An Application of the Bootstrap Variance Estimation Method to the Canadian Participation and Activity Limitation Survey

Éric R. Langlet, Dany Faucher & Éric Lesage
Statistics Canada, R.H.Coats building, 15R, Tunney's Pasture, Ottawa, Canada, K1A 0T6

KEY WORDS: bootstrap method, without replacement design, response propensity model, logistic regression

1. Overview of the problem

The bootstrap method is increasingly used to estimate the variance of estimates obtained from complex survey designs. This method, although computer intensive, has been shown to work well for a wide range of estimators, including medians and quantiles, as well as smooth functions based on totals. Linearization methods, on the other hand, require a distinct derivation for each type of statistic and are applicable only to smooth functions. In addition, the bootstrap can be less computer intensive than the jackknife method for surveys with a very large number of primary sampling units (PSUs) such as the 2001 Canadian Participation and Activity Limitations Survey (PALS) which contains over 5000 PSUs. Also, the use of bootstrap weights in microdata files eliminates the need for stratum and PSU identifiers, reducing somewhat the risk of disclosure. However, for someone knowledgeable about the survey design, the specific patterns of zero in the bootstrap weights for the different bootstrap samples could lead to the identification of the clusters. Yeo, Mantel and Liu (1999) have studied this problem.

The sampling plan of PALS is generally a stratified two-stage design in which PSUs are selected without replacement with probability proportional to size. The survey presents specific challenges to the use of the bootstrap method. For instance, the sampling fraction for PALS is relatively high in many strata, which causes the bootstrap method to overestimate the variance. What is the magnitude of this overestimation? Also, a logistic regression response propensity model is used for the nonresponse adjustment in PALS. Nonresponse classes are formed by grouping individuals in the sample with similar predicted probabilities of responding. Nonresponse rates are then calculated within each class and the sampling weights are then multiplied by the inverse of the response rate within each class. Should a logistic regression model be fitted to each bootstrap sample? How does this method compare with maintaining fixed response classes over all bootstrap samples? This paper will address these issues.

An overview of PALS and its sample design is given in Section 2. Section 3 describes how the bootstrap method was applied to PALS. In Section 4, the bootstrap estimator is compared to the Yates-Grundy estimator to measure the magnitude of the overestimation associated to the bootstrap method. In Section 5, different alternatives of adjusting the bootstrap weights for nonresponse are compared. Finally, some concluding remarks are given in Section 6.

2. Sampling plan of PALS

PALS is a post-censal survey which collects information about Canadian residents whose everyday activities are limited because of a health-related condition or problem. The survey provides essential information on the prevalence of various disabilities, the supports for people with disabilities, their employment profiles, their income and their participation in society. This information will be used by all levels of government, associations, researchers and non-government organisations to support the planning of services needed by people with activity limitations in order to participate fully in society.

PALS is referred to as a post-censal survey because it uses the Census of Population as a sampling frame to pre-identify its target population. It also benefits of the Census infra-structure to reduce its cost The 2001 Canadian Census long form is administered in most regions to a systematic one in five sample of households within each enumeration area (EA) in Canada. Census EAs are small geographical data collection units. In certain remote regions and on Indian reserves, however, the Census long form is assigned to all households. The long form contains two general filter questions on activity limitations and long-term disabilities. The 2001 PALS selected a sample of individuals from respondents on the Census long form who reported a positive response to at least one of these two filter questions. These respondents are said to be “disabled individuals” according to the Census. PALS uses, however, its own definition of disability and “disabled individuals” according to PALS represent a subset of “disabled individuals” according to the Census. PALS followed the groundwork laid by the Health and Activity Limitation Survey (HALS), a Statistics Canada survey about persons with disabilities conducted in 1986 and 1991. However, as opposed to HALS, PALS did not select a sample of individuals reporting a negative response to both filter questions. For more information on differences between PALS 2001 and HALS 1991, see Langlet (2000).

The sampling frame used for PALS consists of estimates of the 2001 Census disabled population by age group and severity of disability within each Enumeration Area (EA). These estimates were obtained from demographic projections of the Canadian population, to which disability rates (using the Census definition) estimated from the 2000 PALS pilot test were applied.

