

COMPARISON OF CROSS-SECTIONAL ESTIMATES FROM TWO WAVES OF A LONGITUDINAL SURVEY

Georgia Roberts and Milorad Kovačević¹

ABSTRACT

The National Population Health Survey (NPHS) has been designed to measure the health status of Canadians over time. In order to accomplish this, a longitudinal sample of people is being followed for up to 20 years. However, it is possible to make use of data from the longitudinal sample and additional sample members at each time point to make cross-sectional estimates of certain health measures. For waves 1 and 2 of NPHS, it has been possible to make cross-sectional estimates of the prevalences of some of the same conditions or diseases. When comparing the estimates from the two different time points, a difficulty arises in estimating the variance of their difference because of the overlap in the samples. This paper describes two approaches to estimating the variance of the difference in cross-sectional prevalence rates from two waves of a longitudinal survey. The proposed approaches rely heavily on the particular composition of the two cross-sectional samples. The research is designed to suit the NPHS; however, the results are applicable to a variety of longitudinal surveys.

KEY WORDS: Variance estimation; longitudinal survey; cross-sectional estimates.

RÉSUMÉ

L'Enquête nationale sur la santé de la population (ENSP) a été conçue pour mesurer l'état évolutif de la santé des Canadiens. Afin d'accomplir ceci, un échantillon longitudinal de personnes est suivi pendant une période allant jusqu'à 20 ans. Cependant, il est possible d'utiliser des données de l'échantillon longitudinal et de l'échantillon de membres supplémentaires à chaque occasion pour effectuer des estimations transversales de certaines mesures de la santé. Pour les première et deuxième vagues de l'ENSP, il a été possible de faire des estimations transversales des prévalences des mêmes conditions ou maladies. Lorsqu'on compare les estimations de deux points temporels distincts, une difficulté surgit lors de l'estimation de la variance de la différence dans les estimations entre deux points temporels en raison de la superposition des échantillons. Cet article décrit deux approches d'estimation de la variance de la différence dans les taux de prévalence transversaux de deux vagues d'une enquête longitudinale. Les approches proposées reposent fortement sur la composition particulière des deux échantillons transversaux. Le plan de recherche est conçu pour répondre à l'ENSP; cependant les résultats s'appliquent à une variété d'enquêtes longitudinales.

MOTS-CLÉS : Estimation de la variance; enquête longitudinale; estimations transversales.

1. INTRODUCTION

The National Population Health Survey (NPHS) has been designed to measure the health status of Canadians over time. A longitudinal sample of people is being followed every two years for up to 20 years. However, because of considerable interest in measuring status at particular points in time, it is possible to make use of data from a "cross-sectional sample" consisting of in-scope longitudinal individuals and, for some provinces, additional sample members (such as single-time provincial buy-ins), to make cross-sectional estimates of certain health measures. For waves 1 and 2 of NPHS, it has been possible to make cross-sectional estimates of

the prevalences of some of the same conditions or diseases since the same questions were asked at both time points. As a next step, it would then seem natural to make inference about the difference in the population prevalence rates at the two time points; such inference would involve the difference in the estimated prevalence rates and the estimated variance of this difference. A difficulty arises in estimating the variance of the difference because of the overlap in the samples at the two time points. This paper describes a Taylor linearization and a bootstrap approach to estimating the variance of the difference of prevalence rates estimated from particular cross-sectional samples. The proposed approaches rely heavily on the particular composition of

¹ Georgia Roberts and Milorad Kovacevic, Data Analysis Support and Research Group, Social Survey Methods Division, 15th floor, RH Coats Bldg., Ottawa, Ontario K1A 0T6. The authors would like to thank Harold Mantel for helpful suggestions.

two cross-sectional samples of the National Population Health Survey (NPHS); however, the ideas are applicable to a variety of longitudinal surveys.

