# State-Level Design-Based Estimates for National Surveys

**Stephen Ash and Brian Shaffer**
U.S. Census Bureau
4600 Silver Hill Road
Washington D.C. 20233

**David Swanson**
Bureau of Labor Statistics
2 Massachusetts Ave NE
Washington D.C. 20212

**Abstract**

National surveys are being asked to do more with their national samples. In addition to providing national estimates, many national surveys are providing different sub-national estimates including state- and metropolitan area-level estimates. Small-area estimation is sometimes used as a last resort to produce sub-national estimates but many surveys are also using design-based methods in creative ways with their national sample designs to produce sub-national estimates.

This paper will give an overview of how several U.S. Census Bureau surveys including the American Housing Survey, National Crime Victimization Survey, and the Consumer Expenditures Survey use different methods within the design-base framework to produce sub-national estimates. We will review and compare the variety of sample designs and weighting methods for the aforementioned surveys. We will also discuss methods for evaluating the resultant estimates to determine whether they are a reasonable set of state-level estimates.

## 1. Introduction

National surveys are being asked to do more with their national samples. In addition to providing national estimates, many national surveys are providing different sub-national estimates including state- and metropolitan area-level estimates. Small-area estimation is sometimes used as a last resort to produce sub-national estimates but many surveys are also finding low-hanging fruit within the national sample designs. By low hanging fruit, we mean states or metropolitan areas within the national sample design that already have large sample sizes. Some of surveys are using their samples and weighting methods in creative ways with their national sample designs to produce sub-national estimates. We include examples of different surveys and how they have addressed this challenge. To begin, we consider two basic questions:

**Q1. What are all of the ways that we use to improve the estimation of state-level statistics?** We will not suggest new methods but will discuss how different surveys have applied a variety of methods to improve their sampling strategy (Särndal *et al*. 1992; p. 30) – the combination of the sample design and estimator (weights). The general areas we will consider include:

- Stratification of first-stage sample
- Adding or pooling sample
- Nonresponse adjustments
- Ratio adjustments
- Variance estimation

**Q2. How do we determine that the resultant state-level estimates are reasonable?** Even after doing everything we can to improve our estimation methods, some states will still cannot produce *reasonable* estimates. We consider both how we can define what is a reasonable and how can we determine whether estimates are reasonable. The latter is a difficult question with no clear or easy answer in some cases but we will suggest some tools that will assist with the decision.

This paper will give an overview of several U.S. Census Bureau surveys and show how they have found creative ways to address our two basic questions. The surveys we consider include:

- American Housing Survey (AHS)
- National Crime Victimization Survey (NCVS)
- Consumer Expenditures Survey (CE)

We focus on national surveys that have a two-stage sample design and whose first priority is producing national estimates. For these surveys, producing state-level estimates was often not an original priority. However, each survey is working towards using their national sample to produce weights that support state-level estimates.

## 2. National Surveys with State Weights

In this section, we provide a brief overview of the three national surveys that are part of our discussion. The overviews of each survey are intended to only provide enough detail for the subsequent discussions of the paper and the reader should consult the referenced publically-available survey documentation for a more complete description of the surveys.

### 2.1 American Housing Survey (AHS)

The AHS provides a comprehensive set of data about a variety of aspects of housing in the U.S. The survey is sponsored by the Department of Housing and Urban Development and conducted by the Census Bureau. AHS is also a longitudinal survey where each round of the survey is conducted every two years with the latest cohort starting in 2015. Although AHS is a national survey, the last redesign of 2015 also incorporated features that facilitate state- and CBSA-level estimates within the national sample design. The sample sizes for the largest 15 Core Based Statistical Areas (CBSAs) were defined so that reasonable CBSA-level estimates could be produced with each round of AHS. Additionally, AHS identified five states – CA, FL, NY, PA, TX – that that have sufficient sample sizes given the national sample allocation and the addition of the sample for the largest 15 CBSAs.

Along with the national survey, AHS also conducts the same survey in additional metropolitan areas that change in each round. The metropolitan areas samples are independent of the national sample and are designed with sample sizes that can support metropolitan-level estimates. In 2015, AHS included 10 metropolitan samples and in 2017, AHS included a different set of 10 metropolitan areas.

### 2.2 National Crime Victimization Survey (NCVS)

The NCVS is one of the nation's primary sources of information on criminal victimization, both reported and not reported to police. The survey is sponsored by the Bureau of Justice Statistics and is conducted by the Census Bureau. NCVS is a monthly panel survey where a sample household is in the survey for seven rounds. Each round of interviewing occurs every six months and the recall period for respondents is the prior six months. See www.bjs.gov for more information about NCVS.

### 2.3 Consumer Expenditures Surveys (CE)

The Consumer Expenditure Surveys (CE) collects data on the expenditures made by households in the U.S. Its primary customer is the Consumer Price Index (CPI) which uses the data to generate weights for the various goods and services in its inflation index, and it is sponsored by the Bureau of Labor Statistics.

The CE survey consists of two sub-surveys, the CE Quarterly Interview Survey (CEQ) which collects data on large and recurring expenditures such as automobiles, major appliances, rent, and utilities; and the CE Diary Survey (CED) which collects data on small, frequently purchased items such as food and apparel.

The CEQ is a rotating panel survey in which a panel of households is interviewed every 3 months for 4 consecutive quarters, after which the panel of households is dropped from the survey and a new panel of households replaces it. It rotates in the sense that one-fourth of the households are new to the survey each quarter. By contrast, the CED is a non-panel survey in which households are asked to keep a diary of their expenditures for two weeks, after which they are dropped from the survey.

