Finite Population Correction (FPC) for NAEP Variance Estimation

Lou Rizzo, Keith Rust
Westat, 1600 Research Boulevard, Rockville, MD 20850

Abstract
The National Assessment of Educational Progress (NAEP) uses jackknife replicate weights for estimating sampling variances. In the presence of nonnegligible finite population corrections, the jackknife requires either special factors attached to each sum of squares or adjustments to be made to the jackknife replicate weights to provide consistent variance estimators. The NAEP sample design has two stages of selection. We needed to incorporate a finite population correction for the first-stage of selection (schools), but not for the second-stage of selection (students). A method is developed for doing this in a way which is simple, allowing for its use by analysts without the need for explicit factors attached to sums of squares in variance estimation (the necessary adjustments are fully incorporated into the weights). The approximation is conservative in that it slightly overestimates the true variance. This paper provides theoretical results, and the companion paper Kali et al. (2011) shows results of its application to NAEP 2011.

Key Words: National Assessment of Educational Progress (NAEP), jackknife replication methods, finite population correction (FPC)

1. Introduction

Finite population corrections are an important part of variances for survey sampling estimates. A widely-used methodology for estimating variances in practice are replication methods, including the jackknife, bootstrap, and balanced repeated replication (see for example Rust and Rao (1996)). Replication methods depend on using the sample itself to provide information about the variability induced by the sampling process. It is difficult to incorporate finite population corrections into replicate variance methods because these factors are not apparent in the sample distribution itself: they are 'external' to the sample distribution. They have to be brought in explicitly in a careful way. This paper describes the new approach for incorporating finite population corrections into replicate variances for the National Assessment of Educational Progress (NAEP)[1]. This methodology will be used in future years of the NAEP program, starting with the current year.

NAEP's variance estimation procedure for many decades has been the jackknife procedure. Without adjustment, the jackknife procedure essentially provides an unbiased variance estimator for totals[2] assuming the finite population correction is negligible. If

[1] The National Assessment of Educational Progress measures the achievement of fourth-, eighth-, and twelfth-graders in the United States through a nationally representative sample. It is sponsored by the US Department of Education. General information about NAEP can be found at http://nces.ed.gov/nationsreportcard/.

[2] Most other estimates such as means (weighted totals divided by the sum of weights), regression coefficients, correlation coefficients, distribution function estimates, etc. can be seen as smooth functions of weighted totals. The jackknife

the finite population corrections are not negligible, then the jackknife variance estimator can be seen as conservative (it is positively biased as an estimator of the true variance of the total).

In order to achieve an unbiased variance estimator where the finite population corrections are all negligible, it is necessary only to work with the first-stage units (the sample units from the first-stage sample, which in NAEP is generally the school level). Jackknife replicate weights that perturb first-stage units only will succeed successfully in providing unbiased variance estimators (see for example Kalton 1979).

The new approach developed in this paper provides jackknife variance estimators that are less conservative through the application of finite population corrections directly in the jackknife weighting perturbations. In principle, the concept is very simple. We start with a set of jackknife perturbations which provide unbiased variance estimators for totals. These perturbations decrease some weights for a particular replicate and increase others. The finite population correction can be accounted for by reducing the size of these perturbations in the exactly appropriate way to achieve variance estimators that are unbiased estimators of the fully fpc-corrected variance. One can cite Rizzo and Judkins (2004), Fay (1984), Fay (1989), Flyer (1987), Judkins (1990), and Rao and Wu (1988), for applications of this technique in a variety of settings.

This paper is one of a pair of papers given in this JSM 2011 session. The second paper is Kali et al. (2011), and provides empirical results for the application of this procedure to NAEP 2011.

2. Should Finite Population Corrections be Applied?

An important starting point in any survey and analysis design is deciding whether it is appropriate to include finite population corrections at all. The fact that the sampling fraction is nonnegligible does not necessarily mean that finite population corrections must be applied. If one for example is planning to analyze the sample by using a superpopulation model, and the inference will be to parameters in the superpopulation model, then incorporating the finite population correction in variance estimators is not the correct answer (see for example Korn and Graubard (1999), Section 5.7). In the NAEP application, there are two stages of selection at the school level and at the student level, with differing sampling rates (and therefore finite population corrections) associated with them.

It appears logical in the NAEP application to incorporate finite population corrections at the school level. At the school level, this will represent the true sampling variance treating the school sample as a without-replacement sample from the frame of schools for that jurisdiction. In particular, when all schools in a jurisdiction are included in the sample, the replicate variance estimator at the school level will be 0, reflecting no sampling variability in this case. And with the previous procedure this does occur for schools selected with certainty. For NAEP, we wish to represent a fixed set of schools

---

estimation theory allows one to assert that the jackknife variance estimator for these other parameters will be consistent (not unbiased), if it is unbiased for the total, and is consistently conservative (i.e., it converges to a value higher than the true value) if the jackknife variance estimator is positively biased for the total. See for example Shao and Tu (1995), Section 2.1.

within each jurisdiction. We also wish to have a procedure that treats certainty schools and non-certainty schools consistently.

But what of students? Should the variance be zero for jurisdictions when all schools and all grade-eligible students are included in the sample? From a strict sampling viewpoint, the answer is yes. There is no sampling, so the variance should be zero. This is certainly the correct answer when the estimators are descriptive, and especially for estimates of totals (total numbers of girls, boys, Blacks, Hispanics, etc.), as the census of schools and students will give the correct number (putting aside nonresponse and measurement error) of grade-eligible students within any demographic domain of interest. These demographic characteristics are 'fixed' characteristics: they will not change at any point.

The student characteristics of direct interest however are the students' proficiency scores, developed from the answers the students give on the NAEP assessment. These proficiencies can be viewed as being drawn for each student from an infinite population of potential proficiencies, for reasons as follows:

- Each student is only given a portion of the full assessment (a set of assigned blocks of questions), with that portion randomly selected, to reduce burden on the student;
- The proficiencies are measured using an Item Response Theory model[3] as a complex function of the actual assessment item results;
- Even given a particular fixed assessment (putting aside the IRT and the randomly assigned blocks), a given student will provide somewhat different results depending on the exact day of the exam (no student's answers will be fully consistent over time: there is some noise coming from the day-to-day differences in the student's capabilities and motivation).

