ON THE PERFORMANCE OF REPLICATION-BASED VARIANCE ESTIMATION METHODS WITH SMALL NUMBER OF PSUS

Mingxiu Hu, Fan Zhang, Synectics, Michael P. Cohen, NCES, and Sameena Salvucci, Synectics
Mingxiu Hu, Synectics for Mgmt Decision Inc., 3030 Clarendon Blvd #305, Arlington, VA 22201

KEY WORDS: Balanced repeated replication (BRR), Fay's method, Jackknife, Bootstrap, Random group, VPLX software

1. Introduction. Most surveys conducted by the National Center for Education Statistics (NCES) apply complex designs. Complex designs which combine some of sampling techniques such as sampling without replacement or with unequal probability, stratification, or multistage sampling, etc., induce a non-iid structure to the data. Conventional variance estimation methods are often difficult to extend to these complex survey data structures or are cumbersome to implement. The standard statistical software packages, such as SAS and SPSS, give inappropriate and usually too small variance estimates. One solution to this difficulty is to use *replication-based variance estimation approaches*, also called *resampling variance estimation approaches* in some cases. A number of replication methods have been proposed over years. Among them, the *jackknife*, *bootstrap*, *BRR*, *Fay's method*, and *random group* have received broad attention.

The problem of variance estimation with small numbers of primary sampling units (PSUs) happens most often with stratified multistage sampling, which is often adopted by NCES surveys. With this type of sampling design, although the total number of PSUs is very large, some strata (explicit and /or implicit) may only have small numbers of PSUs but may contribute substantial numbers of secondary units to the sample. If we are interested in some subpopulation parameters, we may encounter the problems of variance estimation with small numbers of PSUs since many sub-populations will only have small numbers of PSUs.

In case when a large sample of secondary units are drawn from only a few PSUs, it may be able to provide a pretty close point estimator, but it may not be able to provide a reliable variance estimate. This is because direct variance estimators must, explicitly or implicitly, estimate the between PSU component of variance. The precision of this between-PSU variance estimator will be low due to the small number of PSUs. Burke and Rust (1995) conducted a simulation study on a subsample of National Assessment of Educational Progress (NAEP) to examine the performance of two jackknife methods, the usual jackknife and the paired jackknife, with small number of PSUs.

This paper is to evaluate the six replication-based variance estimation approaches stated earlier when only small numbers of PSUs are available. We conducted a simulation study on a subset of 1993-94 Schools and Staffing Survey (SASS). Our simulation population consists of 182 private schools of SASS sample. It differs from Burke and Rust (1995) in five aspects: (1) different variance estimation methods; (2) different evaluation criteria; (3) different software used; (4) different statistics; (5) different simulation populations.

2. Replication-based variance estimation methods The basic idea behind the replication methods is to select subsamples repeatedly from the whole sample, to calculate the statistic of interest for each of these subsamples, and then use the variability among these subsample or replicate statistics to estimate the variance of the full sample statistics. Denote the estimator of the statistic of interest for the r-th replicate sample by $\hat{\theta}_r$ (r=1, ..., K), and the estimator based on the parent sample is $\hat{\theta}$. The design-based estimators $\hat{\theta}_r$ and $\hat{\theta}$ are obtained through standard estimating approaches. Then replication-based variance estimates take the form

$$\hat{v}(\hat{\theta}) = c\sum_{r=1}^{K}(\hat{\theta}_r - \hat{\theta})^2 \qquad (1)$$

or

$$\hat{v}(\hat{\bar{\theta}}) = c\sum_{r=1}^{K}(\hat{\theta}_r - \hat{\bar{\theta}})^2 , \qquad (2)$$

where $\hat{\bar{\theta}} = \sum_{r=1}^{K}\hat{\theta}_r \Big/ K$, and c is an adjusting constant. It is apparent that (1) and (2) are identical for linear estimators, but, for nonlinear estimators, (1) is more conservative than (2). However, in many surveys, the expectation of the difference between (1) and (2), $K(\hat{\theta} - \hat{\bar{\theta}})^2$, is small. The software VPLX of Fay uses the estimator (1), so does our simulation since we mainly used VPLX to implement the methods. Wolter (1985), however, in his discussion on the properties of the replication methods, focuses on estimator (2), which is easier to discuss theoretically.