Section 2 of the paper presents the problem in detail and describes the reasons why there is a problem in estimating the variance of the estimated difference in prevalence rates. In Section 3, a Taylor linearization approach is taken for variance estimation, and under this approach, solutions to various complications are proposed. Finally, in Section 4, the same problem is attacked through a bootstrap variance estimation approach, and, again, a variety of solutions is described.

2. DESCRIPTION OF THE PROBLEM

2.1 What is the problem?

In 1994/1995, Statistics Canada collected data for cycle one of the National Population Health Survey (NPHS), and then collected data for cycle two in 1996/1997. Since the main objective of this survey is to provide measures of health status over time, many of the same variables were collected at both time points. This allows the estimation at both time points of the prevalence of conditions or diseases that were defined from these variables.

One data file at each time point is designated for cross-sectional analysis of variables that were collected on the in-depth health questionnaire which was administered just to a subsample of the individuals who responded to the general component of the questionnaire. In this paper these particular cross-sectional data files will be referred to as the cross-sectional health files.

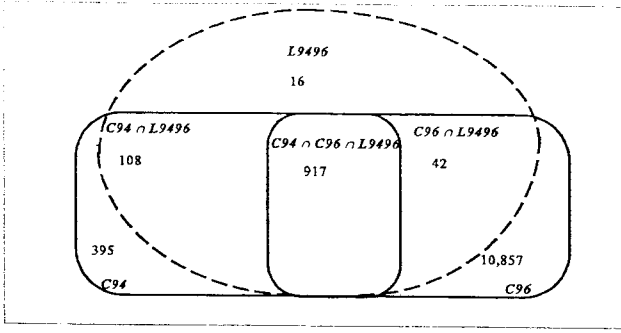
Let s_t be the sampled individuals on the cross-sectional health file at time t , where $t=1$ for cycle one (1994/1995) and $t=2$ for cycle 2 (1996/1997). Suppose that we are interested in the prevalence of a particular condition within a domain of the population at each time point, where the existence of the condition for each of the individuals on each file can be determined from responses to questions on the in-depth health questionnaire. The prevalence rate of the condition within the domain at time t may be estimated by $\hat{\theta}_t = \hat{Y}_t / \hat{X}_t$, with $\hat{Y}_t = \sum_{s_t} w_{it} y_{it}$ and $\hat{X}_t = \sum_{s_t} w_{it} x_{it}$, where w_{it} is the cross-sectional weight of the i th individual in s_t (who will be called the it th individual); $y_{it} = 1$ if the it th individual is in the domain and also has the condition, and $y_{it} = 0$ otherwise; and $x_{it} = 1$ if the it th individual is in the domain, and $x_{it} = 0$ otherwise.

Then $\hat{\Delta} = \hat{\theta}_1 - \hat{\theta}_2$ estimates the net change in the prevalence rate between the two time periods. The main problem addressed in this paper is the estimation of the variance of $\hat{\Delta}$.

2.2 Why is this a problem?

While the NPHS was particularly designed as a longitudinal survey which follows many of the same individuals over time, it also has cross-sectional capabilities. However, the longitudinal underpinnings of the survey contain complexities that cause difficulties when it comes to the estimation of the variance of $\hat{\Delta}$. Some of these complexities, which can vary from province to province, are the following:

- (i) Each of the cross-sectional health files contains all longitudinal individuals who are in-scope for cross-sectional purposes. The cross-sectional samples at the two time points are thus not independent, and actually have a large degree of overlap.
- (ii) At each of the time points, the cross-sectional health files also contain single-time supplementary samples in some of the provinces, that were chosen by random digit dialling (RDD). This adds the complexity of dual frames.
- (iii) The cross-sectional health file for the second time point contains longitudinal individuals who changed province of residence after their selection into the longitudinal sample. For cross-sectional purposes, though, these individuals are considered to be part of the sample for the province in which they reside at the time of the cross-section.
- (iv) The in-depth health questionnaire was administered just to individuals aged 12+ at the first time point. Thus, any domain to be studied at both time points cannot contain people less than 12 years old. The cross-sectional health file for the second time point contains children aged 12 and 13 years who were chosen for the longitudinal sample but who do not appear on the cross-sectional health file for the first time point. Counterintuitively, this means that the longitudinal sample is not entirely within the intersection of the two cross-sectional health files.
- (v) The cross-sectional health files are affected by longitudinal non-response because many longitudinal individuals are included in the cross-sectional files.