The approach then for the NAEP proficiency scores is to view the student assessments as being drawn from an infinite population. In particular, we have chosen not to incorporate a finite population correction at the student level. The student assessment results from a given school are viewed as a with-replacement sample of assessments from an infinite population of such assessments, even when every grade-eligible student in the school is assessed. The student-level sampling fraction should always be viewed as negligible: essentially zero. Again, this is the approach taken previously in NAEP in the case of students sampled from within certainty schools, and again we wish to use a consistent approach across certainty and non-certainty schools.

Given this logic, the variance estimated is both a sampling variance and a model variance: a sampling variance at the school level and a model variance at the student level.

[3] See for example Baker and Kim (2004).

3. Approximate Variances for NAEP Two-Stage Sampling

We approximate the NAEP sample design for each jurisdiction at the school level as a heavily stratified design, with each stratum having a sample size of 2 or 3. The NAEP sample design is in fact a systematic sampling procedure, thus representing it as a heavily stratified design is an approximation. See for example Wolter (2007), Section 8. Using this approximation, we can write the population value of interest as follows (S is the number of strata in the jurisdiction, s designates stratum, N_s is the number of schools in the stratum, M_{si} is the number of grade-eligible students in the school, and μ_{si} is the mean proficiency within the school (a pure model parameter, not a fixed value calculable for any school)):

$$Y = \sum_{s=1}^{S}\sum_{i=1}^{N_s} M_{si}\mu_{si} \qquad \bar{Y} = \frac{\sum_{s=1}^{S}\sum_{i=1}^{N_s} M_{si}\mu_{si}}{\sum_{s=1}^{S}\sum_{i=1}^{N_s} M_{si}}$$

Y corresponds to the expectation (under the model for proficiency) of the total proficiency aggregated over all grade-eligible students in all schools in the jurisdiction. $\bar{Y}$ is the mean proficiency within the jurisdiction.

Within each sampled school, we draw a simple random sample of m_{si} 'proficiencies' from an infinite population with mean μ_{si} and variance σ_{si}^2. We designate y_{sij} as the estimated proficiency of each student j=1,…, m_{si}, n_s as the school sample size in stratum s, and

$$\bar{y}_{si} = \frac{1}{m_{si}}\sum_{j=1}^{m_{si}} y_{sij}$$

The estimator from the two-stage sample of Y is as follows:

$$\hat{Y} = \sum_{s=1}^{S}\sum_{i=1}^{n_s} \pi_{si}^{-1} M_{si}\bar{y}_{si}$$

In the general case of without replacement probability proportional to size sampling, the variance is as follows (see for example Cochran 1977, Equation 11.42[4]):

[4] Note that we are leaving out the finite population correction at the second stage, as the second stage is sampling from an infinite population.

$$Var\left(\hat{Y}\right)=\sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{j>i}^{N_s}\left(\pi_{si}\pi_{sj}-\pi_{sij}\right)\left(\frac{M_{si}\mu_{si}}{\pi_{si}}-\frac{M_{sj}\mu_{sj}}{\pi_{sj}}\right)^2+\sum_{s=1}^{S}\sum_{i=1}^{N_s}\frac{M_{si}^2\sigma_{si}^2}{m_{si}\pi_{si}}=$$
$$=\sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{j>i}^{N_s}\left(\Delta_{sij}\right)\left(w_{si}\mu_{si}-w_{sj}\mu_{sj}\right)^2+\sum_{s=1}^{S}\sum_{i=1}^{N_s}\pi_{si}\frac{w_{si}^2\sigma_{si}^2}{m_{si}}=$$
$$=Var_1\left(\hat{Y}\right)+Var_2\left(\hat{Y}\right)$$

with $\Delta_{sij}=\left(\pi_{si}\pi_{sj}-\pi_{sij}\right)$ and $w_{si}=\frac{M_{si}}{\pi_{si}}$ (1)

π_{si} is equal to the first-stage probability of inclusion for school *i*, and π_{sij} is the joint probability of inclusion for schools *i* and *j* in the same stratum (note that school selection is independent across strata, so that the joint probability of inclusion in that case is the product of the first-stage probabilities of inclusion. σ_{si}^2 is the variance among student assessments in school *si*. The unbiased variance estimator is as follows:

$$v\left(\hat{Y}\right)=v_1\left(\hat{Y}\right)+v_2\left(\hat{Y}\right)=$$
$$=\sum_{s=1}^{S}\sum_{i=1}^{n_s}\sum_{j>i}^{n_s}\frac{\left(\pi_{si}\pi_{sj}-\pi_{sij}\right)}{\pi_{sij}}\left(\frac{M_{si}\bar{y}_{si}}{\pi_{si}}-\frac{M_{sj}\bar{y}_{sj}}{\pi_{sj}}\right)^2+\sum_{s=1}^{S}\sum_{i=1}^{n_s}\frac{M_{si}^2 s_{si}^2}{m_{si}\pi_{si}}$$

with $s_{si}^2=\frac{1}{m_{si}-1}\sum_{j=1}^{m_{si}}\left(y_{sij}-\bar{y}_{si}\right)^2$

$$=\sum_{s=1}^{S}\sum_{i=1}^{n_s}\sum_{j>i}^{n_s}d_{sij}\left(w_{si}\bar{y}_{si}-w_{sj}\bar{y}_{sj}\right)^2+\sum_{s=1}^{S}\sum_{i=1}^{n_s}\pi_{si}\frac{w_{si}^2 s_{si}^2}{m_{si}}$$

with $d_{sij}=\frac{\left(\pi_{si}\pi_{sj}-\pi_{sij}\right)}{\pi_{sij}}$, $w_{si}=\frac{M_{si}}{\pi_{si}}$ (2)

One can cite for example Equation 11.44 in Cochran 1977. d_{sij} can be approximated in a conservative way in most cases using the following approach. If the π_{si} are nearly equal within a stratum (to a common value n_s/N_s), then sampling is nearly equivalent to simple random sampling without replacement, and π_{sij} will be equal in this case to $n_s\left(n_s-1\right)/\left(N_s\left(N_s-1\right)\right)$. After algebra, the value for Δ_{sij} in this case is $\left(n_s/N_s^{\,2}\right)\left((N_s-n_s)/(N_s-1)\right)$. Finally, further algebra (see Appendix) shows that $d_{sij}=\left(1-f_s\right)/\left(n_s-1\right)$, with $f_s=n_s/N_s$. In this case when the sampling probabilities within a stratum are close, a conservative approximation of d_{sij} (in that the approximation is too big, making the variance too big) is $\tilde{d}_{sij}=\frac{1-\min\left(\pi_{si},\pi_{sj}\right)}{n_s-1}$. When π_{si} equals π_{sj} equals n_s/N_s, $\tilde{d}_{sij}$ reduces to the exact d_{sij}.

