosen for illustrative purposes only. Any of the other thresholds discussed earlier could be used as well provided the corresponding levels of precision were deemed adequate.

In terms of impact on national estimates, the $DEFF$ due to clustering remains at $DEFF_{CLU} = 2.0$ using the same argument above. However, it can be shown that under the allocation scheme given in Table 6A columns B and C, there is an increase in the $DEFF$ due to differential sampling rates by states to $DEFF_{UEW} = 1.22$ under this disproportionate allocation compared to $DEFF_{UEW} = 1.00$ under the design of proportionate allocation. Therefore, the overall $DEFF$ at the national level under the proposed allocation is $DEFF = 2.0 * 1.22 = 2.44$. In conclusion, we would expect SEs and MOEs for national estimates to be only 10 percent higher than under a design with proportionate allocation having the same sample size. The allocation scheme in columns D and E would have roughly the same impact on national estimates.

Note that it is also possible to extend the above methodology to obtain LAEs for states and remainders of Census Divisions (CDs). The country is divided into nine CDs; all 50 states and the District of Columbia fall within a distinct CD, and CDs respect state boundaries. States can be presented in descending order of size within the CD to which they belong. An attempt can be made to obtain reliable estimates for as many large states as possible in a given CD using the method described above. After the allocation is made, within each CD, the smaller "leftover" states, each of which has inadequate sample size individually, can be grouped together and labeled "Remainder of CD". The intention is that the group of leftover small states together may have sufficient sample size to produce estimates of crime rates of reasonable precision at the grouped state level. Note that CDs are only one possible way to group states. From the point of view of data users, it may make sense to consider a similar analysis, ignoring geographic location and grouping states together by similar crime and other relevant characteristics instead.

Table 6A. Large states: Proportionate allocation with an increased sample size, where the supplemental sample is disproportionately distributed

State	A: Percentage of the U.S. population	Scheme 1		Scheme 2	
		B: Estimated sample size of individuals - Proportionate allocation with supplemental sample added to bring large states to minimum 2,444 persons	C: Estimated sample size of households corresponding to column B	D: Estimated sample size of individuals corresponding to column E	E: Estimated sample size of households - Proportionate allocation with supplemental sample added to bring large states to minimum 1,528 households
California	12.1	7,018	4,043	7,018	4,043
Texas	7.9	4,582	2,640	4,582	2,640
New York	6.4	3,712	2,139	3,712	2,139
Florida	6.1	3,538	2,038	3,538	2,038
Illinois	4.3	2,494	1,437	2,652	1,528
Pennsylvania	4.1	2,444	1,408	2,652	1,528
Ohio	3.8	2,444	1,408	2,652	1,528
Michigan	3.3	2,444	1,408	2,652	1,528
Georgia	3.2	2,444	1,408	2,652	1,528
North Carolina	3.0	2,444	1,408	2,652	1,528
New Jersey	2.9	2,444	1,408	2,652	1,528
Virginia	2.6	2,444	1,408	2,652	1,528
Washington	2.1	2,444	1,408	2,652	1,528
Massachusetts	2.1	2,444	1,408	2,652	1,528
Indiana	2.1	2,444	1,408	2,652	1,528
Arizona	2.1	2,444	1,408	2,652	1,528
Tennessee	2.0	2,444	1,408	2,652	1,528
Missouri	1.9	2,444	1,408	1,102	635
Maryland	1.9	2,444	1,408	1,102	635
Wisconsin	1.9	1,102	635	1,102	635
Minnesota	1.7	986	568	986	568
Colorado	1.6	928	535	928	535
Alabama	1.5	870	501	870	501
South Carolina	1.5	870	501	870	501

Table 6A. Large states: Proportionate allocation with an increased sample size, where the supplemental sample is disproportionately distributed

State	A: Percentage of the U.S. population	Scheme 1		Scheme 2	
		B: Estimated sample size of individuals - Proportionate allocation with supplemental sample added to bring large states to minimum 2,444 persons	C: Estimated sample size of households corresponding to column B	D: Estimated sample size of individuals corresponding to column E	E: Estimated sample size of households - Proportionate allocation with supplemental sample added to bring large states to minimum 1,528 households
Louisiana	1.4	812	468	812	468
Kentucky	1.4	812	468	812	468
Oregon	1.2	696	401	696	401
Oklahoma	1.2	696	401	696	401
Connecticut	1.2	696	401	696	401
Remainder of States	11.5	6,670	3,843	6,670	3,843
Total	100.0	70,698	40,731	70,671	40,715

2.3.2 Estimates for MSAs with Proportionate Allocation for the Core Sample and with a Supplemental Sample Disproportionately Distributed

If the interest is in producing LAEs for MSAs, an alternate way of using the additional resources of 13,000 cases is to allocate them to the largest MSAs that are most in need, in a manner very similar to that of Variant 3 for states. That is, the current proportionate allocation is used across the country and the largest MSAs receive their proportionate share. Those MSAs that fall short of the threshold minimum sample sizes discussed earlier are supplemented in decreasing order of size. Table 6B presents such an allocation for the largest MSAs.

Columns B and C give an analogous allocation of persons to MSAs as in Table 6A for states, except that the threshold of 2,282 persons is used in place of 2,444 (labeled Scheme 1). In column B, roughly 13,000 additional persons have been allocated to 11 MSAs (Chicago through San Francisco) in such a way that the core sample has been brought up to the level of the threshold for as many MSAs as possible. The intention of this allocation is to bring MSAs to the level where they can produce estimates of violent crimes with an MOE of 5.9. Column B shows that 13 large MSAs have the capacity to do so. This compares with MSAs under the increased sample scenario of Table 4B column D. Table 6B column C gives the corresponding number of households that results from the allocation of supplemental persons in column B. Here 13 MSAs, 14 MSAs and 22 MSAs have a sufficient number of households sampled for estimating property crimes with MOEs of 18.5, 27.8 and 37.1 respectively.

Similarly, Table 6B columns D and E give an analogous allocation of households to MSAs as in Table 6A for states, except that the threshold of 1,015 households is used in place of 1,528 (labeled Scheme 2). In column E, roughly 7,490 additional households have been allocated to 15 MSAs (Dallas through St. Louis) in such a way that the core sample from proportionate allocation is brought up to the threshold of 1,015 for as many MSAs as possible. The intention of this allocation is that these MSAs should be able to produce estimates of property crimes with an MOE of 18.5. Column E shows that 18 large MSAs have the capacity to do so. This compares with three MSAs under the increased sample scenario of Table 4B column E. Table 6B column D gives the corresponding number of persons that results from the allocation of supplemental households in column E. Here only two MSAs have a sufficient number of persons sampled for estimating violent crimes with an MOE of 5.9. This is the same number as under the scenario of increased sample given in Table 4B column D for an MOE of 5.9.

Thus, this disproportionate method of allocating the supplemental sample provides a greater capacity to produce LAEs at the MSA level than that given in Table 4B, where the supplemental sample is allocated proportionately. However, because of the deviation from pure proportionate allocation there will be some impact on the precision of estimates at the national level. Specifically, it can be shown that the $DEFF$ at the national level due to differential sample rates for MSAs will be increased to $DEFF_{UEW} = 1.1$, in comparison to that under proportionate sampling where $DEFF_{UEW} = 1.00$. Thus, we would expect SEs and MOEs for national estimates to be 5 percent higher than under a design with proportionate allocation having the same sample size. The allocation scheme in columns D and E would have roughly the same impact on national estimates.

Table 6B. Large MSAs: Proportionate allocation with an increased sample size, where the supplemental sample is disproportionately distributed

MSA	A: Percentage of the U.S. population	Scheme 1		Scheme 2	
		B: Estimated sample size of individuals - Proportionate allocation with supplemental sample added to bring large MSAs to minimum 2,282 persons	C: Estimated sample size of households corresponding to column B	D: Estimated sample size of individuals corresponding to column E	E: Estimated sample size of households - Proportionate allocation with supplemental sampled added to bring large MSAs to minimum 1,015 households
New York-Northern New Jersey-Long Island, NY-NJ-PA	6.2	3,622	2,087	3,622	2,087
Los Angeles-Long Beach-Santa Ana, CA	4.2	2,453	1,413	2,453	1,413
Chicago-Naperville-Joliet, IL-IN-WI	3.1	2,282	1,315	1,824	1,051
Dallas-Fort Worth-Arlington, TX	2.1	2,282	1,315	1,762	1,015
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	1.9	2,282	1,315	1,762	1,015
Houston-Sugar Land-Baytown, TX	1.9	2,282	1,315	1,762	1,015
Miami-Fort Lauderdale-Pompano Beach, FL	1.8	2,282	1,315	1,762	1,015
Atlanta-Sandy Springs-Marietta, GA	1.8	2,282	1,315	1,762	1,015
Washington-Arlington-Alexandria, DC-VA-MD-WV	1.8	2,282	1,315	1,762	1,015
Boston-Cambridge-Quincy, MA-NH	1.5	2,282	1,315	1,762	1,015
Detroit-Warren-Livonia, MI	1.5	2,282	1,315	1,762	1,015
Phoenix-Mesa-Scottsdale, AZ	1.4	2,282	1,315	1,762	1,015
San Francisco-Oakland-Fremont, CA	1.4	2,282	1,315	1,762	1,015
Riverside-San Bernardino-Ontario, CA	1.4	784	452	1,762	1,015
Seattle-Tacoma-Bellevue, WA	1.1	637	367	1,762	1,015
Minneapolis-St. Paul-Bloomington, MN-WI	1.1	615	355	1,762	1,015

Table 6B. Large MSAs: Proportionate allocation with an increased sample size, where the supplemental sample is disproportionately distributed (Continued)

MSA	A: Percentage of the U.S. population	Scheme 1		Scheme 2	
		B: Estimated sample size of individuals - Proportionate allocation with supplemental sample added to bring large MSAs to minimum 2,282 persons	C: Estimated sample size of households corresponding to column B	D: Estimated sample size of individuals corresponding to column E	E: Estimated sample size of households - Proportionate allocation with supplemental sampled added to bring large MSAs to minimum 1,015 households
San Diego-Carlsbad-San Marcos, CA	1.0	572	329	1,762	1,015
St. Louis, MO-IL	0.9	537	309	1,762	1,015
Tampa-St. Petersburg-Clearwater, FL	0.9	521	300	521	300
Baltimore-Towson, MD	0.9	508	293	508	293
Denver-Aurora, CO	0.8	478	275	478	275
Pittsburgh, PA	0.8	448	258	448	258
Remainder of MSAs	44.2	25,621	14,761	25,621	14,761
nonMSAs	16.5	9,563	5,509	9,563	5,509
TOTAL	100.0	71,461	41,170	71,464	41,172

In conclusion,

- Under proportionate allocation in the core national sample, with a supplemental sample of roughly 13,000 persons allocated disproportionately to the large states falling short of the threshold of 2,444 persons (Scheme 1), 19 states are capable of producing estimates of both violent crimes and property crimes, with MOEs of roughly 5.9 and 27.8 respectively.