The strata are defined by the cross-classification of the ten provinces, four age groups and the Census severity of disability (defined by the response categories "Often" and "Sometimes"). For the purpose of sample selection, each stratum is subdivided into potentially three sub-strata according to PSU sizes (small, medium and large). Independent samples are selected within each sub-stratum. The PSUs are geographically made up of one or more Census EAs. Although a given EA can be selected for more than one age group and severity combination, a PSU is defined in only one stratum. The PSU size is predicted from the projected Census disabled population for the combination of EAs, age group and severity corresponding to the PSU. The sample design is a two-stage stratified design that uses the 2001 Census long-form sample in the second stage. PSUs for which the predicted number of disabled individuals was very small or null (small PSUs) were selected by stratified simple random sampling (stratified SRS). Medium PSUs were selected without replacement using probability proportional-to-size (PPS) sampling (in fact, it is really probability proportional to the estimated size sampling, but for simplicity, it will be referred as PPS sampling). Large PSUs were selected with probability one (*take-all* PSUs). *Take-all* PSUs occur for two reasons. First, a *take-all* of PSUs can be required in small strata. Second, the relative size of the PSU within a stratum can be too large to be selected with probability less than one in PPS sampling. In the second stage of the sample design, all Census long-form respondents in a selected PALS PSU are included in the 2001 PALS sample. The total sample size for PALS is about 43,000 individuals.

3. Application of the bootstrap method for PALS

The variance estimation for PALS was done using the bootstrap method. This method selects a large number of with-replacement samples, called bootstrap samples, from the original sample. The parameter of interest is estimated from each bootstrap sample and the empirical variance of these estimates over all bootstrap samples is used as an estimate of the variance of the parameter estimate. In order to simplify the operational aspects of this method, bootstrap weights are used. These weights represent the weights of the individuals within each bootstrap sample. The bootstrap method used for PALS is due to Rao and Wu (1988). The derivation of bootstrap weights is given in Rao, Wu and Yue (1992).

Calculation of the bootstrap weights is done in several steps. A number, B, of bootstrap samples is selected from the main sample. For PALS, 500 bootstrap samples were selected, which is usually large enough for most statistics produced for the survey (mainly totals and proportions). For a given bootstrap sample, the initial sampling weight of an individual is adjusted as a function of the bootstrap sample sampling fraction as well as the number of times the individual was selected in the bootstrap sample. These initial bootstrap weights are then adjusted for each weighting step performed in the full sample. This method incorporates the variance component coming from each weight adjustment such as nonresponse adjustment and post-stratification for instance.

Bootstrap samples were selected according to three situations, depending on the first stage sampling fraction and the second stage sampling fraction of households who received the long form. The first situation corresponds to *take-some* PSUs (small and medium PSUs). The second situation corresponds to *take-all* PSUs but where households at the second stage were selected with probability less than one. The third situation applies to individuals selected with probability one (*take-all* at both stages). A handful of individuals are in this situation, which will be ignored in this paper.

For *take-some* PSUs at the first stage, no distinction for bootstrap sampling is made regarding the second-stage sampling fraction. More variation is expected between totals of PSUs subsampled at the second stage than between totals of PSUs with full enumeration at the second stage. In sub-strata composed of either small or medium PSUs (*take-some* sub-strata), a sample of $n_h - 1$ PSUs within n_h PSUs is selected with replacement for each bootstrap sample. The particular selection of $n_h - 1$ PSUs within n_h PSUs simplifies the bootstrap weight formula (Rao, Wu, Yue, 1992). All full-sample second-stage units of the $n_h - 1$ selected PSUs are in the bootstrap sample.]

Let w_{hij} denote the initial sampling weight of the j^{th} individual in the i^{th} PSU of the h^{th} sub-stratum.

For a given bootstrap sample, the initial bootstrap weights are given by

$$w_{hij}^{B} = \frac{n_h}{n_h - 1} m_{hi}^{*} w_{hij}.$$

where m_{hi}^{*} represents the number of times that the hi^{th} PSU is selected in the bootstrap sample. The adjustment factors $\frac{n_h}{n_h - 1} m_{hi}^{*}$ are used so that the sum of the bootstrap weights estimates the sub-stratum population total..

For sub-strata composed of *take-all* PSUs (large PSUs) at the first stage, the sample design (a stratified systematic sample of households) can be approximated by a single-stage stratified SRS of individuals. In each such sub-stratum *h*, a with-replacement random sample of $m_h - 1$ individuals within m_h individuals is selected for each bootstrap sample. In this case, the initial bootstrap weights are given by

$$w_{hij}^{B_I} = \frac{m_h}{m_h - 1} m_{hij}^{*} w_{hij}$$

where m_{hij}^{*} represents the number of times that the hij^{th} individual is selected in the bootstrap sample. In this formula, the PSU subscript is only used to classify the individuals within their original PSU.