$$\tilde{v}\left(\hat{Y}\right)=\tilde{v}_1\left(\hat{Y}\right)+\tilde{v}_2\left(\hat{Y}\right)=$$
$$=\sum_{s=1}^{S}\sum_{i=1}^{n_s}\sum_{j>i}^{n_s}\tilde{d}_{sij}\left(w_{si}\bar{y}_{si}-w_{sj}\bar{y}_{sj}\right)^2+\sum_{s=1}^{S}\sum_{i=1}^{n_s}\pi_{si}\frac{w_{si}^2 s_{si}^2}{m_{si}}$$
$$\text{with } \tilde{d}_{sij}=\frac{1-\min\left(\pi_{si},\pi_{sj}\right)}{n_s-1}\quad,\quad w_{si}=\frac{M_{si}}{\pi_{si}} \tag{3}$$

$\tilde{v}\left(\hat{Y}\right)$ in general is a conservative estimator of $Var\left(\hat{Y}\right)$. The actual estimator that is used in practice is the ratio estimator

$$\hat{\bar{Y}}=\frac{\hat{Y}}{\hat{M}_0}\quad \text{with } \hat{M}_0=\sum_{s=1}^{S}\sum_{i=1}^{n_s}\frac{M_{si}}{\pi_{si}}$$

This is an approximately unbiased estimator of $\bar{Y}$. The appendix develops an approximate variance expression for $\hat{\bar{Y}}$, which is as follows:

$$Var\left(\hat{\bar{Y}}\right)\approx\frac{1}{M_0^2}\sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{j>i}^{N_s}\Delta_{sij}\left[\left(w_{si}\left(\mu_{si}-\bar{Y}\right)\right)-\left(w_{sj}\left(\mu_{sj}-\bar{Y}\right)\right)\right]^2$$
$$+\quad\frac{1}{M_0^2}\sum_{s=1}^{S}\sum_{i=1}^{N_s}\pi_{si}\frac{w_{si}^2\sigma_{si}^2}{m_{si}} \tag{4}$$

An approximately unbiased estimator of this expression is as follows:

$$\tilde{v}\left(\hat{\bar{Y}}\right)\approx\frac{1}{\hat{M}_0^2}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\sum_{j>i}^{n_s}\tilde{d}_{sij}\left[\left(w_{si}\left(\bar{y}_{si}-\hat{\bar{Y}}\right)\right)-\left(w_{sj}\left(\bar{y}_{sj}-\hat{\bar{Y}}\right)\right)\right]^2$$
$$+\frac{1}{\hat{M}_0^2}\sum_{s=1}^{S}\sum_{i=1}^{n_s}\pi_{si}\frac{w_{si}^2 s_{si}^2}{m_{si}} \tag{5}$$

The replicate variance estimators developed will be designed to duplicate Equation (3) for totals, and to approximate Equation (5) for means.

4. Replicate Variance Estimators: School Level

Two sets of replicate weights will be developed to implement the variance estimator[5]. The first will be designed to reproduce $v_1(\hat{Y})$. The second will be designed to reproduce $v_2(\hat{Y})$. The first set will consist of perturbations at the school level only (perturbations in the w_s weights). The second set will consist of perturbations at the student level. The replicate variance estimator matching $v_1(\hat{Y})$ is as follows:

[5] This refers to replicate weights before grouping. In practice a particular replicate weight (e.g., replicate weight 1), will include perturbations at the school and at the student level.

$$v_{r1}\left(\hat{Y}\right)=\sum_{r=1}^{R}\left(\hat{Y}(r)-\hat{Y}\right)^2 \quad \text{with } \hat{Y}(r)=\sum_{s=1}^{S}\sum_{i=1}^{n_s} w_{si}(r)\bar{y}_{si}.$$

Before grouping, each replicate r pertains to a stratum s. In this section for simplicity's sake we will assume every n_s equals 2. The case n_s equals 3 will be developed in Section 5 below. For a stratum in which n_s equals 2, there will correspond exactly one replicate r. This replicate weight will be defined as follows:

$$w_{si}(r)=\begin{cases} w_{si}(1+\sqrt{\tilde{d}_{sij}}) & i\in A_r \\ w_{si}(1-\sqrt{\tilde{d}_{sij}}) & i\in D_r \\ w_{si} & \text{otherwise} \end{cases}$$

A_r is the sampled school which is 'retained' for replicate weight r (with no finite population correction this school would receive an r-weight of $2{*}w_{si}$), D_r is the sample school which is 'deleted' for replicate weight r (with no finite population correction this school would receive an r-weight of 0). To incorporate the finite population correction, the replicate weights are 'shrunk back' from $2{*}w_{si}$ and 0, giving a replicate weight which is larger than w_{si} but smaller than $2{*}w_{si}$ in the A_r case, and giving a replicate weight which is smaller than w_{si} but greater than 0 in the D_r case.

Using this replicate weight definition, the square $\left(\hat{Y}(r)-\hat{Y}\right)^2$ is as follows (assuming without loss of generality that $\bar{y}_{s1}$ is the assessment mean for the 'retained unit' for the stratum and $\bar{y}_{s2}$ is the assessment mean for the 'deleted unit' for the stratum).

$$\left(\hat{Y}(r)-\hat{Y}\right)^2=\left(\left(w_{s1}\left(1+\sqrt{\tilde{d}_{s12}}\right)\bar{y}_{s1}\right)+\left(w_{s2}\left(1-\sqrt{\tilde{d}_{s12}}\right)\bar{y}_{s2}\right)-\left(w_{s1}\bar{y}_{s1}+w_{s2}\bar{y}_{s2}\right)\right)^2=$$
$$=(\tilde{d}_{s12})\left(w_{s1}\bar{y}_{s1}-w_{s2}\bar{y}_{s2}\right)^2$$

Assuming one replicate r for each stratum s, the replicate variance estimator is as follows:

$$v_{r1}\left(\hat{Y}\right)=\sum_{r=1}^{R}\left(\hat{Y}(r)-\hat{Y}\right)^2=\sum_{s=1}^{S}\tilde{d}_{s12}\left(w_{s1}\bar{y}_{s1}-w_{s2}\bar{y}_{s2}\right)^2$$

One can see that this is identical to $\tilde{v}_1(\hat{Y})$ as given in Equation (3-3), so that the replicate variance estimator reproduces the unbiased variance estimator.