- Under proportionate allocation in the core national sample, with a supplemental sample of roughly 7,490 households allocated disproportionately to the large states falling short of the threshold of 1,528 households (Scheme 2), 17 states are capable of producing estimates of both violent crimes and property crimes, with MOEs of roughly 5.9 and 18.5 respectively.

- The above two schemes for states offer the most promise in terms of the use of supplemental sample to provide estimates for crime rates with reasonable precision for the largest number of states, while only increasing SEs and MOEs at the national level by 10 percent over a design with proportionate allocation having the same sample size.

- Using a similar approach for MSAs (as an alternative use of the additional resources), reliable estimates for both violent crimes and property crimes with MOEs 5.9 and 18.5 respectively, can be produced for the 13 largest MSAs using a supplemental sample of 13,000 persons allocated to those MSAs wherever there is a shortfall in sample from the threshold of 2,282 persons (Scheme 1). This approach will have only a modest impact on national estimates by increasing SEs and MOEs by 5 percent over a design with proportionate allocation having the same sample size.

- Under proportionate allocation, with a supplemental sample of roughly 7,490 households allocated disproportionately to the large MSAs falling short of the threshold of 1,015 households (Scheme 2), 18 MSAs are capable of producing estimates of property crimes with an MOE of 18.5.

- The above two schemes for MSAs offer the most promise in terms of the use of supplemental sample to provide estimates for crime rates with reasonable precision for the largest number of MSAs.

- Although not discussed explicitly in this section, further gains can be realized by producing multi-year estimates under disproportionate allocation. This is discussed in Section 2.2.5 in the context of proportionate allocation.

The benefits of disproportionate sampling can be summarized in Table 7 showing the number of states and MSAs capable of producing estimates of adequate precision. Adequate precision here is defined by an MOE = 5.9 for violent crimes and 37.1 for property crimes, each corresponding to CVs of about 0.2. This may or may not ultimately be considered an adequate level of precision, however, it serves to illustrate the benefits of the various allocation strategies.

Another possibility, not discussed above, is to concentrate additional sample to a smaller subset of local areas (e.g., regions, cities). The additional sample would be rotated every three or four years to another area. This would significantly increase precision for the estimates and could provide periodic information for local policymakers.

Table 7. Number of states and MSAs capable of producing estimates of adequate precision under allocation schemes for direction estimation

Allocation scheme	Violent crime Number of States	Violent crime Number of MSAs	Property crime Number of States	Property crime Number of MSAs
Proportionate				
Current sample size	5	2	29	22
Increased sample size	7	2	31	29
Three-year combined	24	14	43	67
Disproportionate				
13,000 persons added to core national sample	19	13	29	22

Note: Adequate precision is defined by an MOE = 5.9 for violent crimes and an MOE = 37.1 for property crimes.

2.4 Next Steps

This chapter shows direct estimates for selected states and MSAs can be produced for violent and property crimes. Three or more years of data can be aggregated to greatly improve the precision of the estimates and increase the number of publishable local areas. It should be noted that if estimates are to be produced at this time, the MSAs that were declared certainty PSUs at the time of selection would be eligible for consideration.

At the time the NCVS began in the early 1970's, a separate set of surveys were conducted for the largest central cities in the U.S. This effort eventually was dropped from the NCVS program. Generating estimates for MSA's or even central cities would be similar to those provided by this early program, but it would not involve a separate data collection. By aggregating data over time it is possible to produce data at these levels. Feedback from potential data users, from both interviews conducted for this project and at least one BJS sponsored workshop with users, indicated that periodic release of data for local areas (e.g., every three years) would meet many of the needs for local users.

Some extra work is needed to generate state-level estimates. Since the stratification scheme was designed for national estimates, the sampling strata cross state lines and a poststratification weighting adjustment would be needed to re-align the strata boundaries with state boundaries. The amount of effort would be more than the usual effort to conduct a poststratification adjustment.

Also, in the short term, a discussion is needed to identify the LAEs that are to be produced. The discussion has implications on the next steps, as well as the national sample design. For example, interest in crime statistic areas would require research and collaborative efforts in identifying appropriate reporting areas (such as groups of small states or groups of counties within large states). As another example, if it is determined to produce estimates for all states, it would impact the national design, the design of the supplemental sample, as well as the small area modeling approach.

The discussion of direct estimates revolved around the incidence rates for the major crimes. However, before moving forward there needs to be a discussion on the need to generate estimates for crime characteristics as well. Data users expressed interest in both incidence rates and crime characteristics. Chapter 4 provides a sample size analysis that considers crime characteristics in addition to crime rates, as it relates to a supplemental sample. But there needs to be more consideration of the implications of generating information on crime characteristics when planning the publication of LAEs.

Narrowing the LAE objectives are critical to deciding on the re-design of the national sample. The key decisions in the re-design work are related to stratification strategies for state estimates, and considerations for a more widely scattered sample. The objective of a widely scattered sample is to achieve more accurate direct estimates in the local area, as well as to help facilitate variance estimation for those areas (discussed further in section 2.2.1). The more widely disbursed the sample, the less efficient it is for national estimates. This chapter provides the benefits of disproportionate allocation (as opposed to proportionate allocation as in the current NCVS design) in order to produce direct estimates for more local areas. The impact of increasing the overall sample size is also analyzed according to the level of precision. Chapter 4 discusses possible ways to supplement the sample in ways that might prove more efficient by using less expensive data collection methods.

Indirect Estimation 3

The usual survey sampling direct estimation approach discussed in Chapter 2 has the attractive feature that direct estimates do not depend on the validity of statistical models. The approach is therefore the standard one used for most survey analyses. However, as demonstrated in Chapter 2, there are limits on the ability of direct estimation to produce LAEs of crime rates with acceptable levels of precision. This chapter therefore discusses the possibility of using model-dependent indirect estimation (or SAE) methods for producing local area crime rate estimates. The local areas under consideration include all states, the larger metropolitan areas, and perhaps also some or all counties and smaller metropolitan areas.

A considerable amount of research and development in SAE methods has taken place in the past twenty years or so. The text by Rao (2003) presents a comprehensive overview of the methods, history, and applications of SAE methods. The essence of SAE is to use auxiliary data at the small area level in combination with survey data to model the small area parameters of interest. A wide variety of models has been developed for this purpose. There are two major types of models: area-level and unit-level. The area-level approach models the small area parameter of interest in terms of auxiliary data at the area level, whereas the unit level[5] approach models the underlying variable of interest in terms of unit-level auxiliary data known at the small area level, and then aggregating the individual predictions for each small area.

A number of Federal statistical agencies have produced LAEs using SAE methods. The Task 2 memo listed such programs for both the U.S. and other countries. This chapter describes three examples in more detail. One example is the Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program, which produces annually a number of income and poverty related estimates for states, counties, and school districts. The second example is from the National Center for Education Statistics, which has produced state and county level estimates of the percentages of adults lacking Basic prose literacy skills based on the National Assessment of Adult Literacy (NAAL) survey. The third example is from the Office of Applied Studies at the Substance Abuse and Mental Health Services Administration (SAMHSA), which regularly produces state and substate level estimates of substance abuse based on the National Survey of Drug Use and Health (NSDUH).

[5] 'Unit level' can mean at the individual sample unit level (person or household), or it could mean a geographic area lower than the small area. As discussed in Section 3.2.3, the NSDUH application defines it at the person level.

Section 3.1 describes the models used in these three applications as illustrations of possible approaches that might be adopted for producing LAEs from the NCVS.

The effectiveness of SAE methods depends critically on the availability of auxiliary variables at the local level that are strong predictors of the parameter being estimated. Some potential auxiliary variables for estimating various local area crime rates in conjunction with the NCVS are discussed in Section 3.2.

Section 3.3 describes further aspects of model development, such as details about area-level vs. unit-level models, univariate vs. multivariate models, and methods for benchmarking and validating indirect estimates.

Section 3.4 provides conclusions and discussion for indirect estimation.

Chapter 4 examines how data collected in a supplementary sample survey could be used to facilitate local area estimation, including the possibility of these data being incorporated in an indirect estimation program.

3.1 Some Models for SAE

This section describes the SAE models used in the SAIPE program, the NAAL state and county level estimates, and the NSDUH state and substate estimates.

3.1.1 SAE for SAIPE

The Census Bureau's SAIPE program produces a number of income and poverty related estimates for states, counties, and school districts using area-level models based on the American Community Survey (ACS) and small area auxiliary data from the Internal Revenue Service and other sources (http://www.census.gov/hhes/www/saipe/). Examples of the small area estimates produced by the SAIPE program include median household income and the percentages and numbers of total persons, children, and school-age children below the poverty level. These annual estimates are used to allocate funding for many federal grant programs to state and local jurisdictions, including nutrition assistance, medical assistance, jobs training, housing, and education (see, for example, Citro and Kalton 2000, Chapter 2).

The starting points of the current SAIPE program models are estimates derived from the American Community Survey (ACS)[6]. As an example, one of the models estimates the number of related children ages five to 17 in families in poverty at the county level. The dependent variable z_i in this model is the logarithm of the estimated number of such children for county i obtained from the most recent year of the ACS data collection. The following mixed effects regression model is used:

$$z_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + u_i + e_i$$

Where β_0 through β_5 are fixed regression parameters, u_i is a random intercept representing the difference between the true value of the characteristic for the county and its model-based expectation, and e_i represents the ACS sampling error. The five predictor variables are denoted x_{ji}, where $j = 1, \ldots, 5,$ and are described as follows:

- The log of the number of child exemptions claimed on tax returns of families in poverty;
- The log of the number of Supplemental Nutrition Assistance Program (SNAP) benefits recipients;
- The log of the estimated resident population under age 18;
- The log of the total number of child exemptions on tax returns;
- The log of the Census 2000 estimate of the number of related children in poverty age 5 to 17.

The random effects u_i and e_i are assumed to be normally distributed with mean zero and variances σ_u^2 and σ_{ei}^2 respectively, and are assumed to be independent. The sampling error variance σ_{ei}^2 is estimated directly from the ACS data.

Using the Empirical Bayes approach (see for example Rao 2003, Chapter 9) the small-area estimate for county i has two components, the sampling model and the linking model. The sampling model for the direct estimate $\hat{z}_i$ from the ACS is given by:

---

[6] See http://www.census.gov/did/www/saipe/methods/statecounty/20062008state.html and http://www.census.gov/did/www/saipe/methods/statecounty/20062008county.html.

$$\hat{z}_i = z_i + e_i;$$

and the linking model for the model prediction $\tilde{z}_i$ from the regression model is described as:

$$\check{z}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{\beta}_4 x_{4i} + \hat{\beta}_5 x_{5i}$$

The final estimate $\tilde{z}_i$ (on the log scale) is a linear combination of the direct estimate and the model prediction:

$$\tilde{z}_i = \hat{\gamma}_i \check{z}_i + (1 - \hat{\gamma}_i)\hat{z}_i$$

Where the quantity $\hat{\gamma}_i = \hat{\sigma}_{ei}^2 / (\hat{\sigma}_{ei}^2 + \hat{\sigma}_u^2)$ is an estimate of the relative precision of $\check{z}_i$ as compared to $\hat{z}_i$, based on the estimated variances of each, and $\hat{\sigma}_{ei}^2$ and $\hat{\sigma}_u^2$ are estimators of σ_{ei}^2 and σ_u^2 respectively. Note that if there were no ACS sample in county i, then $\hat{\gamma}_i = 1$ and hence the county estimate would reduce to the model prediction $\check{z}_i$. (In fact, with ACS, this situation does not arise since some ACS sample is drawn in every county). On the other hand, if the ACS sample size is large in county i, the county estimate will be heavily influenced by the direct estimate.