The CE's latest sample redesign started in 2015, and for that a random sample of core-based statistical areas (CBSAs) was drawn from each of the nine Census divisions. Many CBSAs cross state lines, which was not an issue when the sample was drawn since the plan was for the nine Census divisions to be the smallest level of geography for which CE would publish its data. However, the plan changed and now CE wants to publish expenditure estimates in states that have a sufficient amount of sample (e.g., NY, NJ, FL, TX, CA).

## 3. Methods for improving the estimation of state-level estimates

### 3.1 Stratification of first-stage sample design

The large national surveys we consider all use a two-stage sample design where counties are grouped into first-stage units called Primary Sampling Units (PSUs). The PSUs are then designated as Self-Representing (SR) or Non-Self-Representing (NSR), where a PSU is SR if it represents itself and no other PSUs. This means that SR PSUs are in a first-stage stratum by themselves and the NSR PSUs are in a stratum with other PSUs and the sample PSU of the stratum represents the other PSUs of the same stratum. Within each stratum, SR PSUs are selected with certainty (with a probability of selection of 1.0) and NSR PSUs are selected with probability proportional to size. Generally, the SR PSUs are the largest metropolitan areas of the U.S.

Historically, the three large national surveys described in this paper formed strata within either the four Census Regions or later within the nine Census Divisions. However, more recently some have been stratifying their NSR PSUs within state boundaries to facilitate current or future state estimates. Forming strata within each state is a way of reducing the variance of state estimates. Partitioning a state into several strata produces smaller variances for the state-level estimates than partitioning a division or a region into several strata and then backing into the state-level estimates through a post-stratification technique, as it is always better to optimize something directly than indirectly. Forming strata within states will likely result in some loss of precision for the national estimates, but most surveys choose the small loss at the national level in favor of a large gain at the state level.

### 3.2 Pooling sample

Usually the main problem with producing state estimates is that the sample sizes of the national sample that happen to be allocated to a state are too small. The national sample size is sufficient for producing national estimates, but the allocation of the national sample may not be large enough for most states.

For some surveys where state estimates are a very high priority, the surveys have ensured that the states all have sample sizes that can support state estimates. Other surveys have not addressed state-level sample sizes directly, but have improved the sample sizes of states by pooling sample in creative ways. NCVS has identified up to 24 states that they considered as able or being close to being able to support state estimates (Cantor *et al.* 2010). Some of the states require additional sample and to do this NCVS is considering supplemental sample and pooling sample from multiple years of the survey. Combining the sample over multiple years will change the meaning of the victimization rates as compared with the national estimates – the average victimization rate over five years versus an annual victimization rate – but they are able to produce estimates for a larger set of states. Fay and Li (2012) also discuss the preliminary evaluation of states considered by NCVS.

AHS has identified five states (CA, FL, NY, PA, TX) within their national sample that have sample sizes that can support state-level estimates. Additionally, AHS combines the sample of the national sample with the metropolitan area samples for additional states – in 2015 the additional states included OH and CO. Although the Metropolitan Area samples for each round of AHS change, AHS is able to produce additional state-level estimates, even if they are only for the given round.

### 3.3 Nonresponse adjustment

There are several methods that can be used to adjust sample weights for noninterviews including the weighting cell method, propensity modeling or logistic regression, and more recent methods as proposed by Lunstrom and Särndal (1999), Kott and Chang (2008), and others. For national surveys, the weighting cell method is often used due to its simplicity. The issue with any of these methods is whether the adjustment should be calculated within each state and only using sample from the given state or is it permissible or even advantageous to apply the adjustment across states when we are trying to improve state estimates? As with almost all decisions concerning national surveys, the answer concerns the priorities of the given survey. We consider two, sometimes competing, priorities that need to be considered: bias reduction and consistency of state estimates.

We suggest if reducing the bias due to noninterviews is the highest priority, then the noninterview adjustment should be applied across states and this way the adjustment will use the largest pool of sample units possible and potentially have the greatest impact on reducing bias. The reason this is true is that noninterivew adjustments can only account for a finite number of factors in the adjustment. By factors, we mean the variables used in the models or the variables used to define cells with the weighting cell method. If there are other factors that are more associated with response/nonresponse and/or the variable of interest for the survey than state, then including the “state” factor pushes out another factor that is better at reducing bias.

If consistency of state estimates and especially longitudinal consistency is a higher priority than bias reduction, then a survey may choose to apply the noninterview adjustment within the given state. This could happen when the sample size and/or the composition of the national sample fluctuates over time and thereby could impact the state-level noninterview adjustment differently over time.

### 3.4 Adjustments to known totals

Ratio adjustments, calibration, and raking can all be used to adjust weights using known totals and reduce the variance of the resultant estimates. With the development of weights for state estimates, the natural thing to do is to adjust the weights for known state-level totals which will do two things: make the state-level estimates consistent with the known totals and reduce the variance and possibly the bias of the resultant estimators.

We discuss two different adjustments to known totals that are often used together with national surveys: first-stage adjustments and overall adjustments. See also Estevao and Särdnal (2002) and Ash (2003) for further discussions of calibration for two-phase sample designs. The surveys that we consider all use these two ratio adjustments: the first addresses the first-stage sample design and the second ratio adjustment addresses the overall sample design that includes the first and second stage.

#### 3.4.1 First-stage Ratio Adjustments.

With a two-stage sample design, a first-stage ratio adjustment is often used to reduce the variability due to the first-stage sample design. There is sometimes disagreement whether this is needed given that an overall ratio adjustment is also subsequently applied. Our justification of a first-phase adjustment is the same as Deville and Särndal’s (1992) reasoning for calibration in general – “[o]ur objective is to derive new weights that modify as little as possible the original sampling weights $d_k = w_k^{-1}$, which have the desirable property of yielding unbiased estimates; the survey statistician wants to stay close to those weights.” Calibration or ratio adjustments should not have to do a great deal of work to improve or “correct” the weights. If the ratio adjustments are changing the weights a great deal due to the differences in the estimated totals being so different than the known totals, it’s probably a symptom of another problem with the sample design or the weights. We suggest that the variability of the first-stage sample design can be reduced specifically with the first-stage ratio adjustment so that the overall ratio adjustment must work less to modify or improve the weights.