The key difference among the different replication methods is that they draw different replicate samples to form the estimates $\hat{\theta}_r$ (r=1, 2, ..., K). Some methods such as the jackknife use more PSUs each time, and therefore the variation among the replicate statistics is smaller and hence we need larger adjusting constant c in (1) and (2); while some such as the random group use less PSUs each time. Next we will briefly present how each method constructs its replicates and what software is available for it.

Random Group: In this method, the full sample is randomly divided into K parts, called random groups, in a manner designed to represent the major sources of variation arising from the sample design. Each random group is used as one replicate sample. Since, among the six methods, it uses the fewest PSUs in each replicate sample, the variation among the replicate estimates is the largest and thus we need the smallest $c=1/K(K-1)$. This method has been implemented by the following software packages: (1) VPLX of Fay (public domain); (2) OSIRIS IV of Kish (commercial); (3) CLUSTERS of Verma (a normal charge).

Simple and Stratified Jackknife: The simple jackknife creates replicate estimates based on all but one PSU in succession; that is, each replicate estimate omits one PSU while re-weighting the remaining K-1 PSUs by the factor K/(K-1). Since this method uses the most PSUs in each replicate sample, the variation among the replicate estimates is the smallest and thus the adjusting constant $c=(K-1)/K$ is the largest among the methods.

In the stratified jackknife, we assume that S strata have been formed and there are K_s PSUs in the s-th stratum. The (r, s)-th replicate estimate $\hat{\theta}_{rs}$ (r=1, …, K_s, s=1, …, S) is obtained by omitting the r-th PSU and re-weighting the remaining K_s-1 PSUs by the factor $K_s/(K_s-1)$ in the s-th stratum while using the original weights for the PSUs from the other strata. Then the stratified jackknife variance estimate is given by

$$\hat{v}_{sjk}(\hat{\theta}) = \sum_{s=1}^{S} \frac{K_s - 1}{K_s} \sum_{r=1}^{K_s} (\hat{\theta}_{rs} - \hat{\theta})^2 . \qquad (3)$$

The simple and stratified jackknife have been implemented by the following software products: (1) VPLX of Fay(public domain); (2) WesVarPC of Westat (public domain); (3) OSIRIS IV of Kish (commercial); (4) GES V4.0 of Statistics Canada (commercial); (5) BOJA of Boomsma (commercial).

Balanced Repeated Replication (BRR): The BRR is a special half-sample replication method. It uses half of the sample each time and is usually applied to stratified sample designs in which the sample consists of two PSUs from each stratum. If some strata have more than two PSUs, we may either group them into two super-PSUs or divide those strata into smaller (artificial) strata such that each stratum consists of two and only two PSUs. After the desired strata have been created, one PSU from each stratum will be selected to form one replicate. There is a total of 2^S possible half-sample replicates, where S is the number of strata. The BRR uses K (out of 2^S) orthogonal balanced half-sample replicates to obtain variance estimates through Hadamard matrix (Wolter, 1985). The information contained in the 2^S replicates can be captured by K balanced replicates. The minimum number of replicates needed to have full information is the smallest integer greater than or equal to S which is divisible by 4.

The adjusting constant for the BRR is $c=1/K$, which is larger than 1/K(K-1) for the random group but smaller than (K-1)/K for the jackknife.

This method has been implemented by: (1) VPLX of Fay (public domain); (2) WesVarPC of Westat (public domain); (3) OSIRIS IV of Kish (commercial).