5. Replicate Variance Estimators: Student Level

At the student level, there will be one replicate for each sampled school (or combination of schools: see below)[6]. The overall estimator can be written as a weighted total of individual proficiency outcomes:

[6] This is before grouping.

$$\hat{Y} = \sum_{s=1}^{S}\sum_{i=1}^{n_s} w_{si}\bar{y}_{si} = \sum_{s=1}^{S}\sum_{i=1}^{n_s}\sum_{j=1}^{m_{si}} w_{ssi}y_{sij} \qquad \text{with } w_{ssi} = \frac{w_{si}}{m_{si}} = \frac{M_{si}}{\pi_{si}m_{si}}$$

The replicate variance estimator for replicate r' can be written as:

$$\hat{Y}(r') = \sum_{s=1}^{S}\sum_{i=1}^{n_s}\sum_{j=1}^{m_{si}} w_{ssi}(r')y_{sij}$$

For replicate r' corresponding to school *si*, the replicate weights are as follows:

$$w_{ssi}(r') = \begin{cases} w_{ssi}\left(1+\sqrt{\pi_{si}}\right) & j \in A_{si} \\ w_{ssi}\left(1-\sqrt{\pi_{si}}\right) & j \in D_{si} \\ w_{ssi} & \text{otherwise} \end{cases}$$

The set A_{si} is the set of 'retained students' for replicate weight r', and the set D_{si} are the 'deleted students'. A_{si} and D_{si} should each be roughly half of the m_{si} of the students in the school's sample. For simplicity's sake, assume m_{si} is always even, so that A_{si} and D_{si} each have m_{si} /2 students[7]. Write $\bar{y}_{si}(A)$ as the mean of the A_{si} student proficiencies, and write $\bar{y}_{si}(D)$ as the mean of the D_{si} student proficiencies.

The squared difference $\left(\hat{Y}(r') - \hat{Y}\right)^2$ is

$$\left(\hat{Y}(r') - \hat{Y}\right)^2 = \left(\sum_{j\in A_{si}}\sqrt{\pi_{si}}w_{ssi}y_{sij} - \sum_{j\in D_{si}}\sqrt{\pi_{si}}w_{ssi}y_{sij}\right)^2 = \pi_{si}\frac{w_{ssi}^2 m_{si}^2}{4}\left(\bar{y}_{si}(A) - \bar{y}_{si}(D)\right)^2$$
$$= \pi_{si}w_{ssi}^2 m_{si}^2\left(\bar{y}_{si}(A) - \bar{y}_{si}\right)^2$$

Suppose $\mathbf{E_R}$ is the expectation over the replication process, which draws the without replacement sample of A_{si} (m_{si}/2 from a population of m_{si}). Then $E_R\left(\bar{y}_{si}(A) - \bar{y}_{si}\right)^2$ is the variance of a sample mean from a simple random sample without replacement of size m_{si}/2 from a population of size m_{si}. In the appendix it is shown that:

$$E_R\left(\bar{y}_{si}(A) - \bar{y}_{si}\right)^2 = s_{si}^2 \big/ m_{si} \tag{6}$$

Thus

$$E_R\left(v_{2r}\left(\hat{Y}\right)\right) = E_R\left(\sum_{r'=1}^{R'}\left(\hat{Y}(r') - \hat{Y}\right)^2\right) = \sum_{r'=1}^{R'} E_R\left(\hat{Y}(r') - \hat{Y}\right)^2 =$$
$$= \sum_{s=1}^{S}\sum_{i=1}^{n_s}\pi_{si}w_{ssi}^2 m_{si}s_{si}^2 = \sum_{s=1}^{S}\sum_{i=1}^{n_s}\pi_{si}w_{si}^2\frac{s_{si}^2}{m_{si}}$$

[7] The case of odd m_{si} assigns (m_{si} +1)/2 students to A_{si} (or D_{si}). None of the substantive results below are changed.

$E_R\left(v_{2r}\left(\hat{Y}\right)\right)$ is equal to $v_2\left(\hat{Y}\right)$ as desired.

6. Replicate Variance Estimators at the School Level: The Case of Triples

This section develops replicate weights for the case in which n_s equals 3 for the stratum. We can assume for simplicity's sake that n_s=3 for every stratum (the case in which n_s sometimes equals 2 and sometimes equals 3 is a straightforward generalization from Section 4 and this section). We can write $v_1\left(\hat{Y}\right)$ in the case of n_s=3 as follows:

$$v_1\left(\hat{Y}\right)=\sum_{s=1}^{S}\left\{d_{s12}\left(w_{s1}\bar{y}_{s1}-w_{s2}\bar{y}_{s2}\right)^2+d_{s13}\left(w_{s1}\bar{y}_{s1}-w_{s3}\bar{y}_{s3}\right)^2+d_{s23}\left(w_{s2}\bar{y}_{s2}-w_{s3}\bar{y}_{s3}\right)^2\right\}$$

We define $d_{s,\max}=\max\left\{d_{s12},d_{s13},d_{s23}\right\}$. $v_1\left(\hat{Y}\right)$ then is bounded above by

$$v_{1u}\left(\hat{Y}\right)=\sum_{s=1}^{S}\left\{d_{s,\max}\left(w_{s1}\bar{y}_{s1}-w_{s2}\bar{y}_{s2}\right)^2+d_{s,\max}\left(w_{s1}\bar{y}_{s1}-w_{s3}\bar{y}_{s3}\right)^2+d_{s,\max}\left(w_{s2}\bar{y}_{s2}-w_{s3}\bar{y}_{s3}\right)^2\right\}$$