With state estimates, the natural thing is to adjust to state known totals where known totals are available and actually known or there is a good reason to believe that they are of a higher quality than the estimated totals. Using national AHS as an example, ratio adjustments applied to the AHS have been considered as a means to ensure that the housing unit and HUD housing unit estimates derived from the weights are consistent with national, Census Division, and the five states CA, FL, PA, NY and TX. There is no need for AHS to adjust the largest 15 CBSAs because they are all in SR PSUs.

#### 3.4.2 Overall Ratio Adjustments.

After applying the first-stage ratio adjustment, and reducing the variability due to the first stage, the more common and almost expected adjustment is to a set of known overall totals. For national surveys, the last step of calculating the weights if often to adjust the weights to a set of known national totals. Often, these totals are provided at sub-national geographic levels such as region, division, and state. Figure 2 of section 4 provides an example of how the overall ratio adjustments can reduce the variance of the resultant estimates.

For AHS, the way the overall ratio adjustment is applied to the national sample is key to how the national weights also function as state weights. First, the known totals are produced and ratio adjustment cells are formed at the intersection of Census Division, the largest 15 CBSAs, and the five states. For example, the NY CBSA has one set of control totals, the balance of NY state (after removing the counties of the NY CBSA) has a set of totals, and the balance of Mid Atlantic division (PA and NJ) had another set of totals. The control totals include total housing units, HUD housing units, new construction, and demographic characteristics. Raking to separate division, state, and CBSA totals in separate rakes was considered but was ruled out since it would have increased in the overall number of rakes by a factor of 3. When the raking is applied, the result is that the sum of the weights is equal to known totals for all of the geographic levels of interest: Census division, the largest 15 CBSAs, and the five states.

Both NCVS and CE are considering state weights which are ratio adjusted to known state-level totals and are separate from the national weights. An issue CE found is that some of the cells are so small that they have no respondents at the state level, such as black people over 75 years old, which causes the calibration process to fail due to division by zero. The problem was dealt with by combining the independent estimates into fewer cells.

### 3.5 Variance estimation

Most of the surveys that are part of our discussion use a combination of Balanced Repeated Replication (BRR) and Successive Difference Replication (SDR). See Wolter (1995) for a review of BRR and Fay and Train (1992) and Ash (2014) for a review of SDR. Both of these methods work well with national surveys. BRR is applied to sample units in the NSR PSUs and SDR is applied to the sample units in the SR PSUs – both methods are applied with a single set of replicate weights that can be used to estimate the overall variance of an estimator. When applied to state estimates, there may be problems since a state estimate may use a small number of strata which translates to a small number of degrees of freedom. At this time, we have no suggestions for improving the low number of degrees of freedom (the variance of the variance estimator) but we do suggest that more work is needed in this area.

## 4. How to determine whether the state-level estimates are reasonable

Given that we have done everything we can do improve our state estimates, the question we need to ask is: are the estimates reasonable. To answer this question, different surveys will have different criteria based on the different surveys' goals and priorities. We will address two ends of the spectrum – survey's with specific goals and surveys with general goals with respect to reliability of estimates – and we will spend more time on discussing the problem of how to address surveys with general goals.

**Surveys with specific reliability goals**. Some surveys are designed to produce one specific estimate or at most a finite set of key estimates. An example is the Current Population Survey which can be used to produce many different estimates but the key focus of the Current Population Survey sample design is the estimation of the unemployment rate. Likewise, the NCVS can be used to produce estimates about many different facets of victimization, but the highest priority of the survey are the estimates of personal and property crime rates. These surveys and others have a well-defined goals that can be examined straightforwardly – either direct calculations of variances or simulation can be used to examine the expected variances and necessary sample sizes. This is often not a simple task itself, but it is simplified by only needing to consider one estimate or a finite set of estimates.

**General purpose surveys.** Some surveys do not have a key estimate that can serve as the central focus of the sample design. An example is the AHS which can be used to produce estimates about many different aspects of housing in the U.S. Data users that are interested in mortgages will consider mortgage estimates as key estimates and likewise data users that are interested in home improvement will consider their estimates as key. AHS collects information about a wide variety of topics on housing and is used by a wide variety of researchers that often focus on specific topics or conduct research that includes multiple topics. So the question for a general set of estimates is: how do we determine that a given state can do a reasonable job of producing a general set of estimates? We suggest some tools that can be used to examine a large set of estimates. First, we suggest using Generalized Variance Functions and then we suggest some transformations of GVFs that look at sample sizes.

### 4.1 General Variance Functions

We have found that GVFs are both useful for comparing different aspects of a sampling strategy. For example, with GVFs we can vary the sample design by changing its sample size, its allocation method, or its selection method, or we can vary the estimator by changing some of the factors used to construct its weights, and it is easy to see how those changes will affect the survey's variances. Although GVFs can be useful for such analyses, we find that they are often misunderstood and for this reason, we will explain GVFs both in general and as we apply them to examining state estimates. See Wolter (1985; chapter 5) for additional background on GVFs.