3.1.2 SAE for NAAL

The National Center for Education Statistics has produced state and county level estimates of the percentages of adults lacking Basic prose literacy skills. The model used is a Hierarchical Bayes (HB) area-level model based on the 2003 NAAL and auxiliary data from the 2000 Census (see for example Mohadjer et al. 2009). A multi-stage sample design was employed for the NAAL with counties or groups of counties selected as the PSUs . Only counties within sampled PSUs have any sample cases in the NAAL.

The aim of the NAAL SAE model is to estimate the true percentage of adults who are lacking Basic prose literacy skills (as evaluated by the NAAL instrument) within county i in state s, given as p_{si} below. The unbiased estimator of p_{si} within a county with NAAL data is $\hat{p}_{si}$, which is subject to both sampling and measurement error[7],, combined into a single error term e_{si}. The dependent

[7] The measurement error is from the prose literacy test.

variable z_{si} in the mixed-effects regression model is the logit of the true percentage p_{si} (the logarithm of the odds, where the odds is the ratio $p_{si}/(1-p_{si})$. The following sampling model and mixed effects regression linking model were used in NAAL:

$$\hat{p}_{si} = p_{si} + e_{si}$$

$$z_{si} = \beta_0 + \beta_1 x_{1si} + \beta_2 x_{2si} + \beta_3 x_{3si} + \beta_4 x_{4si} + \beta_5 x_{5si} + \beta_6 x_{6si} + v_s + u_{si}$$

where, $z_{si} = \ln\left(p_{si}/(1-p_{si})\right)$, and

where β_0 through β_6 are fixed regression parameters, v_s is a state random intercept representing the difference between the true value of the characteristic for the state and its model-based expectation and u_{si} is a corresponding county random intercept, and the six predictor variables are denoted by x_{jsi}, where $j = 1,\ldots,6,$ and are described as follows:

- Percentage of the population who are foreign-born (and in the U.S. less than 20 years);
- Percentage of age 25 or older adults with a high school education or less;
- Percentage of Blacks and/or Hispanics;
- Percentage of persons below 150 percent of the poverty line;
- A dichotomous zero to one indicator for two Census divisions (New England and North Central).
- A dichotomous indicator for a State Assessment of Adult Literacy state[8].

The random effects v_s, u_{si} and e_{si} are assumed to be normally distributed with mean zero and variances σ_v^2, σ_u^2 and σ_{esi}^2 respectively, and are assumed to be independent. The variance σ_{esi}^2 is estimated from the NAAL.

The six predictor variables listed above were chosen after an assessment of extensive analyses of a large number of possible predictors. The larger set included county information from the 2000 Census and state information from the ACS and Census projections. This larger set was chosen as variables that were thought to be potentially correlated to literacy levels. The process of selecting the final variables from this larger set was based on correlation analyses and stepwise logistic regression

[8] These were states which had separate state literacy studies (partially funded by the state).

procedures. Mohadjer et al. (2009) provide a complete listing of the initial pool of variables and provide details of the variable selection process.

The NAAL small area program used an HB approach (see for example Rao 2003, Chapter 10) rather than the simpler Empirical Bayes approach, because of the use of the logit link in the regression model. This model is an 'unmatched model' as opposed to the SAIPE case of a 'matched model'. In the unmatched model case, as introduced by You and Rao (2002), there is an intermediate function between the unbiased sample estimate and the linear predictive model. The intermediate function (logit) was necessary due to small sample size as well as the low estimated proportions. This is this case as well for the NCVS, which makes the NAAL approach one that may be most applicable to the NCVS. The Task 3 memo provides a demonstration of the NAAL approach as it could be applied to NCVS data. As with the Empirical Bayes approach, the final indirect estimates at the county level are combinations (not necessarily linear) of model predictions and direct estimates for counties with NAAL survey data, and are entirely model predictions for counties without NAAL survey data. The state-level estimates are weighted averages of the county-level estimates.

3.1.3 SAE for NSDUH

The Office of Applied Studies at the SAMHSA regularly produces state level estimates of substance abuse using unit level models based on the NSDUH and a wide variety of sources for the auxiliary data (see for example Hughes et al. 2008). As for the NAAL, a multi-stage sample design was used for the NSDUH. The PSUs[9] are collections of adjacent Census block groups.

The NSDUH small-area model differs from the SAIPE and NAAL models in that it is an individual person level (unit-level) model (person k in age group a in state i, PSU j):

$$z_{aijk} = X_{aijk}\beta_a + \eta_{1i} + \eta_{2ij}$$

Where z_{aijk} is the logit of the probability that person ijk in age group[10] a answers 'yes' for a particular target characteristic (e.g., used alcohol in the past month), X_{aijk} is a vector of characteristics for the individual either from the NSDUH survey or defined for the household at the

[9] (see for example http://www.oas.samhsa.gov/nhsda/2k2nsduh/Results/appA.htm.)

[10] The four age groups a=1,…,4 correspond to ages 12 to 17, ages 18 to 24, ages 25 to 34, and ages over 34.

block group level or above (from auxiliary sources such as the Census), β_a is a vector of age-group specific fixed parameters, η_{1i} is a random intercept at the state level, and η_{2ij} is a random intercept at the PSU level. The random effects η_{1i} and η_{2ij} are assumed to be normally distributed with means zero and variances σ_1^2 and σ_2^2 respectively, and are assumed to be independent. Note that there is no sampling error represented here as in the area-level models. See for example Folsom et al. (1999) and Shah et al. (2000).

The predictor variables are selected from a large set using regression fits, and the selected variables vary widely across the many target characteristics (23 in 2005-2006: see Hughes et al. 2008 for a listing) and the four age groups. The following list gives a selection of some of the predictor variables[11]:

- Race/ethnicity and gender from the NSDUH instrument;
- Percentages from Claritas[12] by age (e.g., percent persons age 45 to 54), by race/ethnicity, and by gender, within block group, tract, and county geographic areas;
- Census 2000 tract-level percentages:
  - by education levels (e.g., percent population with zero to eight years of school);
  - by gender, family structure, marriage status (e.g., percent females head of household, no spouse, with child);
  - by labor force characteristics (e.g., percent males 16 and older in labor force);
  - median household income, median rent, percent homeowners/renters;
  - by poverty (percent families below poverty level, percent households with public assistance income).
- Uniform Crime Report program:
  - Percent drug possession arrest rates (e.g., marijuana, cocaine);
  - Percent illicit drug sale/manufacture arrest rate;
  - Serious, violent crime rate;
  - Driving under influence rate.

---

[11] See Hughes et al. 2008 for a complete listing.

[12] See http://en-us.nielsen.com/tab/product_families/nielsen_claritas.

- Urbanicity from the 2000 Census (MSA status, size of MSA, urban or rural status within MSA status);
- Alcohol, cigarette, drug death rates from NCHS at the county level.

The model predictors were selected from this larger set using a complex series of analyses including step-wise procedures and tree algorithms to select up to three-way interactions between the selected predictors.

The NSDUH small area program uses a Hierarchical Bayes approach (see for example Rao 2003, Chapter 10), as did NAAL. For this unit-level approach, the models were fit using the NSDUH survey weights in order that the parameter estimates would be sample-design consistent, whether or not the model is valid. The fitted model was used to generate predictions at the Census block group level. These block group level percentages were then aggregated up to the substate and state levels.

3.2 Development of Small Area Models: Predictor Variables

The key to producing good indirect estimators with the NCVS is the availability of predictor variables that are well-correlated to the local area target crime rates. Lacking such predictor variables, no degree of model sophistication can make up for this deficit. The quality of indirect estimators for local areas based on the NCVS will in fact depend heavily on the nature of the links between the target crime rates and predictor variables within the local areas. Three possible sources for predictor variables—the American Community Survey, Bureau of Justice Statistics administrative statistics, and the Uniform Crime Report Program—discussed in Sections 3.2.1 through 3.2.3 respectively.

3.2.1 Predictor Variables from the American Community Survey

The Census Bureau's ACS provides up-to-date estimates for many socioeconomic and demographic characteristics. Counties with more than 65,000 people have ACS estimates from a single year (the most recent year). Counties with between 20,000 and 65,000 people have ACS estimates aggregated from a three-year period (the most recent three years). Counties with less than 20,000 people have ACS estimates aggregated from the most recent five-year period. All of these estimates are direct estimates. Care must be taken in understanding the differences in one-year, three-year, and five-year estimates (that they refer to differing time periods), especially when combining or compositing areas with estimates at differing levels. This issue can be resolved by making all areas agree with the smallest area (e.g., if the smallest area has three-year estimates, then three-year estimates are used for every area whether or not a one-year estimate is available).

The ACS can provide percentages of families in poverty, percentages of homeowners, median household income, percentages by race/ethnicity groups, employment rates, education levels and percentages of foreign-born persons, which could be correlated with crime rates. Percentages of persons in particular age categories (e.g., percentage of persons older than 65 years of age, percentage of persons between 0 and 17 years old) may also be correlated to crime rates. Appendix D of the Task 3 Memo includes a listing of candidate demographic characteristics.

3.2.2 Predictor Variables from BJS Administrative Sources

The Bureau of Justice Statistics (http://bjs.ojp.usdoj.gov) provides a wide range of administrative statistics which may be of use in SAE models. Most of these are available at the state level only, except for those delivered from the Census of State and Local Law Enforcement Agencies. Some are available annually, some available biannually, and others available every four years. Statistics are mostly totals, but in all cases these can be converted into more useful per capita statistics, for the purpose of modeling, by dividing by population estimates. Some of the administrative statistics available are the following:

- Census of State and Local Law Enforcement Agencies:
  - Data available every four years (most recent is 2008);
  - Data available at the state and local level (individual law enforcement jurisdictions[13]);
  - Includes total number of sworn law enforcement officers.
- Justice Expenditure and Employment Extract Series:
  - Data available annually;
  - Data are ultimately from Census Bureau's Annual Government Finance Survey and Annual Survey of Public Employment;

[13] These jurisdictions include local police agencies, sheriffs, state police, and special jurisdictions. Local police agencies and sheriffs nest within counties, as well as some special jurisdictions. Others are at the state level (though some special jurisdictions cross states, such as port authorities).

