Survey Weighting Adjustments and the Design Effect: A Case Study

Golshid Chatrchi[1] and François Brisebois
Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa ON K1A 0T6

Abstract
The final sampling weights are the result of various steps of adjustments to the design weights such as frame integration, nonresponse and calibration. Evaluating the Design Effect at each of these adjustments sheds light on their impact on the precision of survey estimates. In this paper, we focus on the Canadian Community Health Survey (CCHS), which uses a complex survey design with multiple frames and multiple stages of selection, and empirically examine the effect of different steps of the weighting process on the overall variability in estimates. The results suggest that the use of unequal person-selection probabilities and the nonresponse adjustment have the most negative impact on the design effect of the CCHS while as expected calibration decreases the design effect and improves the precision of the estimates.

Key Words: bootstrap, calibration, complex survey design, design effect, multiple frames, weighting

1. Introduction

The weighting process is an essential component in the survey process. In general, sampling weights are needed to compensate for the limitations of the sample such as the selection of units with unequal probabilities, the imperfect frame and nonresponse. The first step of the weighting adjustment is usually the creation of the base weight (or design weight) for each sampled unit. The base weights are basically the inverses of the inclusion probabilities. These weights are then adjusted to correct for the limitations in the sample. For complex surveys, the weighting process involves several steps of adjustments. The question of interest is how each of these adjustments affect the precision of survey estimates.

In this paper we focus on the Canadian Community Health Survey (CCHS) and estimate its design effect at different steps of the weighting process. In Section 2, we discuss the sampling design and weighting process of the CCHS. In Section 3, we briefly review the definition and use of design effect. Subsequently, in section 4, we compute the design effects of the CCHS for three variables at different steps of weighting adjustments and discuss the variability of estimates after each adjustment. In section 5 and section 6 we

[1] Author to whom correspondence may be addressed: golshid.chatrchi@canada.ca

highlight the role of the calibration step and the bootstrap method. Finally, we summarize our discussion in Section 7.

2. The Canadian Community Health Survey

The Canadian Community Health Survey (CCHS) is a cross-sectional survey that collects information related to health status, health care utilization and health determinants for the Canadian population. The survey began collecting data in 2001 and was repeated every two years until 2005. Starting in 2007, the CCHS data were collected annually instead of every two years. In 2012, CCHS began work on a major redesign project that involved a series of studies to review the sampling methodology, adopt a new sample frame, modernize the content and review the target population. The redesign was completed and implemented for the 2015 cycle.

One of the objectives of this study was to identify areas of improvement in the weighting process for the purpose of the redesign project. In this paper, we concentrate on the sampling design of the CCHS before the 2015 redesign.

The target population of the CCHS is all persons aged 12 years and over living in private dwellings across Canada. The CCHS relies upon a large sample of respondents and is designed to provide reliable estimates at the health region (HR) level. Health regions refer to health administrative areas with population sizes varying from about 10,000 to 2,250,000 persons aged 12 and over. To provide reliable HR-level estimates, a sample of 65,000 respondents is selected annually. A multi-stage sample allocation strategy gives relatively equal importance to the HRs and the provinces. In the first step, the sample is allocated among the provinces according to the size of their respective populations and the number of HRs they contain. Each province's sample is then allocated among its HRs proportionally to the square root of the population in each HR.

The CCHS used three sampling frames to select a sample of households (Béland et. al., 2005); 49.5% of the sample of households came from an area frame, 49.5% came from a list frame of telephone numbers and the remaining 1% came from a Random Digit Dialing (RDD) sampling frame. The area frame used by the CCHS is the one designed for the Canadian Labour Force Survey (LFS). The LFS design uses a multi-stage stratified cluster design in which the dwelling is the final sampling unit. In the first stage, homogeneous strata are formed and independent samples of clusters are drawn from each stratum. In the second stage, dwelling lists are prepared for each cluster and dwellings are selected from these lists. The CCHS basically follows the LFS sampling design with health regions generally forming the strata. This frame is used in most regions across Canada. As a complement to the area frame, a telephone list frame is used. The telephone frame is the InfoDirect list, an external administrative database of names, addresses and telephone numbers from telephone directories in Canada. Within each stratum (generally the HR), the required number of telephone numbers is selected using a Simple Random Sampling (SRS) design. In four HRs, a Random Digit Dialling (RDD) sampling frame of telephone numbers was used to select a sample of households. The units belonging to the telephone list frame and the RDD are interviewed using computer assisted telephone interviewing (CATI), and area frame units are interviewed using computer assisted personal interviewing (CAPI). The CCHS interviews (CAPI or CATI) are done in two parts. First, a knowledgeable household member is asked to provide basic demographic information on all residents of the dwelling (roster of the household). Then, one member of the household is selected using unequal probabilities of selection to answer a more detailed questionnaire covering health-related topics.