A GVF is simply a regression model describing the relationship between of a set of point estimates and their variances. The point estimates are the independent variables on the x-axis of such a graph, the variances are the dependent variables on the y-axis of such a graph, and the regression line shows the relationship between these two quantities. For example, based on the equation $\hat{V}(\hat{x}) = E(\hat{x}^2) - [E(\hat{x})]^2$ one might model the variance of a survey estimate with the regression line

$$\hat{V}(\hat{x}) = a\hat{x}^2, \text{ or}$$
$$\hat{V}(\hat{x}) = a\hat{x}^2 + b\hat{x}, \text{ or}$$
$$\hat{V}(\hat{x}) = a\hat{x}^2 + b\hat{x} + c.$$

Whatever model is used, we suggest that the estimates should estimate the same quantity. For example, the point estimates may be counts of the number of households, housing units, or people in a population, but they should not be mixtures of those different types of quantities.

A common GVF in household surveys shows the relationship between the estimated number of households in various population sub-domains and their variances. In these situations the estimated number of households in a domain is $\hat{N}_d = \sum_{k \in s_d} w_k$, where $s_d$ is the sample in domain $d$, $k$ indexes the sample units in $s_d$, and $w_k$ is the sample weight for unit $k$. The GVF's graph in Figure 1 shows $\hat{N}_d$ on the $x$-axis and $\hat{v}(\hat{N}_d)$ on the $y$-axis. Then the GVF is a regression line showing $\hat{v}(\hat{N}_d)$ as a function of $\hat{N}_d$ which can be interpreted as the expected value of $\hat{v}(\hat{N}_d)$ given the value of $\hat{N}_d$.

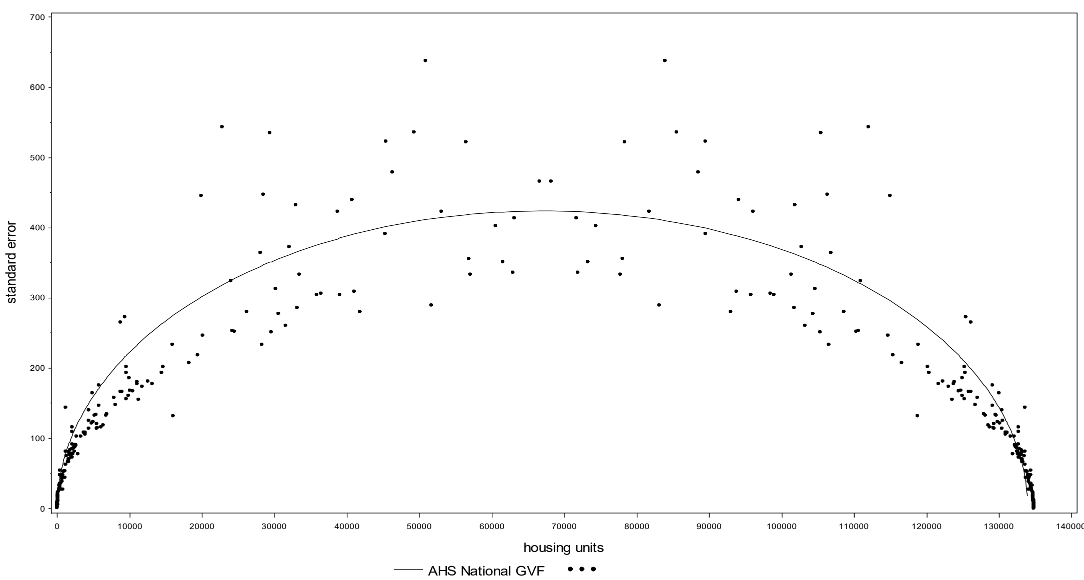


**Figure 1. Example of a GVF**

The Current Population Survey and other surveys use the model (U.S. Census Bureau 2006; p. 14-4):

$$\hat{v}(\hat{N}_d) = a\hat{N}_d^2 + b\hat{N}_d \tag{1}$$

while NCVS uses this model for its GVF (BJS 2017; p. 45) (Krenske 1995):

$$\hat{v}(\hat{N}_d) = a\hat{N}_d^2 + b\hat{N}_d^{3/2} + c\hat{N}_d.$$

Most GVFs do not include an intercept term since it is often assumed that the regression line passes through the origin. That is, $\hat{v}(\hat{N}_d) = 0$ when $\hat{N}_d = 0$. Sometimes it is also assumed that the variance is zero when the whole population is in the domain. That is, $\hat{v}(\hat{N}_d) = 0$ when $\hat{N}_d = N$. We know the variance is zero at $\hat{N}_d = N$ since we ratio adjust the weights of any sample from the set of all possible samples with the known total $N$ and therefore the estimate of the total of the subpopulation is always equal to the known total $N$. With model (1), if $\hat{v}(\hat{N}_d) = 0$ when $\hat{N}_d = N$ we know that $a = -b/N$ which allows us to simplify the model as $\hat{v}(\hat{N}_d) = b(\hat{N}_d - \hat{N}_d^2/N) = bN(\hat{N}_d/N)(1 - \hat{N}_d/N)$, which is similar to the variance formula for a binomial distribution.

An important point is that the GVF should use a large set of estimates and the set of estimates should cover the entire range of values that can be produced. For example, with the AHS we have a GVF modeling housing unit estimates for the state of New York so the GVF should include estimates from the smallest possible estimate of zero to the largest possible estimate – the number of housing units in the state of New York. Because we consider AHS a general purpose survey, we want the whole range of values from the smallest to the largest possible estimate to be "represented" in the model because we want the GVF to cover all of the possible estimates that a data user could produce. In addition, we know of no criteria for the number of estimates, but know that more is better. For AHS, we have used 200+ estimates.

Figure 2 presents the overall GVF of Texas from the 2015 AHS. The horizontal axis represents estimates of $\hat{N}_d$ or for AHS, the number of housing units for different domains $d$ and the vertical axis represents estimates of the standard error $\widehat{se}(\hat{Y}_d) = \sqrt{\hat{v}(\hat{Y}_d)}$. The range of the estimates starts at zero or the smallest number of estimated housing units possible to the total number of housing units in Texas for 2015. In between are different estimates that are representative of the entire range of estimates. The vertical axis represents the standard error of the estimates.