Venn diagram of cross-sectional (C94 and C96) and longitudinal (L9496) samples aged 12+ in Manitoba

Approaches to dealing with some of these complexities when estimating the variance of $\hat{\Delta}$ are discussed in the later sections of this paper.

The chart above illustrates several of these complexities, through graphical presentation of the number of individuals included in the cross-sectional and longitudinal health files for the province of Manitoba. Of the 1420 individuals on the 1994 cross-sectional file, the 11,816 on the 1996 cross-sectional file, and the 1083 individuals on the longitudinal file, 917 individuals are common to all three files. In the 1994 cross-sectional file, 395 individuals are supplementary sample. There are 108 longitudinal individuals in the 1994 cross-sectional sample who are no longer in scope to be cross-sectional health individuals for Manitoba in 1996 for such reasons as institutionalization, death, and nonresponse to the health questionnaire in 1996; also, some of these individuals may have moved from Manitoba between 1994 and 1996 and could be 1996 cross-sectional individuals in another province. On the other hand, the 42 longitudinal individuals who are in the 1996 cross-sectional sample but are not in the 1994 one, could be nonrespondents to the health questionnaire in 1994 or children who became 12+ in 1996. There are 16 longitudinal individuals who are not in either of the cross-sectional samples. These are children that are selected as longitudinal individuals in 1994 in Manitoba but are not considered as cross-sectional due to age, and in 1996 they moved out-of-scope or out of Manitoba, so that they are not considered as cross-sectional there. The 10,857 individuals who are only in the 1996 cross-sectional file are mainly supplementary sample, although some are longitudinal individuals who moved from another province into Manitoba between 1994 and 1996.

3. VARIANCE ESTIMATION: TAYLOR LINEARIZATION APPROACH

3.1 Linearization of $\hat{\Delta}$

One possible approach to obtaining a design-based variance estimate of $\hat{\Delta}$ is Taylor linearization. In developing this approach, for ease of presentation, adjustments to the final weights will be ignored. Since $\hat{\Delta}$ is a non-linear function of the data from both files s_t , $t=1,2$, the first step is to linearize $\hat{\Delta}$ by expansion into a Taylor series around the true net change in prevalence. Assuming that the remainder term is negligible for a sufficiently large sample, the following approximation holds:

$$\hat{\Delta} \approx \Delta + \frac{1}{X_1} \left[(\hat{Y}_1 - Y_1) - \theta_1 (\hat{X}_1 - X_1) \right] - \frac{1}{X_2} \left[(\hat{Y}_2 - Y_2) - \theta_2 (\hat{X}_2 - X_2) \right], \quad (1)$$

where $\theta_t = Y_t / X_t$, $t=1,2$. This implies that

$$\text{Var}(\hat{\Delta}) \approx \text{Var} \left\{ \frac{1}{X_1} \sum_{s_1} w_{1i} (y_{1i} - \theta_1 x_{1i}) - \frac{1}{X_2} \sum_{s_2} w_{2i} (y_{2i} - \theta_2 x_{2i}) \right\}. \quad (2)$$

Sample s_t can be expressed as $s_t = \bigcup_{k=1}^{10} s_{tk}$ where s_{tk}

represents those observations in s_t forming the cross-sectional sample for province k at time t . It then follows that

$$\text{Var}(\hat{\Delta}) \approx \text{Var} \left[\sum_{k=1}^{10} \hat{Z}_1(s_{1k}) - \sum_{k=1}^{10} \hat{Z}_2(s_{2k}) \right], \quad (3)$$

where $\hat{Z}_t(s_{tk}) = \sum_{s_{tk}} w_{ti} Z_{ti}$,

and $Z_{ti} = X_t^{-1} (y_{ti} - \theta_t x_{ti})$, $t=1,2$.