Fay's method: This method is a modified version of the BRR. In the BRR, half of the sample is zero-weighted while the other half is double-weighted. Fay's method assigns weight ρ ($0 \leq \rho < 1$) to one half sample and 2-ρ to the other half. The adjusting constant for this method is $c=1/K(1-\rho)^2$, which is larger than c=1/K for the BRR. In this simulation, ρ=0.5 was used in Fay's method. Fay's method has been implemented by: (1) VPLX of Fay (public domain); (2) WesVarPC of Westat (public domain).

Bootstrap: Bootstrap replicates are created using two steps: (1) using the parent sample, construct an artificial population U^*, assumed to mimic the real but unknown population U; (2) draw K independent bootstrap replicate samples from U^* using a design identical to the one by which the parent sample was drawn from U. The adjusting constant for the bootstrap is c=1/K, which is the same as the one for the BRR.

No software product has yet been developed for the general bootstrap method. So far, *BOJA* written by Boomsma and reviewed by Dalgleish (1995) may be the best software for the bootstrap method. *Resampling Stat for Windows* (Version 4.0) can only be used for the simple random sampling design.

3. Simulation population, sampling scheme and implementation. In the 1993-94 SASS, private schools were first stratified by Affiliation (19 groups), School Level (3 levels) and Census Region (4 regions). Within each stratum, the schools were further sorted by six variables: State, Highest Grade, Urbanicity, First Two Digits of Zip Code, 1991-92 Enrollment and PIN number. Then the school samples were selected with systematic PPS sampling schemes from each stratum. The measure of the PSU (school) size was the square root of the number of teachers from the 1991-92 Private School Survey.

Our artificial simulation population consists of 182 private schools from the four smallest affiliations in the 1993-94 SASS: 26 schools from the Association of American Military Colleges and Schools, 60 from the Friends Council on Education, 44 from the Solomon Schechter Day Schools, and 50 from Other Lutheran affiliation.

The 182 private schools in the artificial population were first divided into three strata by the school level variable: elementary, secondary, and combined. Within each stratum, the schools were further sorted by the same six sorting variables used in the original SASS design. Then the systematic PPS sampling algorithm was used to select the schools.

In our simulation, we employed the systematic PPS sampling scheme used in the original SASS, but did not exactly apply its stratification strategies. A stratified sampling scheme first allocates a sample size to each stratum, then draws a subsample from each stratum, and then combines all the subsamples into one overall sample. In our simulation, in order to find the true variance, we had to compute variance estimates for all possible samples. If we had applied the stratification strategy, the number of all possible samples would have become too large to implement. Therefore we decided not to pre-allocate the sample size to each stratum before performing systematic PPS sampling. Although we did not pre-allocate the sample sizes to the strata, the subsample sizes of the strata obtained through our overall systematic PPS sampling scheme were almost identical to what a stratified PPS sampling scheme would have allocated to the strata if we had employed the stratification strategy. Thus we applied the stratified jackknife method anyway for sample sizes over 12.

For each sample size n (n=2, 4, ..., or 30), there is a total of 182 possible systematic PPS samples, the same number as the artificial population size. This is the case for most systematic PPS sampling designs. We only chose even numbers as sample sizes to make it easier to implement Fay's method and the BRR. For these two method, every two adjacent PSUs were grouped into an artificial stratum. Full orthogonal balanced replicates are generated for the BRR method through the Hadamard matrix.

For the bootstrap, we used a non-systematic PPS sampling scheme to draw re-samples from the artificial population U* constructed by each possible sample, which is actually equivalent to drawing simple random samples with replacement directly from the sample S.

The random group and jackknife methods needed no special treatment to generate replicates.

After all the possible systematic PPS samples had been selected for each sample size through an Excel spreadsheet and SAS program, the re-sample selection for the bootstrap was implemented by Resample Stat for Windows, while it was done automatically for the other methods by VPLX.

4. Analysis of simulation results. In this study, we chose two estimates—student-teacher ratio and total of full-time equivalent teachers—from the 1993-94 SASS private school data. Four criteria have been used in the evaluation: (1) Bias of variance estimates; (2) MSE of variance estimates; (3) coverage probability of covering the true parameter; (4) 95% true confidence intervals of the variance.