Figure 2 is an example of how GVFs can be used to compare different estimators while using the same sample. The set of red and the set of blue points each represent the same set of estimates but they differ by the sample weight used in their estimation. The blue line is the GVF for the variance derived from the sample weight that is adjusted for noninterviews but the overall ratio adjustment was not applied. The red line is the GVF model for the variance derived from the sample weight that is adjusted for both noninterviews and an overall ratio adjustment that used known totals of housing units in Texas. So in Figure 2, there is a blue point and a red point that each represent the estimate of the number of housings units with 1-2 bedrooms. The blue estimate/point uses the weight that is not adjusted with an overall ratio adjustment and the red estimate/point uses the weight that is adjusted with an overall ratio adjustment.

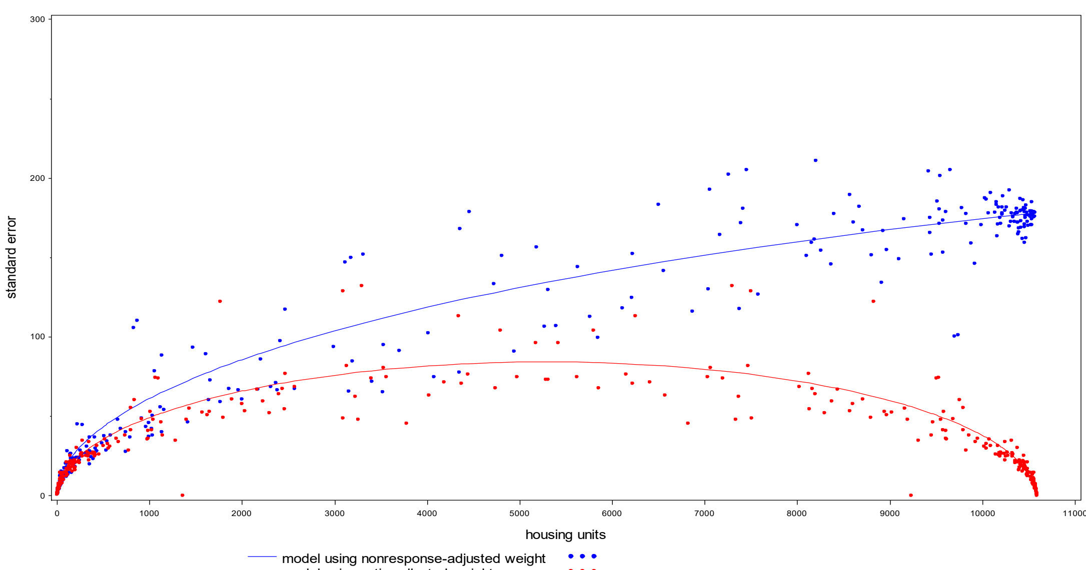


**Figure 2. GVF of a Good State (Texas) from the American Housing Survey**

Although it is difficult to quantify the variance reduction shown by the difference in the red and the blue lines of Figure 2, we can see how the overall ratio adjustment in terms of reducing the variance of the estimates.

GVFs can also be developed for other types of estimates. A different GVF could be produced for estimates of the total household population using the estimators of the form $\hat{Y}_d = \sum_{k \in s_d} w_k y_k$ , where $y_k$ is the number of persons in unit $k$. The graph then has $\hat{Y}_d$ on the horizontal axis and the standard error of the estimate or $\widehat{se}(\hat{Y}_d)$ on the vertical axis. Figure 3 presents a GVF of this type for the household population estimates of Texas. Here $y_k$ is the reported number of people in a sample housing unit. Again the blue and the red lines represent the estimates derived from weights that do not include and do include the overall ratio adjustment, respectively.

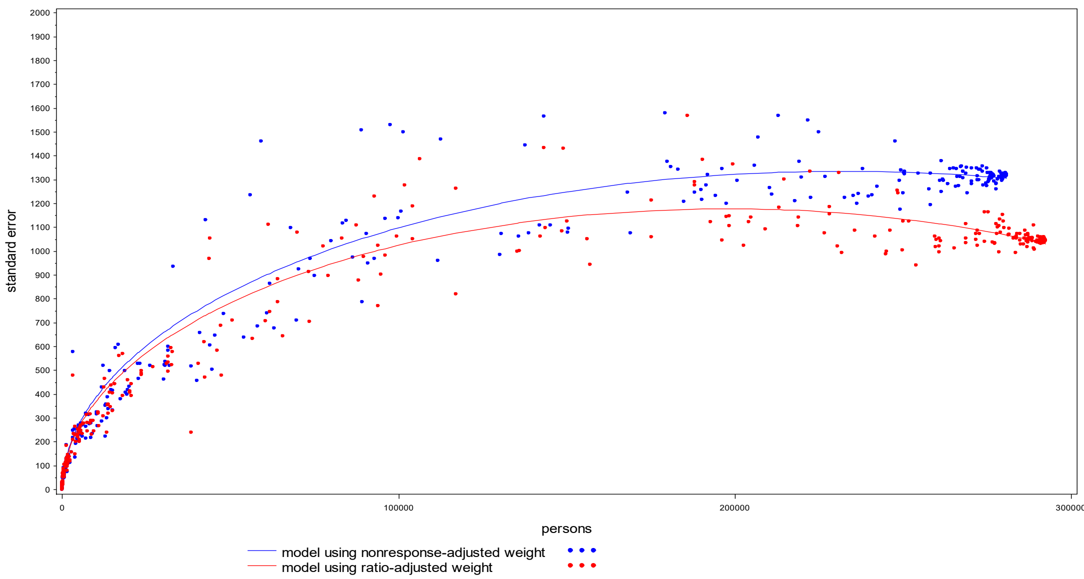


**Figure 3. GVF of Person Estimates (Texas) from AHS**

### 4.2 Sample size versus coefficient of variation

To further inform the sample design and estimation represented in a GVF, we suggest two methods that transform the GVF into a new graph of the sample size and the coefficient of variation. The new graph is useful since we can determine the average CVs of a set of estimates with respect to their average sample size. The first method transforms the model of the GVF itself and the second method transforms the points that were used to model the GVF and then models the CV graph.