If we ignore, for the moment, longitudinal individuals that changed provinces between the two time points, the provincial samples s_{tk} are independent. This then implies that

$$\text{Var}(\hat{\Delta}) \approx \sum_{k=1}^{10} \text{Var} \left[\hat{Z}_1(s_{1k}) - \hat{Z}_2(s_{2k}) \right]. \quad (4)$$

The k -th provincial component of this variance, $\text{Var} \left[\hat{Z}_1(s_{1k}) - \hat{Z}_2(s_{2k}) \right]$, can be expanded further as

$$\begin{aligned} \text{Var}[\hat{Z}_1(s_{1k}) - \hat{Z}_2(s_{2k})] &= \text{Var}[\hat{Z}_1(s_{1k})] \\ &+ \text{Var}[\hat{Z}_2(s_{2k})] - 2\text{Cov}[\hat{Z}_1(s_{1k}), \hat{Z}_2(s_{2k})]. \end{aligned} \quad (5)$$

The problem of estimating the variance of $\hat{\Delta}$ then reduces to estimating the terms on the right hand side of (5).

3.2 Notation and Assumptions Required for Variance Estimation

The following detailed notation is required for explanation of the variance estimation:

H_{tk} = # of strata in the cross-sectional sample in province k at time t ,

n_{tkh} = # of sampled clusters in the h th stratum in province k at time t ,

n_{tkhc} = # of individuals on data file in c th cluster of h th stratum in province k at time t ,

w_{tkhci} = weight on the i th individual in c th cluster of h th stratum in province k at time t , and

$$z_{tkhci} = \hat{X}_t^{-1} (y_{tkhci} - \hat{\theta}_t x_{tkhci}).$$

As well, the following standard assumptions for variance estimation for data from a survey with a stratified multi-stage design are considered to hold for each of the cross-sectional health samples:

- (i) The design of each cross-sectional sample is approximately stratified with selection of psu's with replacement.
- (ii) Each psu is selected at most once (because of small sampling fractions).
- (iii) $n_{tkh} \sum_{i=1}^{n_{tkhc}} w_{tkhci} z_{tkhci} = n_{tkh} z_{tkhc}$ (i.e. $n_{tkh} \times$ weighted cluster total) is approximately unbiased as an estimator for the stratum total Z_{tkh} for any z variable and for any value of t, k, h , and c .

Under these assumptions, there is a straightforward approach to estimate a stratum total and the variance of stratum total at each time point. As well, if the same psu's are represented in a stratum at both time points, there is a straightforward approach to estimating a covariance between stratum totals at the two time points. In particular, under these assumptions:

- (i) An (approximately) unbiased estimate for Z_{tkh} is $\hat{Z}_{tkh} = \sum_{c=1}^{n_{tkhc}} z_{tkhc}$.

- (ii) An (approximately) unbiased estimate of the variance of \hat{Z}_{tkh} is

$$v\hat{a}r[\hat{Z}_{tkh}] = n_{tkh} / (n_{tkh} - 1) \sum_{c=1}^{n_{tkhc}} (z_{tkhc} - \bar{Z}_{tkh})^2, \quad (6)$$

where $\bar{Z}_{tkh} = \hat{Z}_{tkh} / n_{tkh}$.