- Includes total expenditures for law enforcement, judicial functions, and corrections.

- National Prisoners Statistics Program:
  - Data available biannually (most recent 2008);
  - Includes total number of prisoners in state departments of correction.
- Annual probation survey and annual parole survey:
  - Data available annually (most recent 2008);
  - Includes total number of adults on state or federal probation and parole in given year.

The magnitude of the correlation between these statistics and NCVS victimization rates is not known. However, it would be useful to compute these correlations to asses there utility for a SAE.

3.2.3 Predictor Variables from the Uniform Crime Report Program

Another source of potential predictor variables is the Uniform Crime Report (UCR) program of the Federal Bureau of Investigation. The UCR program collects data from U.S. law enforcement agencies on reported crimes for all of the crime types included in the NCVS. These data represent a census of the crimes reported to the police in the reporting agencies. The UCR data are potentially good predictors of crime victimization rates, with some important caveats, as given below:

- The UCR tracks only crimes which are reported to law enforcement agencies. It is an undercount of total crime events as measured by the NCVS.
- The UCR is a voluntary program. Some agencies do not report to it, others report only in some months and/or for some crime types.
- The quality of UCR data is known to vary considerably by jurisdiction. This is a serious limitation in using the UCR data as predictor variables for small area models.

The correlation between the UCR property crime rate and the NCVS property crime rate is high. From work conducted for Task 3, the correlation between the NCVS property crime rate and the reported UCR property crime rate at the MSA level (among 40 of the largest MSAs) is about 0.54, the highest correlation coefficient between the NCVS property crime rate and all the considered auxiliary variables. The correlation to the UCR rates is considerably lower for violent crimes such as rape/sexual assault, simple assault, and simple theft. Simple assault in particular exhibits a low correlation, in large part because the UCR does not compile data on the incidence of simple assault alone so any positive correlation would have to come from a strong relationship between simple assault and the serious violent crimes that the UCR collects (which may or may not be the case). See for example Fay and Li (2010). Thus the UCR may be a viable predictor for property crimes and some violent crimes, but not for simple assault. In addition, there are some jurisdictions where the poor coverage of the UCR will preclude its use.

3.3 Model Development

This section discusses approaches for developing small area models for use with the NCVS, estimating it, and validating it. Section 3.3.1 discusses unit-level vs. area-level approaches. Section 3.3.2 discusses benchmarking and validation. Section 3.3.3 discusses an extension of the modeling to time series models that take advantage of the longitudinal aspect of the NCVS. Univariate versus multivariate modeling is discussed in Section 3.3.4.

A broad issue in model development involves selecting a small set of predictor variables from a larger set of potential predictors (as given in Section 3.2), finding the correct functions of those predictors (linear, quadratic, a polycotomous categorization) and defining important interactions between the selected predictors. Stepwise regression procedures can be used for these purposes, and possibly tree search algorithms to find interactions.

Another issue is deciding on a final model for generating the indirect estimates (Empirical Bayes vs. Hierarchical Bayes), and specifying a methodology for generating the estimates (e.g., Markov Chain Monte Carlo). Rao 2003 provides a good reference for these important details.

3.3.1 Unit-level vs. Area-level Approach

As was briefly discussed in the beginning of Chapter 3, in an area-level model direct estimates produced at the local area level are the prime elements in the modeling process. One part of an area-level model is a ‘sampling model’, where survey-weighted estimates are produced for the small-areas with sample-design based variance estimates. The regression model is developed using predictors at the small-area level only. One can also distinguish between ‘matched’ and ‘unmatched’ models, where the former has the survey weighted estimate directly as the dependent variable in the model regression, and in the latter case, a functional transformation (e.g., the logit function) provides the link to the predictors.

In distinction to this the unit-level model is built at a much lower level such as individual persons or households. For example, the NSDUH program built up its estimates from the individual person level. There is no effort to generate sample-design unbiased estimates and certainly not sample-design based variance estimates at this very low level[14].

Some applications of small-area estimation in U.S. Federal agencies have been area-level approaches, such as the SAIPE Program described in Section 3.1.1. The small area application to NAAL described in Section 3.1.2 similarly took an area-level approach (see also Task 3 Memo). An example case of a unit-level approach is the NSDUH, as described in Section 3.1.3.

Either approach could be used for an NCVS small-area program. The area-level approach is more design-based, as it uses as its basic building blocks the sample-design based estimates at the targeted local level as well as the sample-design based variance estimates at this level. The unit-level approach is more dependent on the validity of the model, as it disaggregates down to the lowest levels. Sampling weights can be used to estimate the parameters of the model, as was done in the NSDUH application, which can make this portion of the estimation process sample-design consistent[15]. Variance estimates are entirely model dependent. Operationally, the area-level approach certainly works with a much simpler data set, with one record for each local area rather than one record for each household or person, and in that sense is easier to work with in practice. This is especially useful as the Bayesian methods require numerous iterations with the data set as an input in each iteration.

3.3.2 Model Validation and Benchmarking

Large-scale small-area estimation programs should employ extensive model validation. The models are never perfect, and systematic errors can manifest themselves. A good small-area estimation program should include both internal and external model validation. Internal model validation

---

[14] The NSDUH estimates did use the survey weights in estimating the parameters of the models, making these parameter estimates design consistent. The totals and the variance estimates were entirely model-based however.

[15] I.e., the estimates are approximately unbiased estimators of the corresponding parameters at the population level, over all possible samples, with this property not dependent on the validity of the model.

consists of checking on the model for its accuracy and robustness. Possible checks include the following (see Citro and Kalton (2000): SAIPE used all of these checks):

- Checks for linearity of the relationship between predictors and target variables;
- Checks of the distribution of residuals:
  - Absence of obvious nonrandom patterns when checked against the predictor variables;
  - Absence of outliers or other deviations from normality.
- Checks of homogeneity of the variance (when checked against the predictor variables).

A failure of these checks indicates a revision of the model might be necessary (for example, adding new predictors, or nonlinear functions of the old predictors such as quadratic and interaction terms).

The NAAL SAE program (Mohadjer et al. (2009)) used a number of other internal model-validating procedures. These include the following:

- Using different priors for the hyperparameters in the HB model to make sure that the model fit is not sensitive to the choice of the prior distribution;
- Using variable sets of predictor variables to make sure that the model fit is not sensitive to the particular set of predictor variables used.

External checks generally consist of comparing aggregations of indirect estimates to direct estimates for larger geographic areas for which reliable direct estimates are available, or to external control totals from other surveys or from administrative data.

Once the model has been finalized after the validation process, benchmarking can be included to bring the small area estimates into full agreement with external controls. This final benchmarking ideally should only be used to provide minor adjustments. The validation procedure should first assure that the model essentially provides agreement with external controls (if not perfectly), with the benchmarking then tidying up residual differences.

3.3.3 Time-Series Modeling

The NCVS has a rotating panel design that yields annual estimates. It may be beneficial to take advantage of this time series for the local area as well as the data gathered for the estimation year. A small area model can be constructed to use the estimates from previous years to help in the estimation of the current year. Rao (2003, in Section 5.4.3) discusses several variations of model structure and implementation that has been put into practice (see for example Rao and Yu (1994) and he suggests that considerable gains can be achieved by borrowing strength across both small areas and time. An appropriate small area model for the NCVS should give considerations to exploring the predictive power of time-series effects.

3.3.4 Multivariate vs. Univariate Models

An indirect estimation program for the NCVS is likely to include more than one target estimate. A limited indirect estimation program might include only percentages of persons victimized by violent crimes and of household victims by property crimes. A broader program could include personal theft, and could include a more detailed breakdown of violent crimes and property crimes respectively.

The SAIPE for example (see Section 3.1.1 for details) has multiple estimates at the state, county, and school district levels. The SAIPE program does not combine the models in a multivariate approach: each is modeled separately with no estimation of covariances between the residuals of the various models.

An example of a multivariate approach was a U.S. Department of Health and Human Services indirect estimation program at the state level for median income of four-person families. See Rao 2003, Section 5.4.1. Two direct estimates were computed from the Current Population Survey (CPS) for each state i: the median income of four-person families $\hat{\theta}_{i1}$, and a linear combination $\hat{\theta}_{i2}$ of the median income of three-person and five-person families (¼ fraction for three-person household CPS median income and ¾ fraction for five-person household CPS median income), for each state i. The 2x2 sampling covariance matrix for these two direct estimates was estimated from CPS and utilized in the indirect estimation process. The final product of this exercise was indirect estimates of true state-level median incomes θ_{i1} for the four-person households: the inclusion of the extra linear combination θ_{i2} was designed only to improve the precision of the four-person household estimates θ_{i1}.

3.4 Conclusions

In Chapter 3, we have explored the many possibilities for indirect estimation, based on precedents from other U.S. Federal surveys and current research on this topic. Indirect estimation can be done with the current NCVS sample as it is, or it can also utilize a supplementary sample (see Chapter 4 below). Predictor variables for an indirect estimation program can include American Community Survey data, Department of Justice administrative statistics, and also data from the supplementary sample. Efforts related to SAE for crime rates would require a thorough search for predictor variables that would likely include other possible data sources. The wide range of models possible allow for considerable flexibility.

Seemingly a low cost effort, the amount of work involved to develop the models would be non-negligible, and therefore the costs involved in producing indirect estimates would be weighed against the resulting gains in usefulness of producing local area crime data (via model-based estimation).

3.5 Next Steps

To implement a small area model to produce indirect estimates, the following steps should be taken:

1. Hold discussions to identify the local areas and key crime rates and characteristics for indirect estimates;
2. Hold discussions on the UCR variables, their differential measurement error across geographic areas, and the acceptability level as predictors by the data users;
3. Conduct an exhaustive compilation of potential predictors that could be used to estimate a model at the state, MSA or central city level;
4. Generate model predictions for the local areas;
5. Validate and evaluate the model.

To achieve a heightened use of the data, further research could be conducted on the formation of about 100 to 150 crime statistic areas (groups of counties within a state, or groups of small states), and evaluate the cost and benefits relating to precision improvements beyond direct estimates at the same geographic levels.

Adding Sample Using Alternative Methods 4

Chapter 2 discussed the ability of the NCVS to make direct estimates of adequate precision for different local areas with current sample sizes. The chapter also assessed the effect of adding approximately 13,000 interviews as part of the direct estimation. The results illustrate the capability and limitations of using the current design, supplemented by additional interviews, to generate estimates for other than the very largest areas of the U.S. In this chapter we explore two alternative uses of the additional funds that would support the extra 13,000 NCVS sample size, both of which involve the selection of a much larger, inexpensive, survey to supplement the core NCVS sample.

One option is to conduct a two-phase survey. The initial phase would use an inexpensive method (e.g., mail) to identify households with at least one victimization. The second phase would subsample households based on the responses to the first phase and follow-up with an interview using the national survey methodology. The aim of this option is to produce estimates from the two-phase survey that are directly comparable with those produced by the NCVS. The data from the two-phase survey would be combined with the national sample for the local area to provide an estimate via composite estimation or some other adjustment. The second option is to conduct a large-scale supplementary survey using a less expensive collection methodology, designed to produce crime rate estimates that are closely correlated, but not directly comparable, to those produced by the NCVS. The supplementary survey estimates would then be incorporated in a small area modeling procedure.