**Method 1: Transform the model**. We can transform the modeled variance into the CV as $\hat{v}(\hat{N}_d) \Rightarrow \widehat{cv}(\hat{N}_d) = \sqrt{\hat{v}(\hat{N}_d)}/\hat{N}_d$ and transform the estimated total into the sample size as $\hat{N}_d \Rightarrow \bar{\hat{n}}_d = \hat{N}_d / \bar{\hat{w}}$ where $\bar{\hat{w}} = \sum_{k \in s} w_k / \sum_{k \in s} 1$. With these transformations of the usual GVF, we can produce a plot of $\left(\bar{\hat{n}}_d, \widehat{cv}(\hat{N}_d)\right)$ – the CV by the average sample size $\bar{\bar{n}}_d$ that will produce the given CV. Figure 4 is an example of a state with a reasonable sample size that can produce reasonable estimates for a set of general estimates.

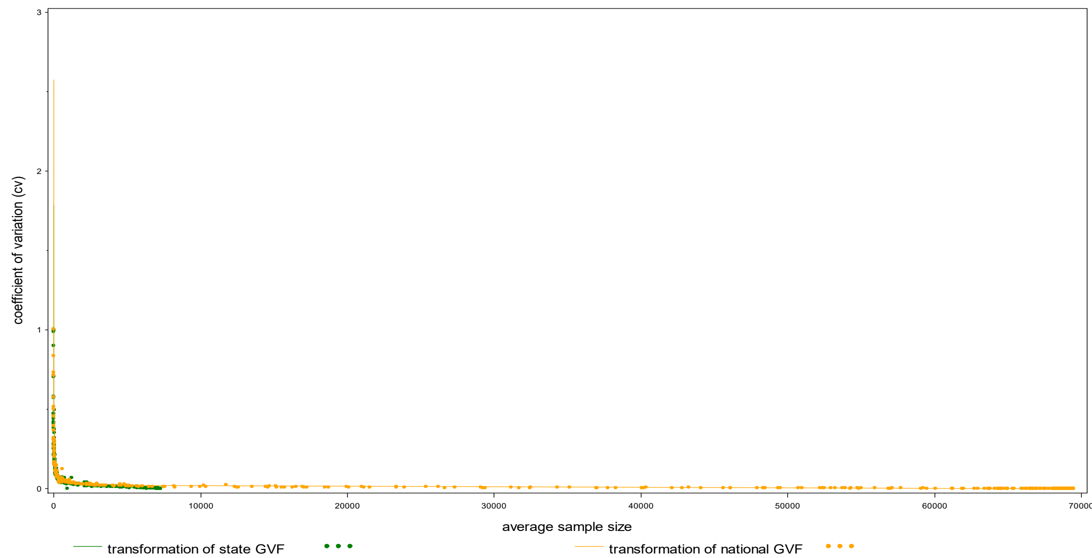


**Figure 4. Sample Size versus Coefficient of Variance for a Good State (Texas)**

In Figure 4, the green line represents the state and the yellow line represents the analogous national estimates. The national estimate is included as a reference and is generally considered as a standard for the state estimates. Unlike with the GVF, we can put the national and state estimates on one graph since the CV's of both estimates are on the same scale. With the GVF, the national and state estimates are usually not comparable due to the differences in their magnitude.

Although Figure 4 is accurate, the overall picture is not very useful. We see that as the sample sizes get large, the CVs get smaller (as expected) and in the other direction, as the sample sizes approach zero, the CVs become very large (as expected). What is interesting about Figure 4 is where and how fast the CVs increase with small sample sizes. Figure 5 is a subset of Figure 4, specifically with sample sizes less than 1,000.