There are two replicates assigned to each stratum *s*, which will be denoted r_1 and r_2. The three sampled schools are assigned to three sets *A*, *B*, and *C* (one school per set). The replicate weights are defined as follows:

$$w_{si}(r_1)=\begin{cases} w_{si}\left(1+\dfrac{\sqrt{d_{s,\max}}}{\sqrt{2}}\right) & si\in A \\ w_{si}\left(1+\dfrac{\sqrt{d_{s,\max}}}{\sqrt{2}}\right) & si\in B \\ w_{si}\left(1-\sqrt{2*d_{s,\max}}\right) & si\in C \\ w_{si} & \text{otherwise} \end{cases} \qquad w_{si}(r_2)=\begin{cases} w_{si}\left(1+\dfrac{\sqrt{d_{s,\max}}}{\sqrt{2}}\right) & si\in A \\ w_{si}\left(1-\sqrt{2*d_{s,\max}}\right) & si\in B \\ w_{si}\left(1+\dfrac{\sqrt{d_{s,\max}}}{\sqrt{2}}\right) & si\in C \\ w_{si} & \text{otherwise} \end{cases}$$

It is shown in the appendix that $v_{r1}\left(\hat{Y}\right)$ defined using these replicates is identical to $v_{1u}\left(\hat{Y}\right)$:

$$E\left\{v_{r1}\left(\hat{Y}\right)\right\}=v_{1u}\left(\hat{Y}\right)\geq v_1\left(\hat{Y}\right) \quad \text{Eq (7)}$$

7. Discussion

This paper describes the new variance estimation approach for NAEP, and provides the philosophy and mathematical derivations underpinning it. The companion paper Kali *et al*. (2011) provides an evaluation of using this approach specifically in the NAEP context. This paper shows that this variance estimation approach is generally successful in reducing significantly the computed variance in 'middling' jurisdictions: those with significant school-level sampling fractions but in which only a minority of schools are certainty selections.

8. Appendix

Proofs of Results in Text

8.1 Proof of Unbiasedness of Variance Estimator in Equation (2)

We will start with the expectation of $v_1(\hat{Y})$. $\mathbf{E_1}$ indicates expectation with respect to the school sample and $\mathbf{E_2}$ indicates expectation with respect to the student sample.

$$\mathbf{E_1}\{\mathbf{E_2}(v_1(\hat{Y})\} = \mathbf{E_1}\left\{\sum_{s=1}^{S}\sum_{i=1}^{n_s}\sum_{j>i}^{n_s} d_{sij}\mathbf{E_2}\left(w_{si}\bar{y}_{si} - w_{sj}\bar{y}_{sj}\right)^2\right\}$$

$$\begin{aligned}
\mathbf{E_2}\left(w_{si}\bar{y}_{si} - w_{sj}\bar{y}_{sj}\right)^2 &= \mathbf{E_2}\left\{\left(w_{si}\mu_{si} - w_{sj}\mu_{sj}\right) + \left(w_{si}\left(\bar{y}_{si} - \mu_{si}\right) - w_{sj}\left(\bar{y}_{sj} - \mu_{sj}\right)\right)\right\}^2 = \\
&= \left(w_{si}\mu_{si} - w_{sj}\mu_{sj}\right)^2 + 2*\mathbf{E_2}\left\{\left(w_{si}\mu_{si} - w_{sj}\mu_{sj}\right)\left(w_{si}\left(\bar{y}_{si} - \mu_{si}\right) - w_{sj}\left(\bar{y}_{sj} - \mu_{sj}\right)\right)\right\} + \\
&+ \mathbf{E_2}\left\{\left(w_{si}\left(\bar{y}_{si} - \mu_{si}\right) - w_{sj}\left(\bar{y}_{si} - \mu_{si}\right)\right)^2\right\} = \\
&= \left(w_{si}\mu_{si} - w_{sj}\mu_{sj}\right)^2 + 2*0 + \mathbf{E_2}\left\{\left(w_{si}\left(\bar{y}_{si} - \mu_{si}\right) - w_{sj}\left(\bar{y}_{si} - \mu_{si}\right)\right)^2\right\} = \\
&= \left(w_{si}\mu_{si} - w_{sj}\mu_{sj}\right)^2 + \mathbf{E_2}\left\{w_{si}^2\left(\bar{y}_{si} - \mu_{si}\right)^2\right\} + \mathbf{E_2}\left\{w_{sj}^2\left(\bar{y}_{sj} - \mu_{sj}\right)^2\right\} + \\
&+ \mathbf{E_2}\left\{\left(w_{si}\left(\bar{y}_{si} - \mu_{si}\right) w_{sj}\left(\bar{y}_{sj} - \mu_{sj}\right)\right)\right\} = \\
&= \left(w_{si}\mu_{si} - w_{sj}\mu_{sj}\right)^2 + w_{si}^2\frac{\sigma_{si}^2}{m_{si}} + w_{sj}^2\frac{\sigma_{sj}^2}{m_{sj}}
\end{aligned}$$

Note that only $\left(\bar{y}_{si} - \mu_{si}\right)$ and $\left(\bar{y}_{sj} - \mu_{sj}\right)$ are random under the student sampling distribution, and are independent of each other (as student sampling between schools is done independently), giving the result (all covariance terms are 0). Thus,

$$\begin{aligned}
\mathbf{E_1}\left\{\mathbf{E_2}(v_1(\hat{Y})\right\} &= \mathbf{E_1}\left\{\sum_{s=1}^{S}\sum_{i=1}^{n_s}\sum_{j>i}^{n_s} d_{sij}\left\{\left(w_{si}\mu_{si} - w_{sj}\mu_{sj}\right)^2 + w_{si}^2\frac{\sigma_{si}^2}{m_{si}} + w_{sj}^2\frac{\sigma_{sj}^2}{m_{sj}}\right\}\right\} = \\
&= \sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{j>i}^{N_s}\left(\Delta_{sij}\right)\left\{\left(w_{si}\mu_{si} - w_{sj}\mu_{sj}\right)^2\right\} + \sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{j>i}^{N_s}\left(\Delta_{sij}\right)\left\{w_{si}^2\frac{\sigma_{si}^2}{m_{si}} + w_{sj}^2\frac{\sigma_{sj}^2}{m_{sj}}\right\} = \\
&= Var_1\left(\hat{Y}\right) + \sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{j>i}^{N_s}\left(\Delta_{sij}\right)\left\{w_{si}^2\frac{\sigma_{si}^2}{m_{si}} + w_{sj}^2\frac{\sigma_{sj}^2}{m_{sj}}\right\} = \\
&= Var_1\left(\hat{Y}\right) + \sum_{s=1}^{S}\sum_{i=1}^{N_s} w_{si}^2\frac{\sigma_{si}^2}{m_{si}}\sum_{j\neq i}^{N_s}\left(\Delta_{sij}\right)
\end{aligned}$$