4.1 Two-Phase Design

An important objective of supplementing the NCVS is to minimize differences in methodology between the supplemental sample and the main NCVS interviews. The bigger the differences between the two, the more reliance the estimates will have on adjustments and some type of small area model. Using small area models, as described in section 4.2 below, proposes using an inexpensive method (e.g., telephone) for supplementation, but would also rely on a statistical model to combine it with the main NCVS interviews. An important advantage of the two-phase methodology is that it uses the full NCVS methodology as part of the second phase. The extent of the adjustments required when combining with the main NCVS will be minimized.

A second important advantage of the two-phase method is that it provides the capability to estimate characteristics related to victimization (e.g., relationship to perpetrator, location of event, consequences of event, reporting to police). While limited by sample sizes, it would be relatively straightforward to use supplemental data collected using national-level methods to generate these estimates.

The discussion below describes this design as a one-time supplement to the main data collection. This is not the only possible application of such a methodology. If a two-phase design were successful, one might consider using it as a substitute for the main NCVS methodology. One might also conceive of the survey as part of a rotating panel design, with the first phase serving as the initial contact. For purposes of exposition, the section below does not expand on these more elaborate designs. As will be noted later, the success of the two-phase approach depends on the ability of the first phase data collection to accurately identify households and persons that have experienced victimizations. If this methodology is of interest to pursue, elaborations of it can be developed once the quality of the information is assessed.

4.1.1 First Phase Survey

As noted above, the first phase survey would be completed using a relatively inexpensive method. By keeping the cost of this phase relatively low, a large number of households could be screened. One possibility would be to use a mail survey. There are other methods that might be used that are also significantly less expensive than an in-person survey, including a telephone interview or some combination of modes. For purposes of the discussion below, we assume a mail survey will be used.

Prior research on the NCVS has discussed using a mail survey (e.g., Biderman, 1981). It is one of the least expensive methods to use and potentially allows for screening a large number of households and individuals. This method is used by some local areas to conduct victimization surveys. The NCVS has not adopted a mail survey approach because this method of data collection is not suitable for a complex questionnaire like the NCVS and it is unlikely to achieve high response rates. However, when viewed in the context of a two-phase design, an initial phase that is conducted by mail may be an efficient way to collect data for supplementing the NCVS.

Using the mail to make initial contact with respondents is becoming much more common. This is reflected in recent studies indicating that a short mail survey, using a comprehensive address frame like the Delivery Sequence File (DSF)[16] and a combination of incentives and special mailings (e.g., priority mail), can achieve better response rates than random digit telephone surveys. For example, the National Household Education Survey (NHES) (Montaquila, et al., 2010) recently achieved a 60 percent response rate to a short screening survey. A second example is the use of a mail survey for recruitment into Gallup's Panel Survey which is replacing the use of random digit dial (RDD) telephone recruitment (Rao, et al., 2010).

The least expensive method would be to send a single questionnaire, asking one person to report on victimizations occurring against everyone 12+ in the household. Using a single respondent to report for the entire household maximizes the household response rate. A short instrument, resembling the current NCVS screener, would be developed to ask about household victimizations, as well as personal victimizations against each household member. There may also be a set of items that would be used to predict victimization risk (demographic items of others in the household; prior victimization experiences). [17]

This first phase survey could also include items asking for contact information that might be used for the second phase survey. For example, asking for a telephone number would provide a way to conduct future interviews over the telephone. Similarly, an e-mail address would provide a way to contact the household over the internet.

Based on responses to the first phase survey, households would be subsampled for the second phase. The sample would be drawn using three basic strata: 1) households reporting at least one victimization; 2) households not reporting any victimization; and 3) households not responding to the first phase survey. Different subsampling rates would be applied to each of these groups. The subsampling rates would be determined to maximize the precision of the estimates produced by the second phase sample.

Using a single respondent to report for the entire household is the simplest and least expensive method to implement a two-phase survey. It has the most promise for measuring household-level crimes, where the household informant would have knowledge of the events. For purposes of illustration, this type of design is further elaborated below. However, using a single informant to

---

[16] The DSF represents all of the residential addresses to which the U.S. Post Office delivers mail.

[17] Another variation of this method would be to ask respondents to fill out the screener on the internet. If they do not respond, then follow-up with a paper questionnaire. This may be less expensive, but research to date has found this to result in a lower response rate when compared to a pure paper survey approach (Messer, 2009).

report on personal victimizations is not likely to work as well. In the final section, issues and alternative methods that might be considered for focusing on personal crimes are discussed.

4.1.2 Second Phase Survey

The second phase survey would administer the entire NCVS interview to each member of the household. This would include both the screener and the detailed incident form. There are a number of variations of this design. The design discussed below uses an in-person interview for the second phase survey. This minimizes the differences between the supplemental interviews and the main NCVS.

4.1.3 Illustration of Methodology for Estimating Household Crimes

To illustrate the possible efficiencies of a two-phase design, this section provides estimates of sampling error and costs for estimating a property crime rate for households. This analysis is for illustration purposes and is conducted under the assumption of simple random sampling, whereas the true design would be more complex with clustering. For example, a clustered design would be necessary due to the follow-up of first phase nonrespondents. The analysis examines the possible benefits of the two-phase design for two different estimates of interest, namely the incidence rate and estimates of incident characteristics.

Optimizing the Precision of Estimates of Crime Incidence

The development of optimal estimates from a two-phase sample comes from Kalton (1993, Section 2.5). Appendix B presents modifications tailored to the NCVS application developed from Cochran (1977), Chapter 2. This illustration assumes there is a SRS of n' households selected at the first phase, with these households then allocated based on the results of this first phase screening data into three strata ($h=1$ for screened crime households; $h=2$ for screener nonrespondents, and $h=3$ for screener no crime households). The optimal sampling fraction f_h for stratum h is[18]:

$$f_h \propto S_h$$

[18] Derived from Equation 12.21 in Cochran 1977

where S_h^2 is the population variance of the number of incidents per household y_i in stratum h. If each household experienced at most one crime, then $S_h = \sqrt{P(1-P)}$, where P is the proportion of households experiencing a crime. However, in practice, S_h will be slightly larger than this to reflect the fact that a few households will experience more than one crime. Nelson (1980) found that a negative binomial distribution fit well to the empirical distribution of the number of crimes experienced. A negative binomial distribution, with mean 0.131 and exponent 0.468, has therefore been used in this illustration.

The first phase screening will be imperfect to some degree, with some households reporting experiencing a crime when in fact they did not (false positives) and some reporting not experiencing a crime when in fact they did (false negatives). Initially, we assume that 10 percent of those reporting a crime are false positives and 1 percent of those not reporting a crime are false negatives. Later we examine the effect of different rates of false negatives.

Table 8 displays the assumed distribution for y_i for screened-for-crime households (i.e., households reporting a crime at the first phase) based on the negative binomial distribution with mean 0.131 and exponent 0.468 (Nelson, 1980). The false positive probability of $y_i = 0$ is set equal to 10 percent, and the remaining probabilities for $y_i = 1$ through $y_i = 9$ are derived from the negative binomial probabilities conditional on $y_i > 0$, adjusted to equal 90 percent for these outcomes.

Table 8. Assumed distribution function for y_i for screened-for-crime households

Incidents	Percentage
0	10.00
1	78.22
2	9.63
3	1.74
4	0.33
5	0.07
6	0.01
7	0.00
8	0.00
9	0.00
Mean	1.04
Variance	0.308

Table 9 presents the assumptions about expected percentages of households experiencing at least one crime, incidence means, and variances within each stratum. It is assumed that 60 percent of households respond to the screener, and that 12.5 percent of the responding households (or 7.3 percent of all sampled households) indicate some victimization in the household.

For screener nonrespondents, the incidence rate $\bar{Y}_2$ is assumed to be 0.135 and the element variance S_2^2 is assumed to be 0.14 (slightly larger than the variance for a dichotomous zero to one random variable with expectation 0.131 to allow for a few household incidence counts greater than one within this stratum). Note that the screener nonrespondents are assumed to have about the same incidence rate than respondent households. There are a set of non-respondents that will have high rates of victimization. For example, highly mobile respondents have higher victimization rates for the in-person NCVS interviews. However, with a mail survey, one would also expect a higher response among those that view the topic as more salient (i.e., crime and victimization). For example, those that have recently been victimized would find the topic more salient than someone who has not had a recent victimization experience.

Households which indicate on the screener that no crimes have taken place are assumed to have a mean incidence (false negative) rate of 0.01, with variance 0.015 (slightly higher than for a dichotomous zero to one random variable with that expected value).

These assumptions are summarized in Table 9, along with the optimal rates f_h given these assumptions.

Table 9. Optimal second-phase sample rates for the two-phase design

	Percentage of first-phase sample W_h	Mean incidence rate $\bar{Y}_h$	Within-stratum variance S_h^2	Optimal sampling fraction f_h (as a percent)
Screened for crime	7.3	1.04	0.308	50.0
Screener nonrespondent	40.0	0.13	0.140	34.0
Screened for no crime	52.7	0.01	0.015	11.1
Total	100.0	0.14	0.086	

Assuming a first phase screening cost of \$75 per sampled household, a cost of \$770 per person in a household selected for the second phase, with all eligible persons over aged 12 being interviewed in a household selected for the second phase, and on average 1.73 eligible persons per household, and taking into account the sampling fractions in the last column of Table 9, the average cost per screened household in this two-phase design is \$383. With a budget of \$10 million, the first-phase sample size n' is thus 26,104. Table 10 below shows the first-phase and second-phase sample sizes under this design. A total of 6,037 households receive second-phase interviews, and 10,444 interviews are conducted.

Table 10. First- and second-phase sample sizes for optimal design

	Percentage of first-phase sample W_h	Optimal sampling fraction f_h (as a percent)	First-phase sample size n'_h	Second-phase sample size n_h	Second-phase interviews
Screened for crime	7.3	50.5	1,893	955	1,652
Screener nonrespondent	40.0	34.0	10,442	3,550	6,141
Screened for no crime	52.7	11.1	13,770	1,532	2,651
Total	100.0		26,104	6,037	10,444

The first panel of Table 11 below presents standard errors for this two-phase design and a single phase design with the same budget (i.e., using the second phase sample data collection approach as a single phase sample – essentially dropping the first phase). The table also gives the budget for a single phase design that would achieve the same standard error as the two-phase design. As shown in the first panel, the standard error for crime rate with the two-phase design is smaller than that for a single phase design with the same overall budget; in terms of variances, the two-phase design gives a gain in precision of about 35 percent. Thus a single phase design that accomplishes the same standard error as the two-phase design with a budget of $10 million would require a budget of $13.5 million.

