



## **National Population Health Survey (NPHS)**

### **Replacement of Bootstrap Weights by Coordinated Bootstrap Weights**

July 2003  
Statistics Canada /HSMD

By virtue of their composition, the cross-sectional samples of the first three cycles of the NPHS cannot be considered completely independent. In fact, the cross-sectional samples of each cycle include, among others, all the longitudinal panel members that responded to the survey in question; the rest of the sample coming from buy-in or top-up samples. Therefore, the fact that the cross-sectional samples include the same people, in part, implies some dependence between the samples over the cycles that must be taken into consideration when estimating variances.

#### ***Current Bootstrap Weights***

The cross-sectional bootstrap weights that were distributed with the data up to now were all created without taking this dependence into consideration. This does not have an impact when the analysis is based purely on one cycle of data, like the computation of a confidence interval for a proportion. However, when more than one cycle of data is used to compute a statistic, like a test for the difference between proportions in Cycles 1 and 2, the impact is considerable. See the list of questions and answers for more details on the impact.

#### ***Coordinating the Weights***

To remedy this situation, that is, to take the dependence between the samples into account when calculating the variances, coordinated bootstrap weights were calculated. Technically, coordinating the bootstrap weights means to preserve the structure of the bootstrap replicates for those persons common to the sample from one cycle to another. The method to compute the variances remains the same, but the use of coordinated bootstrap weights allows the dependence between the samples to be taken into consideration from now on.

Coordinated bootstrap weights were therefore created for the General and Health files for the first three cycles. They will now replace the bootstrap weight files that were originally distributed. To distinguish the coordinated bootstrap weights from the original bootstrap weights, the letter c was added to the beginning of the filename for each of the coordinated bootstrap weights. In order to avoid discrepancies, it is strongly recommended to use these new sets of bootstrap weights for variance estimation. The coordinated bootstrap weight files should be found in the same location as the original weights. For more information on the location of these files, contact the person in your organization responsible for NPHS data access.

## List of Questions and Answers

- ✚ *Which surveys and cycles are affected by this change?*  
Only the bootstrap weights for the NPHS cross-sectional files (General and Health) are affected.
- ✚ *Are previously computed estimates incorrect?*  
No. The estimates themselves are not at all affected. Only certain variance estimations (as well as any measures of precision such as standard deviations, coefficients of variation, statistical tests, etc.) are affected.
- ✚ *Are previously computed variance estimations incorrect?*  
Only those computed for statistics based on more than one cycle. A typical example would be a test of differences between proportions obtained in different NPHS cycles.
- ✚ *Quantitatively, what is the impact of using the coordinated bootstrap weights on the variance?*  
A quick study of the estimates produced with NPHS data showed that the use of the coordinated bootstrap weights reduced the variance by an average of 50% for statistics based on data from two consecutive cycles (either Cycle 1-Cycle 2 or Cycle 2-Cycle 3), or around 30% for statistics based on Cycle 1-Cycle 3 comparisons. These reductions in variance can even reach 80% in some cases. However, note that the reductions in variance were relatively weaker for provincial estimates in provinces with large buy-in samples in Cycle 2 (Alberta, Manitoba and Ontario).
- ✚ *Does this change affect my longitudinal analyses?*  
No. It only affects certain kinds of cross-sectional variance estimates.
- ✚ *How can I be sure that I am using the coordinated bootstrap weights and not the old weights?*  
The names of the bootstrap weight files were created by adding the letter c (for coordinate) to the beginning of the name used in the past. For example, for Cycle 1, the name of the coordinated bootstrap weight file for the General file is cb5h35.
- ✚ *Does this mean that all of the results obtained up to now with the Bootvar were wrong?*  
No. In fact, no version of the Bootvar has yet offered the possibility of computing variances for statistics based on more than one cycle of data.
- ✚ *How can I use the coordinated bootstrap weights with the Bootvar?*  
For variance estimation for statistics involving only one cycle of data, the Bootvar is used as before, but now with the coordinated bootstrap weight files. The Bootvar doesn't currently offer the possibility of variance estimation for statistics involving more than one cycle of data. However, an upcoming version of the Bootvar will allow it.
- ✚ *I've rerun my analyses based on only one cycle of data with the coordinated bootstrap weights and the results are not identical. Is this an error?*  
No. Since the bootstrap method involves a random process during the subsampling, the new coordinated bootstrap weights will not replicate the exact same variance estimates as the original bootstrap weights. However, the difference will be minimal, for example, in the order of decimals for the coefficient of variation (CV).

- ✦ *The bootstrap weight file for the Cycle 2 General file contains 500 weights while the original file contained only 100. Why is there a difference?*  
The Cycle 2 General file contains more than 200,000 respondents. Therefore, the production of 500 bootstrap weights would have been a large burden for the computers of that time. With the development of more powerful computers in recent years, it was decided that it was now reasonable to include 500 bootstrap weights on the file in order to 1-) to be in line with the standard of 500 established for health surveys and, 2-) to improve the precision of the variance estimates, particularly for small domains of interest.
  
- ✦ *Are there any references on the methodology of the coordinated bootstrap weights?*  
Roberts, Kovacevic, Mantel and Phillips (2001) deal with coordinated bootstrap weights in their article presented at the *Federal Committee on Statistical Methodology 2001 Research Conference*. The article is available on the web at the following address: [www.fcsm.gov/01papers/Mantel.pdf](http://www.fcsm.gov/01papers/Mantel.pdf).
  
- ✦ *Who can I contact to get more information about these coordinated bootstrap weights?*  
Contact the Health Statistics Division, NPHS Section by e-mail at [nphs-ensp@statcan.ca](mailto:nphs-ensp@statcan.ca), or by telephone at 1-613-951-1653.