Once the bootstrap samples have been selected, initial bootstrap weights have to be adjusted in the same way as the initial weights of the original sample. The PALS sample selection was a manual operation in the Census Field Control Units (FCUs). Once a PSU was selected, all Census EA boxes corresponding to the PSU were inspected by sample selection clerks to find all individuals with the characteristic of the PSU (given severity level and age group). It was therefore possible for clerks to omit individuals who should have been included and to include individuals by mistake. A few months after the Census, a preliminary Census database was available and it was possible to determine which individuals should have been selected. This constitutes the "theoretical sample". The initial sampling weights in each stratum were adjusted such that the adjusted weights would sum to the stratum total estimated from the theoretical sample.

The next step was the nonresponse adjustment which is described in more detail in Section 5. This adjustment used a response propensity model to predict the probability of response from a set of explanatory variables using a logistic regression model. Nonresponse classes were then formed by combining individuals (respondents and non-respondents) with similar predicted probabilities of response. Within each class, weights for respondents were adjusted by multiplying weights from the previous step by the inverse of the observed response rate for the class. For PALS, the same model was refitted on each bootstrap sample and new classes were formed to produce new adjustments. In section 5, it will be seen that this method produces very similar estimates to those that would have been obtained if the nonresponse classes had remained fixed over all bootstrap samples.

The last step of the weight adjustment was the post-stratification to Census totals estimated from the roughly one-in-five systematic sample of households that received the long form. The weights of each bootstrap sample were post-stratified so that the sum of the weights for each post-stratum would add up to the estimate from the Census.

This bootstrap method is applicable to sample designs in which PSUs are selected with replacement or cases where the PSU sampling fraction is small in most strata. Application of this method to cases in which PSUs are selected without replacement or with a non-negligible sampling fraction will lead to an overestimation of the variance. Although conservative estimates (that is, overestimates) are preferable to liberal estimates of the variance, being overly conservative could be problematic as well.

4. Comparison of the Yates-Grundy and bootstrap estimators

In order to measure the extent of the bootstrap variance overestimation, a comparison was done between the bootstrap estimator and a theoretical estimator of the variance for the PALS sample design. For this comparison some simplifying assumptions about the second stage sample design were made. The first stage of the PALS sample design is a stratified SRS of small PSUs, a stratified PPS sample of medium PSUs and a *take-all* sample of large PSUs. At the second stage, it will be assumed that the PALS sample design can be approximated by a stratified SRS of disabled individuals. It is in fact a stratified systematic sample of households. This last assumption is reasonable if the disabled population is distributed uniformly throughout each PSU. The comparison was done using only the sampling weights and therefore, do not take into account nonresponse and post-stratification adjustments.

Omitting the sub-stratum subscript, the variance estimate for the PSU total Y_i for an SRS of

m_i from M_i disabled individuals <u>within</u> PSU i is given by:

$$v_w\left(\hat{Y}_i\right) = M_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)s_i^2$$

where s_i^2 is the within PSU sample variance. For all PSUs of a given sub-stratum, an estimate of the within PSU variance for an estimated total $\hat{Y}$ is given by:

$$v_w\left(\hat{Y}\right) = \sum_{i\in s}\frac{1}{\pi_i}M_i^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)s_i^2 \qquad (1)$$

where π_i is the inclusion probability of the i^{th} PSU. A between PSU variance estimator for any sample of PSUs under any fixed-size design at the first stage can be obtained from the Yates-Grundy (1953) estimator.

$$v_b\left(\hat{Y}\right) = \frac{1}{2}\sum_{i\neq j}\sum\frac{\left(\pi_i\pi_j - \pi_{ij}\right)}{\pi_{ij}}\left(\frac{\hat{Y}_i}{\pi_i} - \frac{\hat{Y}_j}{\pi_j}\right)^2 \qquad (2)$$

Combining (1) and (2), an estimate of the variance for an estimated total $\hat{Y}$, for a particular sub-stratum is given by

$$v_1\left(\hat{Y}\right) = v_b(\hat{Y}) + v_w(\hat{Y})$$

where π_{ij} is the joint inclusion probability for the i^{th} and j^{th} PSUs. This result can be found in Särndal (1992) for instance.