Using Cochran (1977), Eq. 9A.36, we can reduce the last factor as follows:

$$\sum_{j\neq i}^{N_s}\Delta_{sij}=\sum_{j\neq i}^{N_s}\left(\pi_{si}\pi_{sj}-\pi_{sij}\right)=\pi_{si}\left(n_s-\pi_{si}\right)-\left(n_s-1\right)\pi_{si}=\ \pi_{si}\left(1-\pi_{si}\right)$$

Thus we have:

$$\mathbf{E_1}\left\{\mathbf{E_2}(v_1(\hat{Y})\right\}=Var_1\left(\hat{Y}\right)+\sum_{s=1}^{S}\sum_{i=1}^{N_s}w_{si}^2\,\frac{\sigma_{si}^2}{m_{si}}\left(\pi_{si}\left(1-\pi_{si}\right)\right)$$

The second part is to find the expectation of $v_2\left(\hat{Y}\right)$.

$$\mathbf{E_1}\left\{\mathbf{E_2}(v_2(\hat{Y})\right\}=\mathbf{E_1}\left\{\mathbf{E_2}\left(\sum_{s=1}^{S}\sum_{i=1}^{n_s}\pi_{si}\,\frac{w_{si}^2 s_{si}^2}{m_{si}}\right)\right\}=\mathbf{E_1}\left\{\sum_{s=1}^{S}\sum_{i=1}^{n_s}\pi_{si}\,\frac{w_{si}^2\sigma_{si}^2}{m_{si}}\right\}=\sum_{s=1}^{S}\sum_{i=1}^{N_s}\pi_{si}^2\,\frac{w_{si}^2\sigma_{si}^2}{m_{si}}$$

Thus, we can write:

$$\begin{aligned}\mathbf{E_1}\left\{\mathbf{E_2}(v(\hat{Y})\right\}&=\mathbf{E_1}\left\{\mathbf{E_2}(v_1(\hat{Y})\right\}+\mathbf{E_1}\left\{\mathbf{E_2}(v_2(\hat{Y})\right\}=\\&=Var_1\left(\hat{Y}\right)+\sum_{s=1}^{S}\sum_{i=1}^{N_s}w_{si}^2\,\frac{\sigma_{si}^2}{m_{si}}\left(\pi_{si}\left(1-\pi_{si}\right)\right)+\sum_{s=1}^{S}\sum_{i=1}^{N_s}\pi_{si}^2\,\frac{w_{si}^2\sigma_{si}^2}{m_{si}}\\&=Var_1\left(\hat{Y}\right)+\sum_{s=1}^{S}\sum_{i=1}^{N_s}\pi_{si}\,\frac{w_{si}^2\sigma_{si}^2}{m_{si}}=Var\left(\hat{Y}\right)\end{aligned}$$

8.2 Proof of Equation (4)

$\hat{\bar{Y}}$ is a ratio estimator, and its variance can be approximated using the delta method:

$$Var\left(\hat{\bar{Y}}\right)\approx\frac{1}{M_0^2}\left[\bar{Y}^2Var\left(\hat{M}_0\right)+Var\left(\hat{Y}\right)-2\bar{Y}Cov\left(\hat{M}_0,\hat{Y}\right)\right]$$

The variance of $\hat{Y}$ is given in Equation (1). The variance of $\hat{M}_0$ can be derived as follows. There is no second stage sampling inherent in the estimator $\hat{M}_0$, as the measures of size are computed exactly for each cluster. Thus there is only a first term for first-stage cluster sampling, with the same probabilities of selection for clusters as for $\hat{Y}$. Thus

$$\begin{aligned}Var\left(\hat{M}_0\right)&=\sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{j>i}^{N_s}\left(\pi_{si}\pi_{sj}-\pi_{sij}\right)\left(\frac{M_{si}}{\pi_{si}}-\frac{M_{sj}}{\pi_{sj}}\right)^2=\\&=\sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{j>i}^{N_s}\left(\Delta_{sij}\right)\left(w_{si}-w_{sj}\right)^2\end{aligned}$$

For the covariance between $\hat{Y}$ and $\hat{M}_0$, the first term for first-stage cluster sampling is only applicable, as the lack of second-stage cluster sampling (effectively) for $\hat{M}_0$ will render the covariance at this level equal to 0.

$$Cov\left(\hat{Y},\hat{M}_0\right)=\sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{j>i}^{N_s}\left(\pi_{si}\pi_{sj}-\pi_{sij}\right)\left(\frac{M_{si}\mu_{si}}{\pi_{si}}-\frac{M_{sj}\mu_{sj}}{\pi_{sj}}\right)\left(\frac{M_{si}}{\pi_{si}}-\frac{M_{sj}}{\pi_{sj}}\right)=$$
$$=\sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{j>i}^{N_s}\left(\Delta_{sij}\right)\left(w_{si}-w_{sj}\right)\left(w_{si}\mu_{si}-w_{sj}\mu_{sj}\right)$$

Thus

$$Var\left(\hat{\bar{Y}}\right)\approx\frac{1}{M_0^2}\sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{j>i}^{N_s}\Delta_{sij}\left[\left(w_{si}\mu_{si}-w_{sj}\mu_{sj}\right)^2+\left(w_{si}\bar{Y}-w_{sj}\bar{Y}\right)^2-2\left(w_{si}\bar{Y}-w_{sj}\bar{Y}\right)\left(w_{si}\mu_{si}-w_{sj}\mu_{sj}\right)\right]$$
$$+\frac{1}{M_0^2}\sum_{s=1}^{S}\sum_{i=1}^{N_s}\pi_{si}\frac{w_{si}^2\sigma_{si}^2}{m_{si}}$$
$$=\frac{1}{M_0^2}\sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{j>i}^{N_s}\Delta_{sij}\left[\left(w_{si}\mu_{si}-w_{sj}\mu_{sj}\right)-\left(w_{si}\bar{Y}-w_{sj}\bar{Y}\right)\right]^2+\frac{1}{M_0^2}\sum_{s=1}^{S}\sum_{i=1}^{N_s}\pi_{si}\frac{w_{si}^2\sigma_{si}^2}{m_{si}}$$
$$=\frac{1}{M_0^2}\sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{j>i}^{N_s}\Delta_{sij}\left[\left(w_{si}\left(\mu_{si}-\bar{Y}\right)\right)-\left(w_{sj}\left(\mu_{sj}-\bar{Y}\right)\right)\right]^2+\frac{1}{M_0^2}\sum_{s=1}^{S}\sum_{i=1}^{N_s}\pi_{si}\frac{w_{si}^2\sigma_{si}^2}{m_{si}}$$