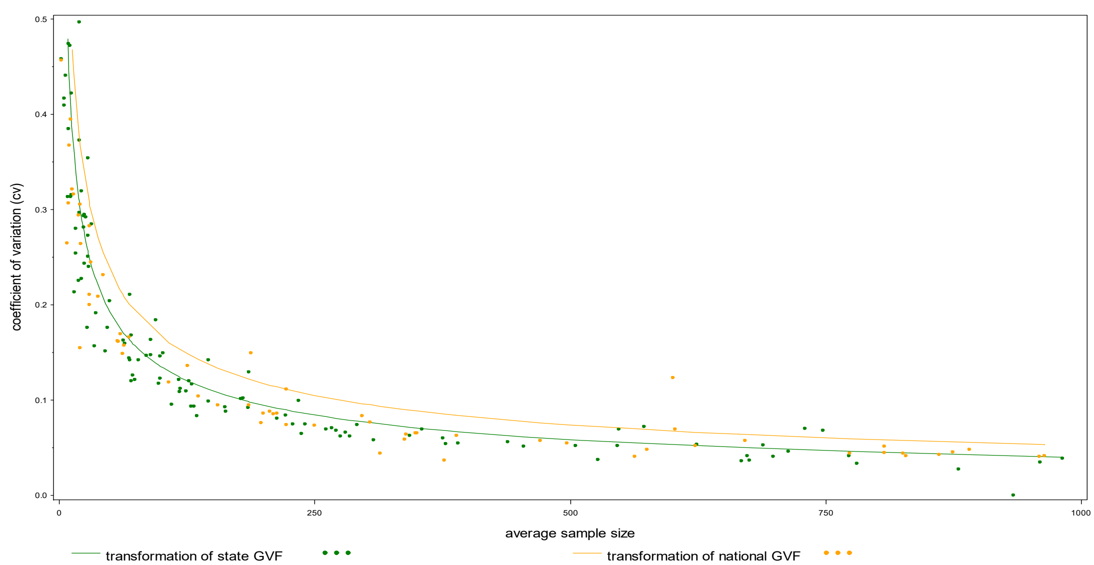


**Figure 5. Sample Size versus Coefficient of Variance for Texas Subset**

We now apply method 1 to evaluate two candidate states. In Figure 6, we can compare the CV graph of the two set of state estimates with the CV graph of the national estimates. The CVs of California (the red line) require less sample size than the national (the yellow line) for a given CV, and the CVs of Georgia (the green line) require more sample size than the national.

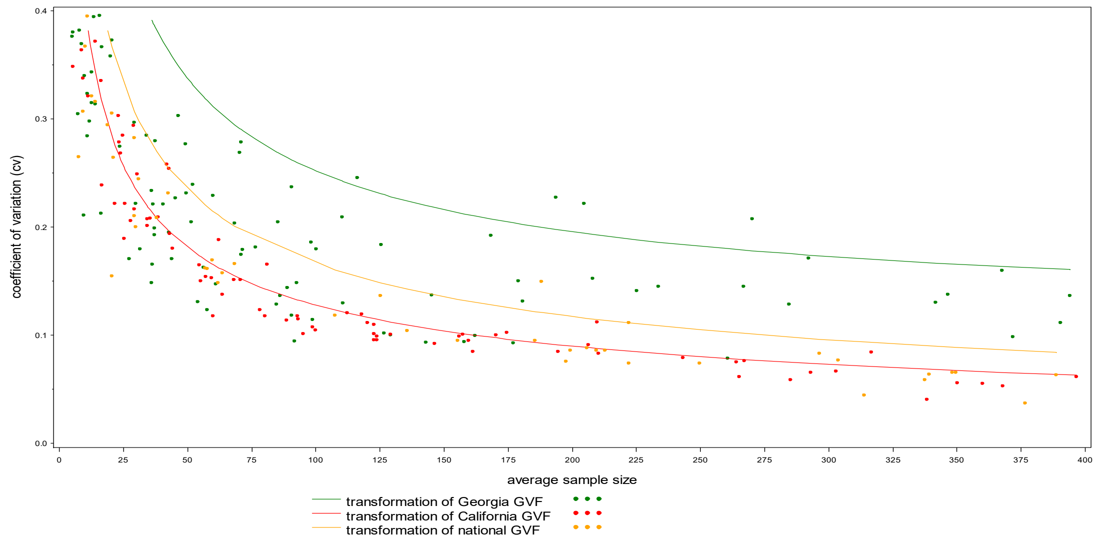


**Figure 6. Sample Size versus Coefficient of Variance for Two States, Method 1**

**Method 2: Transform the points**. We can transform the individual estimates of the variance into a $\widehat{cv}(\widehat{N}_d)$, directly obtain the actual sample size $\hat{n}_d = \sum_{k \in s_d} 1$, and model the points $\left(\hat{n}_d, \widehat{cv}(\widehat{N}_d)\right)$ with a new model that is separate from the model used with the GVF. For our application, we fit the model $\widehat{cv}(\widehat{N}_d) = a + b\hat{n}_d^{-\frac{1}{2}} + c\hat{n}_d^{-1} + d\hat{n}_d^{-\frac{3}{2}}$. The sample size $\hat{n}_d$ is represented on the horizontal axis and $\widehat{cv}(\widehat{N}_d)$ on the vertical axis. As with the first method, the modeled line represents the average sample size needed to produce a given CV. Figure 7 displays the same graph as Figure 6, but we modeled CV as a function of sample size.