In sub-strata composed of small PSUs, $v_b(\hat{Y})$ reduces to the variance of a simple random sample of PSUs. In sub-strata composed of medium PSUs, PPS sampling was done through the use of PROC SURVEYSELECT of SAS with the procedure of Hanurav (1967) and Vijayan (1968). The joint probabilities for this PPS sampling scheme are calculated by PROCSURVEY SELECT. In sub-strata composed of *take-all* PSUs, $v_b(\hat{Y})$ becomes zero because the sample design for these sub-strata reduces to a single-stage design with units selected by stratified simple random sampling.

For the purpose of this comparison, some simplifications were made. PSUs composed only of survey non-respondents or PSUs including only respondents with no limitation (in both cases $\hat{Y}_i = 0$) were assigned a zero within PSU variance in the expression of $v_w(\hat{Y})$. Obviously, PSUs with disabled individuals but for which no one has the characteristic of interest have also a null within PSU variance. Since the size M_i of each PSU is unknown, an estimate $\hat{M}_i$ was used from the long-form Census sample.

For linear statistics such as totals, means, et cetera, the bootstrap variance formula can be approximated by the variance formula of a PPS with-replacement sample of PSU at the first stage with any design at the second stage.

$$v_2\left(\hat{Y}\right) = \frac{1}{n(n-1)}\sum_{i\in s}\left(\frac{\hat{Y}_i}{p_i} - \hat{Y}\right)^2 \qquad (3)$$

where p_i is the probability of selection of the i^{th} PSU. See, for instance, Cochran (1977) for the formula of $v_2\left(\hat{Y}\right)$.

For linear statistics, both variance estimate formula are unbiased for the true variance $V\left(\hat{Y}\right)$ but are not equivalent. The precision of the bootstrap estimator approaches the one of $v_2\left(\hat{Y}\right)$as the number of bootstrap samples tends to infinity (Rust and Rao, 1996). Since both estimators are unbiased, the bootstrap variance estimate will also approach $v_2\left(\hat{Y}\right)$ for an infinite number of bootstrap samples. Since our comparison was restricted to linear statistics, it was not necessary to generate the bootstrap samples. Instead, the bootstrap variance formula in a particular sub-stratum of small or medium PSUs was approximated by the expression in (3).

For small PSUs (drawn by SRS), $p_i = 1/N$ and the formula reduces to the variance of a simple with-replacement random sample of PSUs. For medium PSUs, p_i is the relative size of the PSU within the sub-stratum. For large PSUs (take-all), the formula is replaced by the variance of a simple with-replacement random sample of m from M disabled individuals.

Variance estimates $v_1\left(\hat{Y}\right)$ and $v_2\left(\hat{Y}\right)$ were calculated for several PALS variables, such as being disabled, using a hearing aid, blindness, difficulty speaking, et cetera. Since the prevalence of these limitations differs substantially from one limitation to another, variance estimates were compared in terms of coefficients of variation (CV), which is the standard error of the estimate divided by the estimate itself (also called relative standard error). A regression analysis was performed to summarize the results predicting $cv_2\left(\hat{Y}\right)$ from $cv_1\left(\hat{Y}\right)$. In order to evaluate the extent of the overestimation of the bootstrap estimator for strata with large PSU sampling fractions, a first regression was done for these specific strata and a second regression was done on the remaining strata. The term "strata" here corresponds to the cross-classification of province, age group and severity of limitation. Therefore, it combines all PSU size sub-strata. Large sampling fractions were defined as sampling fractions (*f*) larger than 20%, which constitutes 28% of the strata. A third regression was done for all strata combined. Each unit in the regression corresponds to a particular stratum and a particular estimate. Results are shown in Table 1 below.

Table 1. Results of the regression
$CV_{boot} = \beta_0 + \beta_1 CV_{yates}$

Model	β_0	β_1	R^2
Strata with high sampling fraction (>20%)	0.0157	0.9502	0.9669
Other strata	0.0065	0.9719	0.9770
All strata	0.0089	0.9657	0.9740