Table 11. Summary of two-phase design and of one-phase design comparison for incidence rates

Description	Two-phase-design Phase 1	Two-phase-design Phase 2	Single phase design
Non-victim strata incidence = .01			
Sample size	26,104	10,444	13,000
Standard error of incidence mean	0.00297		0.00345
Cost for same precision ($mn)			13.5
Non-victim strata incidence = .035			
Sample size	20,944	10,947	13,000
Standard error of incidence mean	0.00327		0.00345
Cost for same precision ($mn)			11.0
Non-victim strata incidence = .05			
Sample size	18,198	11,214	13,000
Standard error of incidence mean	0.00345		0.00345
Cost for same precision ($mn)			10.0

The second and third panels of Table 11 provide similar estimates once assuming the screener does not classify as accurately the population into victim and non-victim groups. The middle panel assumes that the incidence rate of the non-victim strata is .035 (rather than .01). As can be seen, the efficiency drops significantly to where the standard errors are considerably closer. The third panel gives the case where the two methods yield equivalent precision.

Optimal Designs for Collecting Information on Incident Characteristics

A design that is optimal for estimating incident characteristics is different from one that is optimal for estimating incidence. Suppose the scenario is identical to that described in Section 4.1.3, but the focus is on estimating a prevalence P^* among incidents (for example, the proportion of incidents that are reported to the police). Let P_h^* be the incident characteristic prevalence in stratum h and assume that P_h^* is 50 percent for each stratum. The approximate optimal sampling fractions f_h for each stratum that maximize the precision of the prevalence estimate for a fixed budget are given in Table 12 (see Appendix B for details).

Table 12. Optimal second-phase sample rates for collecting crime victimization information

	Percent of first-phase sample W_h	Mean incidence $\bar{Y}_h$	Percentage of victims A_h	Optimal sampling factor f_h (as a percent)	First-phase sample size	Second-phase sample size	Second-phase crime victims
Screened for crime	7.3	1.04	56.1	100.0	1,398	1,398	1,459
Screener nonrespondent	40.0	0.13	40.0	48.0	7,712	3,702	500
Screened for no crime	52.7	0.01	3.9	13.0	10,171	1,322	13
Total	100.0	0.13	100.0		19,281	6,422	1,972

The major difference in results from estimating incidence is that the optimal second-phase sampling factor for the screened-for-crime households stratum is 50 percent compared to 100 percent in for incident characteristics. For estimating prevalence of a characteristic it is best to interview all in the screened for crime stratum. As a result, the first-phase sample size is 19,281 households rather than the 26,104 households in Table 10. These results demonstrate that there are different optimal designs for two different types of estimates. A decision will therefore need to be made as to whether to focus the design on one type of estimate or whether to find a compromise design that serves both types of estimate reasonably well.

Table 13 summarizes the standard errors and costs for the prevalence estimator for the two-phase and single phase design for three different assumptions on the false negative rate. Focusing on the first panel, the standard error of the prevalence estimate is about 24 percent smaller with two-phase design (i.e., the variance of the estimate is about 55 percent smaller). In terms of costs, the single phase design would be about 50 percent more expensive to achieve the same level of precision. The difference between the two designs decreases as the assumption on the false negative rate increases. If one assumes an incidence rate as high as .05 in the non-victim strata, the two designs differ by around 10 percent.

Table 13. Summary of one-phase and two-phase comparison relating to incident characteristics

Description	Two-phase design	One-phase design
Non-victim strata incidence = .01		
Standard error of incidence mean	0.0129	0.0157
Cost for same precision ($mn.)	$10.0	$15.1
Non-victim strata incidence = .035		
Standard error of incidence mean	0.0142	0.0157
Cost for same precision ($mn.)	$10.0	$12.1
Non-victim strata incidence = .05		
Standard error of incidence mean	0.0148	0.0157
Cost for same precision ($mn.)	$10.0	$11.3

4.1.4 Combining Estimates

One straightforward method for combining the data collected in the main NCVS and the supplemental survey would be to produce separate estimates from each survey and combine them using a composite estimator, similar in form to that discussed in Chapter 3 for the SAIPE program. Thus the composite estimate $\hat{p}$ would be a linear combination of main NCVS estimate $\hat{p}_1$ and the estimate from the supplementary survey $\hat{p}_2$:

$$\hat{p} = \alpha \hat{p}_1 + (1-\alpha)\hat{p}_2$$

where the quantity $\alpha = \hat{\sigma}_2^2 / (\hat{\sigma}_1^2 + \hat{\sigma}_2^2)$ is an estimate of the relative precision of $\hat{p}_2$ as compared with $\hat{p}_1$, based on the estimated variances of each. This form of composite estimator is appropriate when the estimates from the two surveys are both approximately unbiased for the same population parameter. In the current case, careful consideration needs to be given to possible sources of difference between the two estimates beyond sampling error. For example, under the current NCVS design, one would suspect that combining the two estimates would need to consider:

1. Time-in-sample. The main NCVS survey would be fully balanced with respect to the average rotation group. The supplemental survey would be unbounded and have inflated rates relative to the main NCVS. To make the estimates comparable, the supplementary sample estimate would need to be adjusted as, for instance, is currently done with the NCVS wave 1 data.

2. Non-response bias. The second phase of the supplemental survey would likely have a lower response rate than the NCVS. To the extent response is correlated with victimization, the rates would be different from the main NCVS. A thorough non-response adjustment would need to be conducted in order to reduce the potential bias to a level comparable to the NCVS in order to allow the simple compositing approach.

If a telephone survey was used at the second phase of the supplementary survey, there may also be differences due to mode. Hubble and Wilder (1988) found that CATI interviews lead to more reports of victimization (see also Cantor and Lynch, 2005).[19]

The above differences pertain to the current NCVS. They may not apply with an NCVS that is redesigned to incorporate different features, such as revised screener, a 12 month reference period and other modes of interviewing (e.g., ACASI).

4.1.5 Issues When Measuring Personal Crimes

There are several complicating issues for a two-phase design when moving from household to person level victimizations. One is that the process related to identifying victimized individuals for the second phase is not as simple as for household crimes. If a single individual within a household is reported by a household informant as being a personal crime victim in the first phase, one would have to decide whether it is worth interviewing the entire household (as with the normal NCVS) or interviewing the person identified by the household informant. Logistically it would be easier to interview every person. If this is done, then a more elaborate stratification design would be needed to distinguish between households that are expected to yield 1) no crimes; 2) a household crime; 3) a personal crime; and 4) both types of crimes.

Perhaps a more important issue is that of using an informant to report on personal crimes for the entire household. The NCVS is designed under the assumption that it is necessary to interview every eligible individual to accurately collect information on personal victimizations. Research conducted during the development of the NCVS in the early 1970's found a household informant can provide

[19] It isn't clear if this effect also applies to the current methodology of administering CAPI interviews from the interviewer's home.

fairly accurate information for crimes that are likely to be known to the entire household (e.g., "household theft"[20]; Burglary, Motor Vehicle Theft), but underreported personal crimes (Dodge and Turner, 1981; Kalish, 1981). An experiment was conducted that compared rates of reporting for a single, household informant design to one that interviewed each individual within the household. The results indicated that the single respondent design produced significantly lower rates of personal victimizations. The ratio of multiple vs. single respondent households reports of robbery among household members ranged from 1.18 to 2.2, depending on the type of event. Similarly for assault, this ratio ranged from 1.24 to 1.73.

The standard of accuracy of the initial phase for a two-phase method is not as high as for the main NCVS. The second phase of the procedure uses more accurate methods. Nonetheless, the efficiency of the two-phase methodology depends on being able to use the first phase to accurately classify households. As shown above, the less accurate this phase is, the less effective the methodology will be.

One option to using a single household informant in the first phase is to send multiple first-phase surveys to the same household and ask all eligible individuals to return a survey. This procedure has been used for general population surveys (Battaglia, et al., 2008; Cantor, et al., 2008). This introduces additional non-response within multiple person households which would have to be followed up during the second phase.

4.1.6 Next Steps

The success of a two-phase methodology depends on the extent victimizations are captured in the first phase. If the stratification based on the first phase is not accurate, it will not be efficient. The example for estimating incidence rates for household crimes illustrated the effects of varying levels of false negative reports on the screener. As false negatives increase, the efficiency of the two-phase methodology goes down.

To move forward with this methodology, there are some preliminary analyses that could be done to assess feasibility. For example, one might analyze current NCVS data to look at the correlation between the responses to the NCVS screener filled out by the household respondent and the amount of household crime reported once using the full NCVS data-set.

[20] Theft that occurs on household property.

It may also be worth searching for other victimization surveys that have used a single respondent design. For example, as noted above, the early NCVS studies did some experimentation with using a single respondent (Kalish, 1981). There may be other surveys, perhaps done at a local level, which also might inform the design of a single respondent to proxy for the entire household.

To provide a more concrete idea of the requirements of a two-phase design, it would be useful to compute the efficiencies with different assumptions about the design of the first phase survey (e.g., single vs. multiple respondents) and the ability to identify particular types of victimizations (e.g., personal vs. household crimes). This information could be used to further assess which methodology might be most promising.

If the two-phase method seems feasible, a series of pilot tests should be conducted to assess the approach. The pilot tests would seek to address questions such as:

1. How accurately can the first phase identify households that contain victims of different types?
2. What would be the sampling parameters used to conduct the second phase survey?
3. How comparable are the final estimates to the main NCVS methodology?

An initial pilot could test the basic mail methodology, its response rate and its ability to get household and/or person-level measures of victimization. If this approach is promising, subsequent pilots would carry through with both phases of the design.

4.2 A Supplementary Survey

In this section, we discuss uses of a supplementary survey that collects the information using a relatively inexpensive method in a single phase. Because the methodology relies on a small area model, it is not necessary that the estimators generated from the supplementary survey be unbiased, as the methodology can adjust for the bias under certain conditions (see Section 4.2.3 below). This provides freedom to keep the instrument and mode of administration inexpensive, allowing for larger sample sizes. Sections 4.2.1 and 4.2.2 present two alternatives along these lines, and Section 4.2.3 discusses how the indirect estimation program might incorporate the supplementary survey.

4.2.1 A Supplementary Mail Survey

As discussed in Section 4.1 above, a mail survey can be an inexpensive, but effective way of administering a survey. It has its drawbacks as a final stand-alone instrument (as opposed to being the initial screening phase in a two-phase approach, as discussed in Section 4.1). The mail questionnaire must be limited in its scope, without complex skip patterns.

A mail survey could be restricted to this mode or one could follow up with a telephone contact for initial mail nonrespondents. Telephone numbers would be available for the roughly 50 percent of households for which a telephone number can be associated with the address.

4.2.2 A Supplementary Telephone Survey

A second option is a telephone survey, selecting telephone numbers from landline and cell phone exchange frames. This option has the drawback of not covering non-telephone households, and has the complexity of dealing with landline vs. cell phone issues. However, it allows for a more extensive instrument with more complicated skip patterns. For example, one could randomly select one person within the household to be the targeted individual of the questionnaire, and the full NCVS interview could be asked of that person. Alternatively, the entire household could be interviewed.