8.3 Proof of Equation (6)

In this section we show that $E_R\left(\bar{y}_{si}(A)-\bar{y}_{si}\right)^2=s_{si}^2/m_{si}$

From Cochran 1977, Section 2.6, the variance of a sample mean $\bar{y}$ of size n from a population of size N, with population variance S^2 is

$$Var(\bar{y})=\frac{N-n}{Nn}S^2$$

In this case for the sample mean $\bar{y}_{si}(A)$, the population size is m_{si}, the sample size is $m_{si}/2$, and the population variance is s_{si}^2, thus

$$E_R\left(\bar{y}_{si}(A)-\bar{y}_{si}\right)^2=\frac{1}{2}\frac{2}{m_{si}}s_{si}^2=\frac{s_{si}^2}{m_{si}}$$

8.4 Proof of Equation (7)

In this section we show that $E\left(v_{r1}(\hat{Y})\right)=v_{1u}(\hat{Y})$ when $n_s=3$

Using the replicate weight definition in Section 6, the sum of squares from the two replicates r_1 and r_2 is as follows:

$$\left(\hat{Y}(r_1)-\hat{Y}\right)^2 + \left(\hat{Y}(r_2)-\hat{Y}\right)^2 = \left\{\begin{array}{l}\left(w_{s1}\left(1+\frac{\sqrt{d_{s,\max}}}{\sqrt{2}}\right)\overline{y}_{s1}\right)+\left(w_{s2}\left(1+\frac{\sqrt{d_{s,\max}}}{\sqrt{2}}\right)\overline{y}_{s2}\right)+\left(w_{s3}\left(1-\sqrt{2*d_{s,\max}}\right)\overline{y}_{s3}\right)\\ -\left(w_{s1}\overline{y}_{s1}+w_{s2}\overline{y}_{s2}+w_{s3}\overline{y}_{s3}\right)\end{array}\right\}^2$$

$$+\left\{\begin{array}{l}\left(w_{s1}\left(1+\frac{\sqrt{d_{s,\max}}}{\sqrt{2}}\right)\overline{y}_{s1}\right)+\left(w_{s2}\left(1-\sqrt{2*d_{s,\max}}\right)\overline{y}_{s2}\right)+\left(w_{s3}\left(1+\frac{\sqrt{d_{s,\max}}}{\sqrt{2}}\right)\overline{y}_{s3}\right)\\ -\left(w_{s1}\overline{y}_{s1}+w_{s2}\overline{y}_{s2}+w_{s3}\overline{y}_{s3}\right)\end{array}\right\}^2 =$$

$$= d_{s,\max}\left(\frac{w_{s1}\overline{y}_{s1}+w_{s2}\overline{y}_{s2}}{\sqrt{2}}-\sqrt{2}w_{s3}\overline{y}_{s3}\right)^2 + d_{s,\max}\left(\frac{w_{s1}\overline{y}_{s1}+w_{s3}\overline{y}_{s3}}{\sqrt{2}}-\sqrt{2}w_{s2}\overline{y}_{s2}\right)^2$$

If we take the expectation over the replication process, each of the three school level means has equal chance of being unit 1, unit 2, and unit 3: so that

$$E_R\left[\left(\hat{Y}(r_1)-\hat{Y}\right)^2+\left(\hat{Y}(r_2)-\hat{Y}\right)^2\right]=$$

$$=\frac{2}{3}\ d_{s,\max}\left\{\begin{array}{l}\left(\frac{w_{s1}\overline{y}_{s1}+w_{s2}\overline{y}_{s2}}{\sqrt{2}}-\sqrt{2}w_{s3}\overline{y}_{s3}\right)^2+\left(\frac{w_{s1}\overline{y}_{s1}+w_{s3}\overline{y}_{s3}}{\sqrt{2}}-\sqrt{2}w_{s2}\overline{y}_{s2}\right)^2\\ \quad +\left(\frac{w_{s2}\overline{y}_{s2}+w_{s3}\overline{y}_{s3}}{\sqrt{2}}-\sqrt{2}w_{s1}\overline{y}_{s1}\right)^2\end{array}\right\}=$$

$$=\frac{1}{3}\ d_{s,\max}\left\{\begin{array}{l}\left(w_{s1}\overline{y}_{s1}+w_{s2}\overline{y}_{s2}-2w_{s3}\overline{y}_{s3}\right)^2+\left(w_{s1}\overline{y}_{s1}+w_{s3}\overline{y}_{s3}-2w_{s2}\overline{y}_{s2}\right)^2\\ \quad + \quad \left(w_{s2}\overline{y}_{s2}+w_{s3}\overline{y}_{s3}-2w_{s1}\overline{y}_{s1}\right)^2\end{array}\right\}=$$

$$=\frac{d_{s,\max}}{3}\left\{\begin{array}{l}\left(w_{s1}\overline{y}_{s1}\right)^2+\left(w_{s2}\overline{y}_{s2}\right)^2+4\left(w_{s3}\overline{y}_{s3}\right)^2+\quad\left(w_{s1}\overline{y}_{s1}\right)^2+4\left(w_{s2}\overline{y}_{s2}\right)^2+\left(w_{s3}\overline{y}_{s3}\right)^2\\ +\ 4\ \left(w_{s1}\overline{y}_{s1}\right)^2+\left(w_{s2}\overline{y}_{s2}\right)^2+\left(w_{s3}\overline{y}_{s3}\right)^2+\\ +\ 2\left(w_{s1}\overline{y}_{s1}w_{s2}\overline{y}_{s2}\right)+\ 2\left(w_{s1}\overline{y}_{s1}w_{s3}\overline{y}_{s3}\right)+\ 2\left(w_{s2}\overline{y}_{s2}w_{s3}\overline{y}_{s3}\right)-\\ -\ 4\left(w_{s1}\overline{y}_{s1}w_{s3}\overline{y}_{s3}\right)-\ 4\left(w_{s2}\overline{y}_{s2}w_{s3}\overline{y}_{s3}\right)-\ 4\left(w_{s1}\overline{y}_{s1}w_{s2}\overline{y}_{s2}\right)\\ -\ 4\left(w_{s2}\overline{y}_{s2}w_{s3}\overline{y}_{s3}\right)-\ 4\left(w_{s1}\overline{y}_{s1}w_{s2}\overline{y}_{s2}\right)-\ 4\left(w_{s1}\overline{y}_{s1}w_{s3}\overline{y}_{s3}\right)\end{array}\right\}=$$