The relationship between the CVs obtained from the bootstrap and Yates-Grundy estimators is quite strong for the three analyses, with R^2 values above 96%. The slope is significantly different from one in each case and the intercept is also significantly different from zero in each case (not shown in Table 1). Although $cv_2(\hat{Y})$ is higher than $cv_1(\hat{Y})$ for strata with large sampling fractions, particularly for lower CVs, the differences are not large globally. It seems that $cv_2(\hat{Y})$ is not systematically higher than $cv_1(\hat{Y})$, since the intercept is greater than zero but the slope is lower than one. Under the fitted models, the positive intercept combined with the slope slightly smaller than 1 indicates that $cv_2(\hat{Y})$ is larger than $cv_1(\hat{Y})$ especially for small to moderate CVs. For large CVs, $cv_2(\hat{Y})$ is on average slightly smaller than $cv_1(\hat{Y})$. This apparent anomaly could be explained by the large variability of the estimates $cv_1(\hat{Y})$ and $cv_2(\hat{Y})$ for large CVs. These cases correspond to estimates of rare characteristics. It could also be partially due to the simplifying assumptions that were made. As expected, it was also found, that on average, the higher the sampling fraction at the PSU level the larger the positive difference between $cv_2(\hat{Y})$ and $cv_1(\hat{Y})$. This seems to indicate that the larger the sampling fraction, the larger is the magnitude of the overestimation with the bootstrap method. It should be noted, however, that the bootstrap estimator was compared to the Yates Grundy estimator, which is not the true variance. It is rather a "reasonable" direct estimate of the true variance which does not assume that the PSUs are drawn with replacement. Therefore, the term "overestimation" should be used with caution since two estimates are compared.

Figure 1 presents the fitted regression on all strata.

5. Adjustment of bootstrap weights for nonresponse

As mentioned in Section 3, the full-sample weights were adjusted for nonresponse through the use of a response propensity model, a logistic regression model that predicts the response probability. A criterion used in the logistic regression was to minimize the Hosmer and Lemeshow (HL) statistic (Hosmer, D.W. Jr. & Lemeshow, S., 1989). This method subdivides the individuals (respondents and non-respondents) into usually 10 groups of approximately the same size (method of equal deciles) based on the predicted probabilities of response. The 10% of individuals with the lowest predicted probabilities of response form the first class, the next 10% with the lowest predicted probabilities of response form the second class, and so on. The HL statistic is based on the differences in each class between the sum of the predicted probabilities of response and the observed response rate within each class. If the two quantities are equal in every class, the HL statistic takes a value of 0. The HL statistic is approximately distributed as a chi-square with *g-2* degrees of freedom, where *g* is the number of classes formed. This statement is valid under the condition that the number of cells defined by the cross-classification of the different explanatory variables is approximately equal to the sample size (no replication within the cells).

This step involves quite a bit of modelling to find the most appropriate model. It requires fitting many different models possibly including different interaction terms, et cetera. Since this aspect requires manual intervention, it is not operationally possible to completely remodel each bootstrap sample. On the other hand, since all bootstrap samples come from the same main sample, they should follow the model used for the main sample. The second-best option would be to re-estimate the parameters of the same model on each bootstrap sample (Method A). New parameter estimates would be produced for each bootstrap sample, leading to new nonresponse classes and new adjustment factors. Since the fitting of a logistic regression model is an iterative process, this method still can be quite computer intensive when conducted on 500 different samples. Moreover, there is no guarantee that convergence of the parameter estimates will be attained for all bootstrap samples. If convergence is not attained for some of the bootstrap samples, the initial model has to be somewhat simplified to accommodate all bootstrap samples, which means that the nonresponse adjustment model may not be optimal for the main sample.

An alternative to this method (Method B) would be to maintain fixed nonresponse classes over all bootstrap samples. The bootstrap samples being different from one another, the number of respondents and non-respondents within each class would be different and different nonresponse adjustment factors would be applied to each bootstrap sample. How much do we lose by using this alternative strategy? In particular, how much will variance estimates vary between Method A and Method B?

A comparative study was done to measure the differences. It should be noted that the post-stratification adjustment was done on each bootstrap sample following the two nonresponse adjustments of Method A and Method B. The CVs of certain key statistics are compared in Table 2. This table presents estimates of children aged 5 to 14 with vision disability. The table is broken down by severity level. The CVs for both methods and the relative differences are presented.

Table 2. Estimates and CVs of children aged 5-14 with vision disability

Severity	Estimate	CV(A)	CV(B)	Relative Diff.
Mild	1400	20.58%	20.48%	0.45%
Moderate	4200	15.57%	15.61%	-0.30%
Severe	3900	13.66%	13.65%	0.11%
very severe	5000	13.91%	13.87%	0.27%

As can be seen from this study, for PALS, it is not necessary to perform the nonresponse weight adjustments in the bootstrap sample in exactly the same way as this was done for the full sample. In this example, the relative difference between the CVs obtained from both methods is clearly negligible, the maximum relative difference being less than 0.5%. Other tables not included in this paper showed comparable relative differences. Richard Valliant (2002) presents an excellent paper on the effect of multiple weighting steps on variance estimation. His study compares different variance estimators such as replicated variance estimators (various forms of the jackknife) and linearization methods that account for some or all of the weighting steps in their derivation. The replication variance methods have the advantage of being able to take into account of all the weighting steps by repeating each adjustment on each replicate. His paper also includes a number of very useful references on the topic.