- (iii) If, at times $t=1$ and $t=2$, the same psu's are observed in a stratum sample, (which implies that $n_{1kh} = n_{2kh}$), an (approximately) unbiased estimate of the covariance of \hat{Z}_{1kh} and \hat{Z}_{2kh} is given by

$$\begin{aligned} c\hat{o}v[\hat{Z}_{1kh}, \hat{Z}_{2kh}] &= n_{1kh} / (n_{1kh} - 1) \sum_{c=1}^{n_{1kh}} \\ &(z_{1khc} - \bar{Z}_{1kh})(z_{2khc} - \bar{Z}_{2kh}). \end{aligned} \quad (7)$$

These results can then be readily applied to the various types of cross-sectional health samples found in the different provinces. The following 2 subsections illustrate the applications to situations which would be reasonable approximations for the different provinces.

3.3 Variance Estimation for Provinces with the Same Strata and Psu's at Each Time Point

For provinces other than British Columbia, Ontario, Manitoba and Alberta, it would be reasonable to expect that the cross-sectional health files for the two time points would consist of the same strata and psu's, since the two cross-sectional health samples consisted totally of the longitudinal respondents to the in-depth health questionnaire who were cross-sectionally in scope or (for New Brunswick) consisted of the in-scope longitudinal respondents plus some supplemental sample at time 1, where the supplements were chosen from the same psu's as the longitudinal sample. While the psu's are the same, there are several reasons to expect that the individuals within a particular psu would not be exactly the same at the two time points, such as nonresponse of a longitudinal individual to the health questionnaire at one of the time points or a longitudinal person entering an institution between the two time points. The following variance and covariance estimates would follow in a straightforward manner from the results of the previous subsection:

$$v\hat{a}r[\hat{Z}_t(s_{tk})] = \sum_{h=1}^{H_{tk}} n_{tkh} / (n_{tkh} - 1) \sum_{c=1}^{n_{tkhc}} (z_{tkhc} - \bar{Z}_{tkh})^2, \quad (8)$$

$$c\hat{o}v[\hat{Z}_1(s_{1k}), \hat{Z}_2(s_{2k})] = \sum_{h=1}^{H_{1k}} n_{1kh} / (n_{1kh} - 1) \sum_{c=1}^{n_{1kh}} (z_{1khc} - \bar{Z}_{1kh})(z_{2khc} - \bar{Z}_{2kh}), \quad (9)$$

while z_{1khc} and z_{2khc} would consist of weighted sums over different individuals if the khc -th psu contained different individuals at the two time points.

3.4 Variance Estimation for Provinces with Supplemental Sample at the Second Time Point

Ontario, Manitoba and Alberta had extensive supplemental samples at the second time point, where the design of these additional samples was RDD. The cross-sectional sample at time 2, s_{2k} , is thus made up of two pieces:

- (a) a longitudinal portion, s_{2Lk} , with the same strata and psu's as the cross-sectional sample at time 1; and
- (b) an RDD portion, s_{2Rk} , with independent strata and psu's.

Since it is reasonable to assume that the frames for the two portions were both exhaustive and that no individuals were chosen for both samples, then $\hat{Z}_2(s_{2k})$ is approximately equal to $\gamma \hat{Z}_2(s_{2Lk}) + (1-\gamma) \hat{Z}_2(s_{2Rk})$, where γ is a constant arising from the use of an adaptation of the dual-frame technique of Skinner and Rao (1996) and where the weights assumed in the definitions of $\hat{Z}_2(s_{2Lk})$ and $\hat{Z}_2(s_{2Rk})$ are those that would be available if s_{2Lk} or s_{2Rk} respectively was the only sample available for cross-sectional estimation at time 2. (The dual-frame approach proposed by Skinner and Rao is suited to the situation of dual frames where at least one of the samples is drawn by a complex design, and has the property of using the same weights for all estimates.) It then follows that the Skinner-Rao dual frame approach may be used to estimate the variance of $\hat{Z}_2(s_{2k})$. Also, because of the selection of the RDD sample at time 2 independently from the longitudinal sample, $Cov[\hat{Z}_1(s_{1k}), \hat{Z}_2(s_{2k})] = Cov[\hat{Z}_1(s_{1k}), \gamma \hat{Z}_2(s_{2Lk})]$. This covariance can be estimated in the same manner as in Section 3.3 since the γ multiplier is actually incorporated into the final cross-sectional weights for time 2 on s_{2k} . The fact that the cross-sectional samples of Ontario and Manitoba at time 1 contained supplemental sample within the same psu's as the longitudinal sample does not contribute additional complications to this covariance estimation, as noted in Section 3.3. The estimation of the variance of $\hat{Z}_1(s_{1k})$ is straightforward for these three provinces. However, since for all three provinces, $s_{2Rk} \gg s_{2Lk}$, the covariance may not be large and thus consideration could be given to ignoring it.