The first column of the tables 1 and 2 gives the true variances for all the sample sizes under study. Generally, we would expect the variance to decrease as sample size increases. But some cases obviously violate this trend. For sample sizes 18, 22, and 24, the true variances for both the student-teacher ratio and the total of full-time equivalent teachers are unexpectedly small. This is probably because the systematic sampling scheme hits some pattern in the population so that the average variation among all possible PPS systematic samples are much smaller than the average variation among all possible PPS random samples. However, for sample size 26, the true variance is unexpectedly large for the student-teacher ratio, but is unexpectedly small for the total of full-time equivalent teachers. Similar reasons are responsible for the results. We should keep it in mind that we try to estimate the design-based variance, the variance among all possible systematic samples, and have no interest in the variance among all possible random samples since our estimates are based on systematic samples.

4.1 Bias of variance estimates. From Figure 1, it is evident that all the replication methods tend to overestimate the variance of the student-teacher ratio. One reason for this is that our simulation samples are drawn without replacement (*WOR sample*), while the replication methods assume that the samples are drawn with replacement (*WR sample*). Generally, a WOR sample has larger within-sample variation. If we treat a WOR sample as a WR sample, we will overestimate the true variance. Actually, Efron and Stein (1981) and Fay (1989) show that, even if the samples are drawn with replacement, the jackknife, random group, and half-sample methods still tend to overestimate the variance.

For the student-teacher ratio, the random group method always has the highest positive bias, while Fay's method always has the lowest negative bias. Since all the methods tend to overestimate the variance, Fay's method appears to be the best or close to the best in terms of bias except for the sample sizes 2 and 4. For those two cases, Fay's method seriously underestimates the variances. The other four methods are comparable.

All six methods have very large positive biases when sample size equals 18, 22 and 24. As we stated earlier, these cases have very small true variance. The true variance actually measures the variation among all possible parent samples, while each replication variance estimate is based on resamples from one parent sample. If the resamples mimic the parent samples well, we expect the replication variance estimate to be close to the true variance. However, if the within-parent-sample variation is much larger than the between-parent-sample variation (the variation in the population), then the variation between the resamples will be much larger than the variation between the parent samples, and therefore the replication method will overestimate the true variance. This is what happens for sample sizes 18, 22 and 24.

For the total of full-time equivalent teachers, the simple jackknife and the random group are identical, while the BRR and Fay's method are indistinguishable. Figure 2 shows that all methods tend to overestimate the variance, and all methods except the stratified jackknife are comparable. The stratified jackknife always has the largest positive biases except for sample size 30. This is probably because the within-stratum variations are not significantly smaller than the overall sample variations for the linear statistic, and the overall sample size is not large enough, and consequently some strata have too few PSUs, which leads to large variance estimates within those strata.

4.2 MSE of variance estimates. For the student-teacher ratio, the random group provides much less accurate variance estimates than any other methods. In many cases, the MSEs of the random group variance estimates are more than ten times larger than those of the other replication variance estimates. The random group's large biases account for a major part of its large MSEs. The BRR behaves very poorly when the sample size is less than or equal to 12, but after then it catches up with the other methods. Overall, Fay's method is again the best. It almost always has smaller MSEs than the BRR. Sample size 22 seems to be a breakdown point for all methods but the BRR and Fay's method. The stratified jackknife is among the best except for sample size 22. The simple jackknife is a little worse than the stratified jackknife but a little better than the bootstrap.

For the linear statistic, all methods except the stratified jackknife are comparable in terms of MSE. The stratified jackknife method has the largest MSEs except when the sample size is 30, in which it has the smallest MSE.