6. Summary and conclusion

Specific challenges to the application of the bootstrap variance estimation method to PALS were presented in this paper. In particular, the relatively high PSU sampling fraction in some of the strata could cause the bootstrap method to overestimate the variance. A study was done to compare the bootstrap estimate to an approximate direct variance estimate using the Yates-Grundy estimator. A few simplifications had to be done to make this comparison. Results indicated that although the bootstrap variance estimator tends to produce a larger variance estimate than the one obtained from the Yates-Grundy formula, the difference is relatively small. This seems to indicate that the magnitude of the overestimation associated to the bootstrap estimator is relatively small in our situation. Here, the term "overestimation" should be used with caution since the Yates-Grundy formula is not the true variance but rather an estimate of it which does not assume with replacement of the PSUs. Therefore, the bootstrap method was judged appropriate for PALS. In this case, it was felt that the variance estimates were in general only slightly conservative, which is preferable to the opposite. Of course, these results are limited to our particular application. The extent of the over-estimation will depend of course on the magnitude of the PSU sampling fraction in each stratum.

Another challenge in using the bootstrap method was the fact that a response propensity model was used for the nonresponse adjustment of the main sample. Re-estimating the parameters of the same logistic regression model on each bootstrap sample can be quite computer intensive and convergence may not be obtained on all bootstrap samples. This may require simplifying the initial logistic regression model to accommodate all bootstrap samples. The drawback is a loss of optimality for the nonresponse adjustment of the full sample. The study showed that instead of re-estimating the same logistic regression model on each bootstrap sample, an alternative approach where the nonresponse classes are fixed over all bootstrap samples gives almost exactly the same results in terms of variance estimation. Therefore, the alternative approach is recommended to avoid excessive computer time and possible convergence problems. Moreover, a more optimal model can be used for the main sample.

References

Cochran, W.G. (1977). *Sampling Techniques*, New York: John Wiley & Sons, 254.

Hanurav, T.V. (1967), "Introduction to Survey Sampling", *Sage University Paper series on Quantitative Applications in the Social Sciences*, series no. 01-035, Beverly Hills and London: Sage Publications Inc.

Hosmer, D.W. Jr. & Lemeshow, S, (1989). *Applied Logistic Regression*, New York: John Wiley & Sons.

Langlet, E.R. (2000), "Estimation of Proxy, Time and Content Effect for the Health and Activity Limitations Survey," *2000 Proceedings of the American Statistical Association*, Section on Survey Research Methods, Alexandria, VA: American Statistical Association, 89-97.

Rao, J.N.K. and Wu, C.F.J. (1988), "Resampling Inference With Complex Survey Data," *Journal of the American Statistical Association*, March 1988, Vol. 83, No 401, Theory and Methods.

Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992), "Some Recent Work on Resampling Methods for Complex Surveys," *Survey Methodology*, Vol. 18, 2, December 1992.

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

Sarndäl C.-E., B. Swensson and J. Wretman (1992). *Model Assisted Survey Sampling*, Springer-Verlag, New-York.

SAS Institute Inc. (1999), *SAS/STAT User's Guide, Version 8*, Vol. 2, 3273-3320.

Valliant, R. (2002), "The effect of multiple weighting steps on variance estimation," *2002 Proceedings of the American Statistical Association*, Section on Survey Research Methods [CD-ROM], Alexandria, VA: American Statistical Association, 3547-3552.

Vijayan, K. (1968), "An exact PPS Sampling Scheme: Generalization of a method of Hanurav," *Journal of the Royal Statistical Society*, Series B, 30, 556-566.

Yates, F., and Grundy, P.M. (1953), "Selection without replacement from within strata with probability proportional to size," *Journal of the Royal Statistical Society*, Series B, 15, 235-261.

Yeo, D., Mantel, H., and Liu, T.P. (1999), "Bootstrap Variance Estimation for the National Population Health Survey", *1999 Proceedings of the American Statistical Association*, Section on Survey Research Methods, Alexandria, VA: American Statistical Association, 778-783.

Figure 1. Regression on all the strata