Variance and covariance estimation for British Columbia

would use a similar approach as that for Ontario, Manitoba and Alberta, except that the RDD supplemental sampling in British Columbia occurred at time 1 rather than time 2, and was restricted to one health unit rather than covering the full province.

3.5 Accounting for Movers between Provinces

Also needing consideration for the Taylor linearization variance approach are movers, that is, longitudinal people who are on the cross-sectional health files for both time points, but who changed province between those time points.

Since a design-based variance measures the variability in an estimate due to the selection of the sample, then, for variance estimation, individuals in the cross-sectional files who are part of the longitudinal sample must be associated with their "design" province, which was their province of residence at time 1. This is analogous to redefining the s_{tk} in (3) as the sample at time t that was selected in province k at time 1.

4. VARIANCE ESTIMATION: BOOTSTRAP APPROACH

4.1 Using currently-available bootstrap weights

A bootstrap approach to variance estimation for the NPHS is being supported by Statistics Canada through the provision of a file of 500 bootstrap weights to accompany each data file. Separate files of such weights are thus available for the cross-sectional health files for times 1 and 2, and for the longitudinal file for the two time points.

Each file is the result of independently drawing $B=500$ bootstrap samples from the individuals on these files through a bootstrap procedure for stratified multistage samples where, for a single replicate, for each stratum h , a simple random sample of $n_h - 1$ PSU's is chosen with replacement from the n_h PSU's in the sample.

By definition,

$$Var[\hat{\Delta}] = Var[\hat{\theta}_1] + Var[\hat{\theta}_2] - 2Cov[\hat{\theta}_1, \hat{\theta}_2]. \quad (10)$$

The first two components on the right hand side of (10) can be estimated readily using the bootstrap weight files for the two cross-sectional health files. For estimating the third component, it must be kept in mind that covariance between $\hat{\theta}_1$ and $\hat{\theta}_2$ only comes from the overlap in the samples on which the two different estimates are based, and that all individuals in the

overlap are from the longitudinal sample. The challenge is to express both estimates in terms of longitudinal weights and sums over the longitudinal sample, so that the bootstrap weights for the longitudinal sample can be used for estimating the covariance term. First, through linearization, we approximate $\hat{\theta}_t - \theta_t$ by

$$\begin{aligned} \hat{\theta}_t - \theta_t &\approx \sum_{i \in s_t \cap s_L} w_{ii} X_t^{-1} (y_{ii} - \theta_t x_{ii}) \\ &+ \sum_{i \in s_L \cap s_t} w_{ii} X_t^{-1} (y_{ii} - \theta_t x_{ii}) \\ &= \sum_{i \in s_t \cap s_L} w_{ii} Z_{ii} + \sum_{i \in s_L \cap s_t} w_{ii} Z_{ii} \\ &= \varepsilon_t(s_t \cap s_L) + \varepsilon_t(s_L \cap s_t), \end{aligned} \quad (11)$$

where s_L is the longitudinal sample. Then, using (11), and assuming that $s_L, s_1 \cap s_L$, and $s_2 \cap s_L$ are independent,