4.3 Coverage probability. The primary interest of Burke and Rust (1995) is the coverage probability of the 95% confidence interval $\hat{\theta}_{0i} \pm t(0.975)\sqrt{\hat{v}_i}$ covering the true parameter, where *t(0.975)* is the 97.5th percentile of the t-distribution. In our simulation, the t-distribution has $n_1 + n_2 + n_3 - 3$ degrees of freedom for the stratified jackknife and K-1 for the others, where K is the number of replicates, and n_s (s=1, 2, 3) is the number of PSUs in the *s*-th stratum. $\hat{\theta}_{0i}$ (i=1, ..., 182) is the estimator for the *i*-th parent sample and does not depend on the replication methods, while the $\hat{v}_i$ varies from one replication method to another; that is, the above intervals have the same center but different widths for different methods. Larger variance estimates will lead to higher coverage rates. In this situation, since all the methods tend to overestimate the variance, higher coverage rates almost always imply larger positive biases of variance estimates, which in turn means a worse replication method.

Our simulation results show that: (1) for sample sizes 18, 22 and 24, the coverage rates are too high (almost always 100%), which is because the variance estimates are too large (leads to too wide intervals); (2) for the student-teacher ratio, the random group has the highest coverage rates in most cases since it has the largest positive biases (leads to the widest intervals); (3) for the total of full-time equivalent teachers, the stratified jackknife almost always has the highest coverage rates since it always has the largest positive biases; (4) most of the coverage rates are very high since all the methods tend to overestimate the variance.

We do not think this is a good criterion for the evaluation of the replication-based variance estimation approaches due to three reasons: (1) the replication methods tend to overestimate variance, and therefore this type of coverage rate is high and not worrisome as seen in both Burke and Rust's simulation and our simulation; (2) in most cases, higher coverage rates imply a worse approach, which contradicts the usual sense of coverage probabilities; (3) if the normality assumption of the estimates does not hold, it is not appropriate either to compare the coverage rates to 95%, the nominal level.

4.4 95% confidence intervals Table 1 presents the 95% confidence intervals for the variances of the student-teacher ratio estimates obtained through the actual distribution of the variance estimates based on all possible PPS systematic samples. In table 1, the highlighted confidence intervals do not cover the true variances. In all of these cases, the true values sneak out of the intervals from the lower limits, which means that at least 97.5% of variance estimates are larger than the true variance. They are seriously positively biased. The random group and the simple jackknife both have three such bad cases with sample sizes 18, 22 and 24, the stratified jackknife has two with sample sizes 22 and 24, and the bootstrap has one with sample size 24. However, for these three cases, the BRR and Fay's method cover the true variances with convincingly shorter intervals.

For the student-teacher ratio, Fay's method is the obvious choice. It provides very sharp confidence intervals which always cover the true variances. The BRR's intervals always cover the true variances, but they are not sharp. Both jackknife methods sometimes give very sharp intervals but they can also be seriously biased. The confidence intervals of the bootstrap are considerably wider than those of Fay's method. The random group has much wider confidence intervals which still can not cover the true variances sometimes.

For the total of full-time equivalent teachers, Table 2 shows that Fay's method/BRR again has the overall best performance. Its 95% confidence intervals always cover the true variances, and it more likely provides shorter confidence intervals than any other method. The random group/simple jackknife sometimes provide very short intervals for the true variances, but they are not robust as shown by the two seriously positively biased cases in which the 95% confidence intervals can not cover the true values. All confidence intervals of the bootstrap cover the true variances, but, again, this method does not seem very sharp.

The stratified jackknife obviously has the worst overall performance for the linear statistic. It has three seriously biased cases. Its lower confidence limits always have the highest values, but it never leads to short intervals. This implies that it has a great tendency to overestimate the variance, which agrees with our findings in the bias analyses. The random group/simple jackknife always have the second largest lower confide-nce limits following the stratified jackknife.