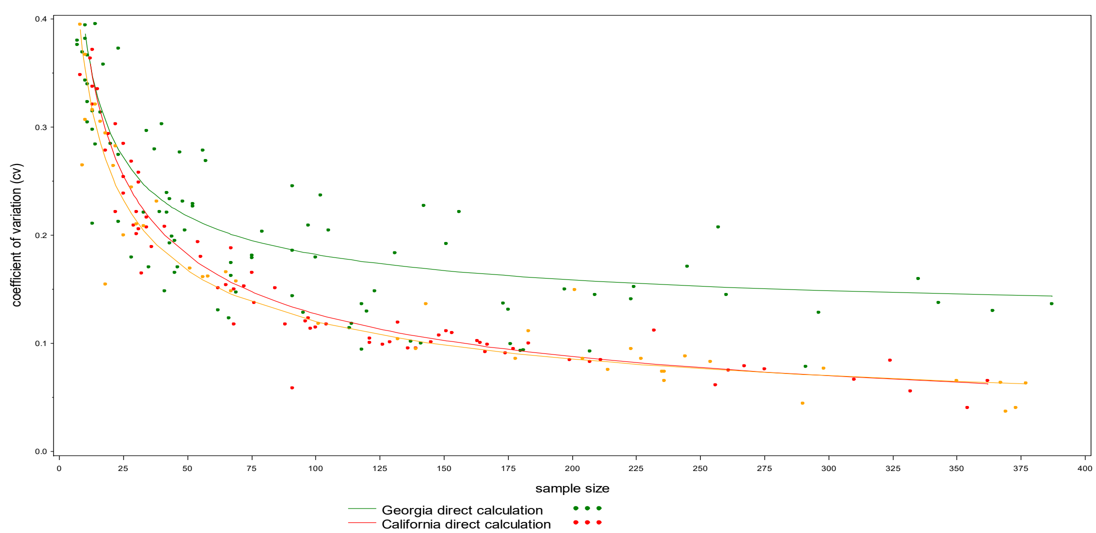


**Figure 7. Sample Size versus Coefficient of Variance for Two States, Method 2**

Although the results of method 2 seem to be consistent with the graph of method 1 in Figure 7, we are currently recommending method 1 over method 2. When the GVF is constrained, for the case of $\hat{v}(\hat{N}_d) = 0$ when $\hat{N}_d = 0$ or the case of $\hat{v}(\hat{N}_d) = 0$ when $\hat{N}_d = N$, we know that the CV graph is also constrained in the same way since the CV graph is a transformation of the constrained GVF. With method 2, we are not using the GVF and we have not found a way to similarly include the constraints in the CV graph. Without the constraints, we think method 2 is more sensitive to the model selected.

### 4.3 Evaluating State Estimates

Returning to our question of interest – are the state estimates reasonable – we show how the graph of the CVs can contribute to this question. Table 4.1 summarizes the state CV graphs for several states considered by AHS. From the CV models, we were able to determine the sample sizes necessary for the three CVs of interest: 30% (almost bad), 15% (good), and 5% (great). The results of Table 4.1 can help us identify very good and very bad states; however, it does not provide an explicit rule classifying states in the middle. At some point, we need to "draw a line" for the criteria ourselves.

**Table 4.1. Average Sample Size for a Given CV**

| | Coefficient of Variation | | |
|---|---|---|---|
| State | <30% | <15% | <5% |
| U.S. | 26 | 103 | 917 |
| California | 17 | 67 | 574 |
| Texas | 20 | 78 | 646 |
| Florida | 30 | 116 | 880 |
| New York | 13 | 53 | 408 |
| Pennsylvania | 23 | 89 | 642 |
| Illinois | 30 | 160 | 574 |
| New Jersey | 13 | 49 | 344 |
| Massachusetts | 38 | 196 | nd |
| Georgia | 56 | 470 | nd |

Good states were determined to have sample sizes for a given CV that are comparable to the sample sizes of the U.S. estimates. Generally, the good states included CA, TX, FL, NY, PA, IL, and NJ. We excluded states starting with MA and GA because these states could not produce estimates with CVs less than 5% and we think a state should be able to support CVs that are less than 5%. States after MA and GA had worse GVFs. In the end, we also excluded NJ and IL due to the large CVs of specific estimates.

We also suggest a secondary use for the CV models: giving simple and explicit guidance to data users with respect to the reliability of estimates. Instead of providing guidance to data users with respect to CVs, we can provide average sample sizes that will produce a given CV. For example, in PA, estimates with sample sizes less than 23 should be used with great caution because they will on average have CVs of 30% or more. Similarly, if an estimate has a sample size of 400 or more in NY, we could say that the estimate will be very good because the CV will be 5% on average. We see that this type of information could be valuable for different data users.

### 5. Conclusion

We have shown how three large surveys are applying innovative methods to produce state-level estimates using a national surveys. In all three instances, state estimates were not an original priority. Some of the methods are incorporated directly into the sampling strategy of the national estimates and other methods were applied outside of the national estimates, e.g., state weights are produced.

GVFs and graphs of CVs were shown to be useful for examining a set of estimates and are useful for surveys with a general purpose surveys. In our case, we applied the GVFs and graphs of CVs to our state estimates, but they can also be applied to other comparisons of different aspects of a sample design.

*Views expressed in this paper are those of the authors and do not reflect the views or policies of the U.S. Census Bureau or the Bureau of Labor Statistics.*

**References**

Ash, S. (2003). "Simultaneous Calibration for Two-Phase Samples," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 395-400.

Ash, S. (2014). "Using successive difference replication for estimating variances," Survey Methodology, 40, 47-59.

Bureau of Justice Statistics. (2017). "National Crime Victimization Survey, 2016 – Technical Documentation," NCJ 251442, December 8, 2017.

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

Deville, J.-C. and Särndal, C.E. (1992). "Calibration Estimators in Survey Sampling," Journal of the American Statistical Association, 87, 418, 376-382.

Estevao, V.M. and Särndal C.-E. (2002). "The Ten Cases of Auxiliary Information for Calibration for Two Phase Sampling," *Journal of Official Statistics*, 18, 2, 233-255.

Fay, R.E. and Li, J. (2012). "Rethinking the NCVS: Subnational Goals through Direct Estimation," presented at the 2012 Federal Committee on Statistical Methodology Conference, Washington D.C., January 10-12, 2012.

Fay, R.E., and Train, G.F. (1995). "Aspects of survey and model-based postcensal estimation of income and poverty characteristics for states and counties." *Proceedings of the Section on Government Statistics*, American Statistical Association, 154-159.

Kott, P. and Chang, T. (2010). "Using Calibration Weighting to Adjust for Nonignorable Unit Nonresponse," Journal of the American Statistical Association, 105, 1265-1275.

Krenske, T. (1995). Reevaluating Generalized Variance Model Parameters for the National Crime Victimization Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 327-332.

Lundström, S. and Särndal, C.-E. (1999). "Calibration as a Standard Method for Treatment of Nonresponse," Journal of Official Statistics, 15, 305-327.

Särndal, C.-E., Swenson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*, Spring-Verlag.

U.S. Census Bureau (2006). "Design and Methodology: Current Population Survey" Technical Paper 66, October 2006.

Wolter, K.M. (1985). *Introduction to Variance Estimation*. Springer-Verlag.