$$=\frac{d_{s,\max}}{3}\left\{6\left(w_{s1}\overline{y}_{s1}\right)^2+6\left(w_{s2}\overline{y}_{s2}\right)^2+6\left(w_{s3}\overline{y}_{s3}\right)^2-6\left(w_{s1}\overline{y}_{s1}w_{s2}\overline{y}_{s2}\right)-6\left(w_{s1}\overline{y}_{s1}w_{s3}\overline{y}_{s3}\right)-6\left(w_{s2}\overline{y}_{s2}w_{s3}\overline{y}_{s3}\right)\right\}=$$

$$=2*d_{s,\max}\left\{\left(w_{s1}\overline{y}_{s1}\right)^2+\left(w_{s2}\overline{y}_{s2}\right)^2+\left(w_{s3}\overline{y}_{s3}\right)^2-\left(w_{s1}\overline{y}_{s1}w_{s2}\overline{y}_{s2}\right)-\left(w_{s1}\overline{y}_{s1}w_{s3}\overline{y}_{s3}\right)-\left(w_{s2}\overline{y}_{s2}w_{s3}\overline{y}_{s3}\right)\right\}\ldots$$

We can also show:

$$\left\{\left(w_{s1}\overline{y}_{s1}-w_{s2}\overline{y}_{s2}\right)^2+\left(w_{s1}\overline{y}_{s1}-w_{s3}\overline{y}_{s3}\right)^2+\left(w_{s2}\overline{y}_{s2}-w_{s3}\overline{y}_{s3}\right)^2\right\}=$$

$$=\left\{\begin{array}{l}\left(w_{s1}\overline{y}_{s1}\right)^2-2\left(w_{s1}\overline{y}_{s1}w_{s2}\overline{y}_{s2}\right)+\left(w_{s2}\overline{y}_{s2}\right)^2+\left(w_{s1}\overline{y}_{s1}\right)^2-2\left(w_{s1}\overline{y}_{s1}w_{s3}\overline{y}_{s3}\right)+\left(w_{s3}\overline{y}_{s3}\right)^2+\\ +\left(w_{s2}\overline{y}_{s2}\right)^2-2\left(w_{s2}\overline{y}_{s2}w_{s3}\overline{y}_{s3}\right)+\left(w_{s3}\overline{y}_{s3}\right)^2\end{array}\right\}=$$

$$=\left\{\begin{array}{l}2\left(w_{s1}\overline{y}_{s1}\right)^2+2\left(w_{s2}\overline{y}_{s2}\right)^2+2\left(w_{s3}\overline{y}_{s3}\right)^2-\\ -2\left(w_{s1}\overline{y}_{s1}w_{s2}\overline{y}_{s2}\right)-2\left(w_{s1}\overline{y}_{s1}w_{s3}\overline{y}_{s3}\right)-2\left(w_{s2}\overline{y}_{s2}w_{s3}\overline{y}_{s3}\right)\end{array}\right\}$$

The two expressions together give us:

$$E_R\left[\left(\hat{Y}(r_1)-\hat{Y}\right)^2+\left(\hat{Y}(r_2)-\hat{Y}\right)^2\right]=$$
$$d_{s,\max}\left\{\left(w_{s1}\overline{y}_{s1}-w_{s2}\overline{y}_{s2}\right)^2+\left(w_{s1}\overline{y}_{s1}-w_{s3}\overline{y}_{s3}\right)^2+\left(w_{s2}\overline{y}_{s2}-w_{s3}\overline{y}_{s3}\right)^2\right\}$$

Thus we have:

$$E_R\left[\sum_{r=1}^{R}\left(\hat{Y}(r)-\hat{Y}\right)^2\right]=v_{1u}(\hat{Y})$$

References

Baker, F. B., and Kim, S-H. (2004). *Item Response Theory: Parameter Estimation Techniques, 2nd ed.* New York: Marcel Dekker.

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

Fay, R. E. (1984). Some properties of estimates of variance based on replication methods. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 495-500.

Fay, R. E. (1989). Theory and application of replicate weighting for variance calculations. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 212-218.

Flyer, P. (1987). Finite population correction for replication estimates of variance. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 732-736.

Fuller, W. (1998). Replicate variance estimation for two-phase samples. *Statistica Sinica 8*, 1153-1164.

Judkins, D. (1990). Fay's method for variance estimation. *Journal of Official Statistics 6*, 223-240.

Kali, J., Burke, J., Hicks, L., Rust, K. and Rizzo, L. (2011). Incorporating a first-stage finite population correction (FPC) in variance estimation for a two-stage design in the National Assessment of Educational Progress (NAEP). *Proceedings of the Section on Survey Research Methods, American Statistical Association*,.

Kalton, G. (1979). Ultimate cluster sampling. *Journal of the Royal Statistical Society A, 142, Part 2*, 210-222.

Korn, E. L., and Graubard, B. I. (1999). *Analysis of Health Surveys.* New York: John Wiley & Sons.

Rao, J. N. K., and Wu, C. F. J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association 83*, 231-241.

Rizzo, L., and Judkins, D. (2004). Replicate variance estimation for the National Survey of Parents and Youths. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 4257-4263.

Rust, K. F., and Rao, J. N. K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research 5*, 283-310.

Shao, J., and Tu, D. (1995). *The Jackknife and Bootstrap.* New York: Springer.

Wolter, K. M. (2007). *Introduction to Variance Estimation, 2nd ed.* New York: Springer.