There are some options regarding putting together the landline and cell phone samples:

1. All households retained from landline and cell phone samples;
2. Cell phone-only households retained from cell phone sample only and all landline households retained;
3. Cell phone-only or cell phone-mostly households retained from cell phone sample; cell phone-mostly households dropped from landline sample.

Option 1 is the most efficient in retaining the most households, but generates complications in weighting. Option 2 simplifies weighting, but requires considerable screening-out from the cell phone frame sample. Option 3 requires more detailed screening questions, but allows for greater retention from the cell phone sample. The cell-mostly households are those which have a landline telephone, but almost never use it (making response through the landline frame very problematic).

4.2.3 Combining the Main Survey and the Supplementary Survey

Regardless of how supplementary sample is drawn, an important question is how the local area information it produces is used in making small area estimates. Each survey would provide separate estimates for small areas which are covered by both surveys. The estimates from the supplementary survey would not be directly comparable with those from the NCVS but, to be effective in the modeling, the underlying parameters should be correlated with the NCVS parameters.

One natural approach for borrowing strength from the supplementary sample would be to use the supplementary survey as predictors in an area level small area model, of the type discussed in Chapter 3. However, in such a use, the sampling error of the supplementary survey small area estimates has to be taken into account. Otherwise the small area regression model estimates will be biased because of the variable error in one of the predictor variables. Ybarra and Lohr (2008) have recently developed methods for taking variable errors in a predictor in a small area model into account and these methods could be applied in this case.

A different approach is to model the NCVS and supplementary survey small area parameters jointly in a small area model, incorporating an allowance for a correlation between the parameters as a way to borrow strength from making the NCVS small area model-dependent estimates. Raghunathan et al. (2007) provide an example of this approach, with the National Health Interview Survey (NHIS) as the main survey and the Behavior Risk Factor Surveillance Survey (BRFSS) as the supplementary survey. The National Health Interview Survey is an in-person survey that collects data in only a set of sampled PSUs while the BRFSS is a large scale telephone survey with sample cases in nearly all counties in every state, but covering only telephone households. Raghunathan et al. (2007) develop a single small area model that jointly estimates three separate county parameters using a set of county-level auxiliary variables as predictors: an NHIS telephone household parameter, an NHIS non-telephone household parameter, and a BRFSS telephone household parameter. The county estimates for the NHIS telephone and non-telephone parameters are then combined to form the overall county estimates. A Hierarchical Bayes methodology was used for the estimation. See Appendix A for a more technical description of this type of model.

4.2.4 Next Steps

The success of the supplemental sample in SAE models will depend on the variation explained by the supplemental sample estimates beyond the variation already explained by readily available information (refer to the covariates listed in Chapter 3).

As an initial investigation, correlations could be computed between existing screener data (serving as the supplemental sample data from a mail survey), NCVS crime rates and other covariates for different subgroups to gain insights as to the predictive power of a supplementary screener beyond what is gained by including other variables as predictors.

A simulation or sensitivity study could be conducted to better understand the impact of:

1. the supplemental sample design on the prediction model; and
2. the correlation structure on the prediction model

The results would inform decisions on whether or not to pursue the supplemental sample/SAE approach, and if pursued, how the supplemental sample should be designed.

A possible simulation study could involve the large MSAs (or large states). Because the screener estimates and the NCVS estimates should not come from the same sample, each MSA could be randomly split into two samples, one for the screener estimates and one for the NCVS estimates. To better understand the relationship between the number of areas to select, and the number of persons to select within the areas, the study sample could be subset in various ways: 1) large sample in many areas; 2) small sample in many areas; 3) large sample in few areas; and 4) small sample in few areas.

While the correlation structure between the NCVS screener results, NCVS crime rates, and other covariates would be informative, it would not replicate the correlation structure involving the supplemental survey results. The NCVS screener results could be modified to provide a range of correlations with the other variables. Each simulated dataset could be processed through a small area model to test the impact on the precision of the model predictions, with and without the NCVS screener results as predictors. This would help determine the benefits of a supplemental sample with similar correlation structure with other covariates and the NCVS crime rates.

A pilot study would be necessary to produce a correlation structure at the area-level between the supplemental estimates, NCVS crime rates and other covariates. Bonet and Wright (2000) provide sample sizes that are needed to compute correlations in general. If the correlation is about 0.9, then the sample size could be less than 30 if you allow for a relatively wide confidence interval (such as 0.3). The sample size grows to over one thousand as the correlation decreases. The correlation structure could be mapped back to the simulation study results to interpolate the impact on the precision of the model predictions. The correlation structure would also help to inform how many areas and how much sample to select in each area.

Depending upon the results of the above analysis, and depending upon the results of the pilot study below, to incorporate the supplemental sample into an SAE model, the modeling approaches discussed above would need to be developed. This could use existing data. The development of these models would be completed if supplemental data (e.g., from a pilot study) are highly correlated with crime rates.

If an SAE approach is considered promising, a series of pilot studies would be conducted with the goal of identifying the best data collection approach for the supplemental sample. The objectives would include the following:

1. To develop and test a screener that could be used as the supplemental survey;
2. To compare to an RDD survey that interviewed entire household or a randomly selected respondent;
3. To use the results to compute correlations between supplemental survey crime rates, UCR, and NCVS crime rates, and using the correlation to assess the additional predictive power through interpolation using the results from the non-pilot work; and
4. To develop modeling programs toward generating predictions using the supplemental data and predictors.

The pilot studies should be conducted in areas where the current sample is also being collected. This would allow for developing and evaluating the statistical models. The large MSAs are prime candidates.

References

Bankier, M.D. (1988). Power allocations: Determining sample sizes for subnational areas. *American Statistician,* 42, 174-177.

Battaglia, M.P., Link, M.W., Frankel, M.R., Osborn, L., and Mokdad, A.H. (2008). An evaluation of respondent selection methods for household mail surveys. *Public Opinion Quarterly,* 72, 459-469.

Biderman, A.D. (1981). Notes on the methodological development of the National Crime Survey (1970-73). In R.G. Lehnen and W.G. Skogan (Eds.), *The National Crime Survey: Working papers. Volume 1: Current and historical perspectives* (NCJ-75374; pp. 30-33). Washington, DC: Bureau of Justice Statistics, U.S. Department of Justice.

Bolling, K., Grant, C., and Donovan, J. (2008). *2007-08 British Crime Survey (England and Wales). Technical report volume I* (pp. 1-37). Available at http://www.homeoffice.gov.uk/rds/pdfs08/bcs0708tech1.pdf.

Bonett, D., and Wright, T. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Biometrika,* 65(1), 23-28.

Bureau of Justice Statistics, U.S. Department of Justice. (2002). *National Crime Victimization Survey: Criminal victimization, 2001* (Bureau of Justice Statistics Bulletin NCJ 194610, written by Callie Rennison). Washington, DC: Author.

Bureau of Justice Statistics, U.S. Department of Justice. (2004). *National Crime Victimization Survey: Criminal victimization, 2003* (Bureau of Justice Statistics Bulletin NCJ 205455, written by Michael R. Rand). Washington, DC: Author.

Bureau of Justice Statistics, U.S. Department of Justice. (2007). *National Crime Victimization Survey: MSA data, 1979-2004* [Computer file]. Conducted by the Census Bureau, U.S. Department of Commerce. ICPSR04576-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor]. doi:10.3886/ICPSR04576.

Bureau of Justice Statistics, U.S. Department of Justice. (2009). *National Crime Victimization Survey: Criminal victimization, 2008* (Bureau of Justice Statistics Bulletin NCJ 227777, written by Michael R. Rand). Washington, DC: Author.

Cantor, D., and Lynch, J. (2005). Exploring the effects of changes in design on the analytic uses of the NCVS data. *Journal of Quantitative Criminology*, 21, 293-320.

Cantor, D., Sigman, R., and Han, D. (2008, May). *Pilot of a mail survey for the Health Information National Trends Survey.* Paper presented at the annual meeting of the American Association for Public Opinion Research, New Orleans, LA.

Citro, C.F., and Kalton, G. (2000). *Small-area income and poverty estimates: Priorities for 2000 and beyond.* Washington, DC: National Academy Press.

Cochran, W.G. (1977). *Sampling techniques,* 3rd ed. New York: John Wiley & Sons.

Dodge, R.W., and Turner, A.G. (1981). Methodological foundations for establishing a national survey of victimization. In R.G. Lehnen and W.G. Skogan (Eds.), *The National Crime Survey: Working papers. Volume 1: Current and historical perspectives* (NCJ-75374; pp. 2-9). Washington, DC: Bureau of Justice Statistics, U.S. Department of Justice.

Elliot, M.R., and Davis, W.W. (2005). Obtaining cancer risk factor prevalence estimates in small areas: Combining data from two surveys. *Applied Statistics,* 54(3), 595-609.

Fay, R., and Li, J. (2010). *Study of the association between UCR and NCVS measures at the county level* (Report prepared under contract to Bureau of Justice Statistics). Rockville, MD: Westat.

Folsom, R., Shah, B., and Vaish, A. (1999). Substance abuse in states: A methodological report on model-based estimates from the 1994-1996 National Household Surveys on Drug Abuse. *Proceedings of the American Statistical Association Section on Survey Research Methods*, pp. 371-375.

Hanson, M., Hurwitz, W., and Madow, W. (1953). *Sample survey methods and theory, Volume I.* New York: John Wiley & Sons.

Hubble, D., and Wilder, B.E. (1988). Preliminary results from the National Crime Survey CATI experiment. *Proceedings of the American Statistical Association Survey Research Methods Section.*

Hughes, A., Sathe, N., and Spagnola, K. (2008). *State estimates of substance use from the 2005-2006 National Surveys on Drug Use and Health* (DHHS Publication No. SMA 08-4311, NSDUH Series H-33). Rockville, MD: Office of Applied Studies, Substance Abuse and Mental Health Services Administration. Available at: http://www.oas.samhsa.gov/2k6state/AppA.pdf.

Hughes, T.R., and Griffin, D. (2010, March). *Managing a mixed-mode survey: The American Community Survey's experience.* Paper presented at the Federal CASIC Workshops, Washington, D.C.

Kalish, C.B. (1981). The Dayton-San Jose methods test. In R.G. Lehnen and W.G. Skogan (Eds.), *The National Crime Survey: Working papers. Volume 1: Current and historical perspectives* (NCJ-75374; pp. 28-29). Washington, DC: Bureau of Justice Statistics, U.S. Department of Justice.

Kalton, G. (1993). *National Household Survey Capability Programme: Sampling rare and elusive populations.* New York, U.S.: United Nations Department for Economic and Social Information and Policy Analysis.

Kish, L. (1965). *Survey sampling.* New York: John Wiley & Sons.