5. Summary. All the replication methods tend to overestimate the variance on average for both linear and non-linear statistics, and the confidence intervals $\hat{\theta}_{0i} + t_{0.975}\sqrt{\hat{v}_i}$ generally have very high coverage rates for covering the true parameter. Since higher coverage rates in this situation are almost equivalent to higher positive biases, we do not think that this is a good criterion for evaluating replication variance estimation methods. When the systematic sampling design hits some underlying pattern in the population such that the average variation among all possible systematic samples is much smaller than the average variation among all possible random samples, the replication variance estimates will be seriously positively biased.

For non-linear statistics, the random group should not be considered a candidate for variance estimation. It always gives much larger biases and MSEs, and much broader confidence intervals for the variances which are sometimes still unable to cover the true variance. Although our simulation is for small sample sizes, we do not recommend using this method even for large sample sizes since no evidence shows that the random group gets closer to the other methods. We believe that the random group will not perform so poorly if more PSUs are included in each random group, but it requires a large number of PSUs since each PSU is used only once by the random group method.

For non-linear statistics, Fay's method has the best overall performance in terms of bias, MSE, and 95% confidence interval for variance estimation, and should be recommended as long as the number of PSUs in the sample from the sub-population of interest is larger than 4. The BRR performed poorly and should not be used when the sample size is smaller than 14. The bootstrap variance estimates have slightly larger MSEs, slightly broader confidence intervals compared to the best method in most cases. The stratified jackknife gives very sharp variance estimates in some occasions, but also provides seriously positively biased estimates in a few other cases. The simple jackknife is slightly worse than the stratified jackknife.

For linear statistics, the random group/simple jackknife has the overall best performance in terms of MSE, but they lose to the BRR/Fay's method and the bootstrap in terms of 95% confidence intervals. The stratified jackknife has the overall worst performance according to all the criteria used in the simulation.

Reference

Burke, J. and Rust, K. (1995). On the performance of jackknife variance estimation for systematic samples with small numbers of primary sampling units. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 321-326.

Dalgleish, L. I. (1995). Software review: bootstrapping and jacknifing with BOJA. *Statistics and Computing* **5**, 165-174.

Efron, B. and Stein, C. (1981). The jackknife estimates of variance. *Ann. Statist.* **9**, 586-596.

Fay, R. E. (1989). Theory and Application of Replicate Weighting for Variance Calculations. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 212-217.

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

Figure 1. *Bias of variance estimate for student-teacher ratio (JK--simple jackknife, SJK--stratified jackknife)*

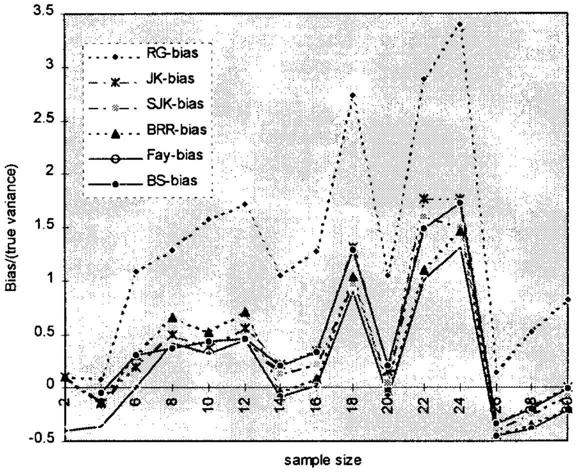

Figure 2. *Bias of variance estimate for total of full-time equivalent teachers*

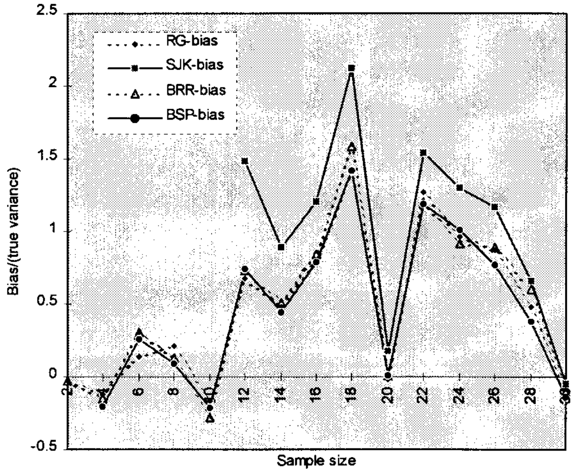

Table 1 *95% confidence interval for variance of the student-teacher ratio estimate*

