Appendix 1 (as supplied by author): Calculating age-sex-standardized weighted prevalence using the bootstrap method

The National Population Health Survey (NPHS) 1996 and Canadian Community Health Survey (CCHS) Cycles 1.1, 2.1, 3.1, and 4.1 used multi-stage stratified cluster sampling to select participants. As a result, the sampling variance of sample statistics (such as the age-sex standardized prevalence of a risk factor) cannot be calculated using standard formulas for the variance of an estimate. Instead, the complex survey design must be accounted for when estimating the sampling variance of a statistic. Researchers using the NPHS and CCHS are unable to do this directly, because Statistics Canada, for the purposes of protecting the identity of survey participants, does not disclose the cluster to which each survey participant belongs. Therefore, to calculate variance estimates, one must use bootstrap methods. For each survey cycle, Statistics Canada performed the bootstrap method by drawing 500 sub-samples from the full sample, using simple random sampling with replacement of n-1 clusters among the n clusters selected within each stratum. An adjusted weight (i.e. bootstrap weight) was assigned to each observation belonging to the bootstrap sub-sample and these bootstrap weights were provided to the authors by Statistics Canada. Using these bootstrap weights, we calculated the weighted prevalence of each cardiovascular risk factor or disease for each age-sex stratum (i.e. male: 12-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, ≥85; and female: 12-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, ≥85). This was done 500 times using the 500 sets of bootstrap weights. Within each bootstrap sample, direct age-sex standardized rates were computed using the 2001 Ontario census population as the standard population. The variance of the sample estimate of the prevalence of a specific risk factor was estimated as the simple variance of that age-sex standardized prevalence across the 500 bootstrap samples. The P value was derived from the standard z-test in which the prevalence difference was divided by the standard deviation of this statistic across the 500 bootstrap samples. All tests were two-sided and P<0.05 was considered statistically significant.

Reference: BOOTVAR user guide (BOOTVAR 3.1 SAS version), Statistics Canada, April 2005. http://prod.library.utoronto.ca:8090/datalib/codebooks/cstdli/gss/gss18/sasbootdoc_eng.pdf. Accessed November 28, 2009.