Krenzke, T., and Cantor, D. (2009, January 5). *NCVS Task 2: Review of surveys while assessing the feasibility of local area estimates* (draft memorandum from Tom Krenzke and David Cantor of Westat to Michael Rand and Erika Harrell of the Bureau of Justice Statistics). Rockville, MD: Westat.

Krenzke, T., Li, J., and Cantor, D. (2009, June 2). *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.

Link, M.W., Battaglia, M.P., Frankel, M.R., Osborn, L., and Mokdad, A.H. (2008). A comparison of address-based sampling (ABS) versus random-digit dialing (RDD) for general population surveys. *Public Opinion Quarterly,* 72(1), 6-27.

Messer, B.L. (2009, May). *Improving survey response in mail and Internet general public surveys using address-based sampling and mail contact procedures.* Master's thesis, Washington State University.

Mohadjer, L., Kalton, G., Krenzke, T., Liu, B., Van de Kerckhove, W., Li, L., Sherman, D., Dillman, J., Rao, J.N.K., and White, S. (2009). *National Assessment of Adult Literacy indirect county and state estimates of the percentage of adults at the lowest literacy level for 1992 and 2003* (NCES 2009-482). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Montaquila, J.M., Brick, J.M., Hagedorn, M.C., and Williams, D. (2010, May). *Maximizing response in a two-phase survey with mail as the primary mode.* Paper to be presented at the Annual Meeting of the American Association for Public Opinion Research, Chicago, IL.

National Research Council of the National Academies. (2008). *Surveying victims: Options for conducting the National Crime Victimization Survey.* Washington, DC: Panel to Review the Programs of the Bureau of Justice Statistics (R.M. Groves and D.L. Cork, Eds.).

Nelson, J.F. (1980). Multiple victimization in American cities: A statistical analysis of rare events. *American Journal of Sociology,* 85(4), 872.

Office of Management and Budget. (2009). *Update of statistical area definitions and guidance on their uses* (OMB Bulletin No. 10-02). Available at: http://www.whitehouse.gov/omb/assets/bulletins/b10-02.pdf.

Raghunathan, T.E., Xie, D., Schenker, N., Parsons, V.L., Davis, W.W., Dodd, K.W., and Feuer, E.J. (2007). Combining information from two surveys to estimate county-level prevalence rates of cancer risk factors and screening. *Journal of the American Statistical Association,* 102, 474-486.

Rao, J.N.K. (2003). *Small area estimation* (Wiley Series in Survey Methodology). New York: Wiley.

Rao, J.N.K., and Yu, M. (1994). Small area estimation by combining time series and cross-sectional data. *Canadian Journal of Statistics,* 22, 511-528.

Rao, K., Kaminska, O., and McCutcheon, A.L. (2010). Recruiting probability samples for a multi-mode research panel with Internet and mail components. *Public Opinion Quarterly,* 74, 68-84.

Shah, B.V., Barnwell, B.G., Folsom, R., and Vaish, A. (2000). Design consistent small area estimates using Gibbs algorithm for logistic models. *Proceedings of the American Statistical Association Section on Survey Research Methods*, pp. 105-111.

Wolter, K.M. (1985). *Introduction to variance estimation.* New York: Springer-Verlag.

Ybarra, L.M.R., and Lohr, S.L. (2008). Small area estimation when auxiliary information is measured with error. *Biometrika*, 95(4), 919-931.

You, Y., and Rao, J.N.K. (2002). Small area estimation using unmatched sampling and linking models. *Canadian Journal of Statistics,* 30(1), 3-13.

Appendix A
Technical Description of the Raghunathan et al. Small Area Model

A

This section provides an overview of the area-level model developed by Raghunathan et al. (2007) for the combination of the Behavior Risk Factor Surveillance Survey (telephone survey covering every county in every state in the country) and National Health Interview Survey (in-person survey covering sampled PSUs only). BRFSS covers every county, but only telephone households. NHIS covers only selected PSUs, but covers both telephone and nontelephone households. Write p_{xj} as the percentage in county j (having some characteristic of interest) estimated within telephone households from the in-person survey, write p_{yj} as the same percentage in county j estimated from non-telephone households from the in-person survey, and p_{zj} as the percentage in county j within telephone households from the telephone survey.

Raghunathan et al. (2007) transform these percentages using an arcsin square root transformation (this transformation has variances which are approximately equal across differing values of the parameter):

$$\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} = \begin{pmatrix} \sin^{-1}\left(\sqrt{p_{xj}}\right) \\ \sin^{-1}\left(\sqrt{p_{yj}}\right) \\ \sin^{-1}\left(\sqrt{p_{zj}}\right) \end{pmatrix}$$

The model for the transformed sample percentages is parameterized as follows:

$$\begin{pmatrix} x_j \\ y_j \\ z_j \end{pmatrix} \sim N_3\left[\begin{pmatrix} \theta_j \\ \varphi_j \\ \left(1+\delta_j\right)\theta_j \end{pmatrix}, S\right]$$

$$S = \frac{1}{4}\begin{bmatrix} \tilde{n}_{xj}^{-1} & \rho_{xy}\tilde{n}_{xj}^{-1/2}\tilde{n}_{yj}^{-1/2} & 0 \\ \rho_{xy}\tilde{n}_{xj}^{-1/2}\tilde{n}_{yj}^{-1/2} & \tilde{n}_{yj}^{-1} & 0 \\ 0 & 0 & \tilde{n}_{zj}^{-1} \end{bmatrix}$$

where N_3 is a trivariate normal distribution, $\tilde{n}_{xj}$, $\tilde{n}_{yj}$ and $\tilde{n}_{zj}$ are the effective sample sizes for the three percentages, and ρ_{xy} is the correlation between the non-telephone and telephone household estimates from NHIS. Note that the parameter δ_j measures the difference between the in-person and the telephone survey estimates for the telephone households. It contains a mode effect and a differential response effect.

The second part of the model then is as follows:

$$\begin{pmatrix} \theta_j \\ \varphi_j \\ \delta_j \end{pmatrix} \sim N_3\left(\beta U_j, \Sigma\right)$$

U_j is a vector of predictors at the county level. This part of the model allows a link to the predictor variables.

One modification of this approach would be to compute the sampling variance matrix S computing the sampling variances of the vector of transformed percentages directly, and then smoothing it using a special smoothing model, as was done for a univariate untransformed percentage for the small area program for the National Assessment of Adult Literacy survey (Mohadjer et al. (2009)). Each of the three sampling variances (of x_j, y_j, and z_j) and the sampling correlation between x_j and y_j would be smoothed under this approach. As in NAAL, the smoothed variances from this special model would be then used as fixed variances in the following model-fitting step. The smoothing model for NAAL included as predictor variables the sample size and the percentage. Using the arcsin-square root transformation, the percentage would not need to be a predictor (because of the variance stabilizing transformation).

Appendix B
Optimizing Estimate for Crime Incidence and Incident Characteristics

B

Optimizing Estimates of Crime Incidence

The development in this appendix comes from Kalton (1993), Section 2.5, with modifications tailored to the NCVS application developed from Cochran (1977), Chapter 2. The application assumes there is a SRS of n' households selected at the first phase, with these households then allocated based on the results of this first phase screening process into H strata. In the NCVS application, $H = 3$ strata ($h = 1$ for screened crime households; $h = 2$ for screener nonrespondents, and $h = 3$ for screener no crime households). The sample units allocated to each strata by the first-phase screening are n'_h, with first-phase proportions $w_h = n'_h / n'$; where w_h has population expectation W_h. Using a sampling rate f_h, we subsample $n_h = f_h n'_h$ second-phase household units in each stratum to determine numbers of incidents in this subsample for each stratum. An average of $\bar{m}$ persons is interviewed in each household (we assume $\bar{m}$ is 1.78 across the three strata).

The population characteristic considered in this section is incidents of thefts reported during the last twelve months. The count for each household y_i can be 0, 1, 2, …. The population value estimated is the population mean $\bar{Y}$ over all households. $\bar{Y}$ is assumed to be 0.14 in the development in this section (corresponding roughly to recent U.S. theft rates). The within-stratum mean values are $\bar{Y}_h$. S_h^2 is defined to be the population variance across households of y_i, the total variance within each stratum h, and

$$S^2 = \sum W_h S_h^2 + \sum W_h \left(\bar{Y}_h - \bar{Y}\right)^2 = S_w^2 + S_b^2$$

The second-phase sample household-level mean estimate for stratum h is $\bar{y}_h$. The overall mean incidence estimate after both phases is $\bar{y} = \sum w_h \bar{y}_h$, and an approximate variance of $\bar{y}$ is[21]

$$V\left(\bar{y}\right) \approx \sum \frac{w_h s_h^2}{n' f_h} + \frac{1}{n'} \sum W_h \left(\bar{Y}_h - \bar{Y}\right)^2$$

[21] This is equation 12.14 from Cochran 1977 (with some simplification by dropping finite population correction terms).

A linear cost function can be utilized:

$$C = n'\left(c' + \sum c\bar{m}W_h f_h\right)$$

where c' is the cost of a completed screener, c is the cost of a completed second-phase (in-person) interview, and $\bar{m}$ is the expected number of second-phase interviews per household. Overhead costs are left out. With these formulas for cost and variance, the optimal sampling fractions are[22]:

$$f_h = S_h\sqrt{\frac{c'}{c\bar{m}S_b^2}}$$

Optimal Designs for Collecting Information on Incident Characteristics

A design that is optimal for estimating incident characteristics is different from estimating incidence. Suppose the scenario is identical to that described in Section 4.1.3, but the focus is on estimating a prevalence P^* among incidents (for example, the proportion of incidents that are reported to the police). As given in Kalton (1993) an important parameter here is the proportion of incidents in each stratum A_h, with

$$A_h = \frac{\bar{Y}_h n'_h}{\sum \bar{Y}_h n'_h} = \frac{\bar{Y}_h W_h n'}{\sum \bar{Y}_h W_h n'}$$

The parameters $\bar{Y}_h$, W_h, n'_h and n' are defined in Section 4.1.3. Table 10 presents the values of A_h under the assumptions of Section 4.1.3. The incident characteristics prevalence is:

$$P^* = \sum A_h P_h^*$$

where P_h^* is the incident characteristic prevalence in stratum h. The estimator of this from the second-phase sample is:

$$p^* = \sum A_h p_h^*$$

[22] Derived from Equation 12.21 in Cochran 1977

Where p_h^* is the estimator of the crime victimization prevalence in stratum h obtained from second-phase interviews from victims in subsampled households. The sample size for this within each stratum is m_h, with expected value

$$E\left(m_h\right) = \bar{Y}_h n_h = \bar{Y}_h f_h n_h'$$

The approximate variance of p^* is

$$v\left(p^*\right) = \sum A_h^2 \frac{P_h^*\left(1 - P_h^*\right)}{E\left(m_h\right)}$$

The cost function is the same as given above. The values of f_h for each stratum that minimize cost times variance were found numerically. The optimizing formula found through calculus has too much complex nonlinearity to be useable. The solution is accurate to the nearest 2.5%.