Smpl size	True Variance	Random group	Simple jackknife	Stratified jackknife	BRR	Fay's method	Bootstrap
2	9.8274	.011~38.4	.011~38.4		.011~ 38.4	.011~28.2	
4	5.0131	.196~21.3	.158~21.7		.077~20.0	.067~16.1	.126~20.5
6	1.9082	.477~13.8	.366~7.11		.138~9.55	.110~6.91	.283~10.2
8	1.2428	.368~13.7	.336~4.66		.207~6.68	.205~5.68	.190~5.18
10	0.8926	.450~13.1	.407~2.80		.245~3.51	.235~3.02	.331~3.46
12	0.7122	.482~10.0	.365~2.96	.326~2.64	.323~3.60	.308~2.33	.330~2.58
14	0.7858	.387~7.16	.286~2.63	.258~2.55	.168~2.01	.167~1.85	.297~1.99
16	0.6202	.275~5.98	.354~1.82	.299~1.54	.257~1.74	.236~1.68	.280~1.94
18	0.3367	**.345~4.71**	**.384~1.94**	.236~1.38	.223~1.33	.254~1.21	.269~2.07
20	0.5485	.494~3.97	.338~1.31	.284~1.16	.163~1.19	.148~1.08	.197~1.55
22	0.2622	**.315~3.12**	**.322~2.99**	**.274~2.66**	.238~1.42	.229~1.29	.196~2.73
24	0.2117	**.399~3.02**	**.307~1.51**	**.219~1.46**	.204~1.53	.198~1.26	**.247~1.49**
26	0.7385	.294~2.91	.221~1.05	.225~.876	.134~.789	.134~.752	.137~1.13
28	0.5227	.299~2.11	.257~.672	.217~.616	.133~.589	.132~.568	.182~.743
30	0.4070	.282~1.70	.250~.785	.228~.746	.096~.669	.094~.643	.176~.908

Table 2 *95% confidence interval for variance of total of full-time equivalent teachers (in millions)*

Smpl size	True variance	Random group/ Simple jackknife	Stratified jackknife	BRR/ Fay's method	Bootstrap
2	2.4807	(.003, 14.9)		(.003, 14.9)	
4	1.3399	(.034, 4.94)		(.013, 4.43)	(.006, 5.17)
6	0.7288	(.083, 3.50)		(.034, 4.26)	(.048, 4.21)
8	0.5151	(.127, 2.90)		(.054, 1.97)	(.058, 1.92)
10	0.5776	(.136, 1.87)		(.073, .997)	(.129, 1.36)
12	0.2512	(.079, 1.51)	(.181, 1.99)	(.069, 2.01)	(.066, 1.61)
14	0.2417	(.098, 1.09)	(.166, 1.37)	(.051, .920)	(.064, 1.13)
16	0.1756	(.071, .877)	(.141, .925)	(.058, .743)	(.064, .931)
18	0.1168	**(.137, .618)**	**(.154, .732)**	(.075, .611)	(.091, .685)
20	0.2493	(.086, .708)	(.128, .823)	(.055, .726)	(.073, .781)
22	0.1004	**(.105, .468)**	**(.132, .491)**	(.096, .416)	(.088, .584)
24	0.1060	(.069, .549)	(.105, .601)	(.061, .684)	(.051, .658)
26	0.1023	(.088, .319)	**(.111, .354)**	(.073, .412)	(.061, .412)
28	0.1197	(.066, .392)	(.085, .393)	(.053, .357)	(.048, .399)
30	0.1863	(.070, .297)	(.103, .360)	(.059, .451)	(.056, .393)