$$\begin{aligned} \text{cov}(\hat{\theta}_1, \hat{\theta}_2) &\approx \text{cov}(\varepsilon_1(s_L \cap s_1), \varepsilon_2(s_L \cap s_2)) \\ &= \text{cov} \left[\sum_{i \in s_L \cap s_1} w_{Li} (w_{1i}/w_{Li}) Z_{1i}, \right. \\ &\quad \left. \sum_{i \in s_L \cap s_2} w_{Li} (w_{2i}/w_{Li}) Z_{2i} \right] \\ &= \text{cov} \left(\sum_{i \in s_L \cap s_1} w_{Li} \varepsilon_{1i}^*, \sum_{i \in s_L \cap s_2} w_{Li} \varepsilon_{2i}^* \right), \end{aligned}$$

where $\varepsilon_{ii}^* = (w_{ii}/w_{Li}) Z_{ii}$. Finally, the b th bootstrap replicate is $\hat{\varepsilon}_t^{(b)} = \sum_{i \in s_L \cap s_t} w_{Li}^{(b)} \hat{\varepsilon}_{ii}^*$ and the bootstrap estimate of covariance is

$$\begin{aligned} B^{-1} \sum_{b=1}^B (\hat{\varepsilon}_1^{(b)} - \hat{\varepsilon}_1(s_L \cap s_1)) (\hat{\varepsilon}_2^{(b)} - \hat{\varepsilon}_2(s_L \cap s_2)), \quad \text{where} \\ \hat{\varepsilon}_t(s_L) = \sum_{i \in s_L \cap s_t} w_{Li} \hat{\varepsilon}_{ii}^* \text{ and} \\ \hat{\varepsilon}_{ii}^* = (w_{ii}/w_{Li}) \hat{Z}_{ii} = (w_{ii}/w_{Li}) \hat{X}_t^{-1} (y_{ii} - \hat{\theta}_t x_{ii}). \end{aligned}$$

A second possible approach to using the currently-available bootstrap weights is to ignore the dependence in the samples due to the overlap, and to randomly pair bootstrap weights for the two cross-sectional health files. The b th bootstrap estimate of Δ is constructed as the difference in the estimates of θ_1 and θ_2 derived from the corresponding bootstrap weights in the b th pair. The separate estimates of Δ would be combined in the usual manner to obtain a bootstrap estimate of variance. The theoretical properties of this variance estimate are currently unknown.

4.2 "Coordinated" Bootstrap

The major difficulty with using the currently available bootstrap weights for estimating the variance of $\hat{\Delta}$ is due to the fact that independent bootstrap samples were taken to generate the bootstrap weights for the different files. As proposed by Mantel (1998), a "coordinated" bootstrapping approach would be better for estimating variances of quantities based on both cross-sectional files. In this approach, for the b th bootstrap estimate of Δ , the same selection of PSU's at both times are used for the strata that are in common to both cross-sectional files, and a separate selection is done for the bootstrap sample of PSU's from the strata that are due to the RDD supplemental sample. The different bootstrap estimates of Δ are combined in the usual manner to obtain the bootstrap variance estimate for $\hat{\Delta}$.

5. CONCLUSION

The variance estimation for the difference of two cross-sectional estimates obtained from a longitudinal sample with possible supplements requires the estimation of the covariance over the overlapping part. A major challenge is to identify members of the overlapping part and their "design" origin. This paper has shown how, by using Taylor linearization, some complexities can be simplified. An alternative straightforward method for using the available bootstrap weights is considered as well. Two other bootstrap variants are also discussed.

6. REFERENCES

- Mantel, Harold (1998). Variance Estimation for Estimates of Change. *Internal Memo, Statistics Canada*.
- Skinner, C.J. and Rao, J.N.K. (1996). Estimation in dual frame surveys with complex designs. *Journal of the American Statistical Association*, 91, 433, 349-356.
- Stukel, D.M., Mohl, C.A., and Tambay, J.-L. (1997). Weighting for Cycle Two of Statistics Canada's National Population Health Survey. *Proceedings of the Survey Methods Section, Statistical Society of Canada*, 111-116.