2.1 The weighting process of the CCHS

Given this complex survey design, several steps of weighting adjustments are required. Wilder and Thomas, S. (2010) summarized the weighting process of the CCHS in Figure 1.

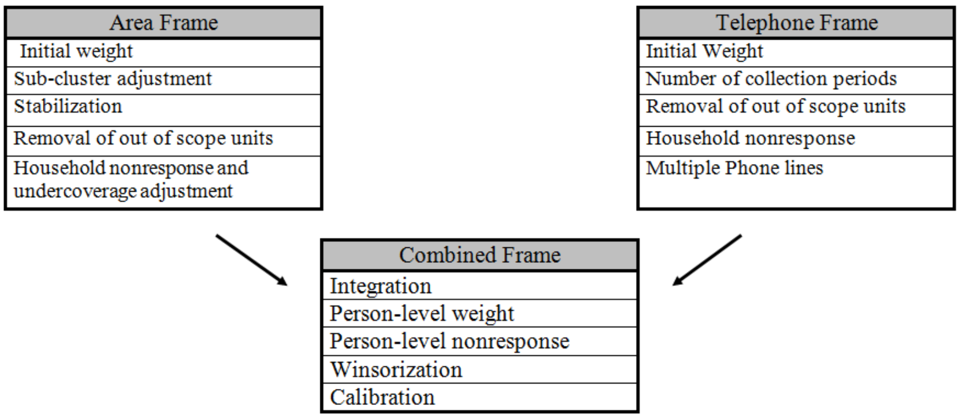


Figure 1: The CCHS weighting process

As mentioned earlier, the CCHS was redesigned in 2015 and the above weighting process is no longer used. However, for the purpose of this study we follow the steps of adjustments presented in Figure 1.

The goal is to assess the impact of the following steps of the weighting process on the variability of estimates.

1. Removal of out-of-scope units in the area frame and the list frame
2. Household nonresponse for units in the area frame and the list frame
3. Integration
4. Person-level weight
5. Person-level nonresponse
6. Winsorization
7. Calibration

It should be mentioned that the RDD sample, Yukon, Northwest Territory and Nunavut were not included in this study.

3. Design Effect

The design effect is widely used to compare the efficiency of survey designs and to develop sampling designs. Kish (1965) introduced the term “design effect” in survey sampling, and defined it as the ratio of the sampling variance of an estimator under a given design to the sampling variance of an estimator under simple random sampling

(SRS) of the design. The concept of design effect received much attention by survey samplers, and it became a useful tool for assessing the efficiency of the sampling design. If the design effect is less than one, this indicates that the sample design is more efficient than an SRS design; a greater than one value indicates that the sample design is less efficient than an SRS design. The design effect measure can also be used to determine the sample size required to do an analysis. The larger the design effect, the more sample is required to obtain the same precision of an estimate as would have been obtained under an SRS design.

In practice, an estimate of the design effect is calculated using the corresponding variance estimators for the sample data set (Lehtonen and Pahkinen, 2004, p.15).

$$DEFF_{p(s)}(\hat{\theta}) = \frac{\hat{v}_{p(s)}(\hat{\theta})}{\hat{v}_{srs}(\hat{\theta})}$$

Where subscript p (s) refers to the actual sampling design and $\hat{\theta}$ denotes the estimator of a population parameter θ. For complex designs, the estimation of the sample variance, $\hat{v}_{p(s)}$, is complicated. It is usually computed using resampling techniques, such as bootstrap and Jackknife.

Assuming that finite population correction (fpc) factor can be ignored, Kish (1987) proposed the design effect decomposition as a function of two independent components associated with clustering and unequal weighting.

$$DEFF_{p(s)}(\hat{\theta}) = DEFF_{unequal\ weighting} \times DEFF_{clustering}$$

$$= (1 + cv_w^2) \times \left[1 + \left(\bar{b} - 1\right)\rho\right]$$

where cv_w^2 denotes the relative variance of the sample weights, $\bar{b}$ is the mean of cluster size and ρ is the intra-class correlation coefficient.

Kish (1992) presented another formula for expressing the first component of the above formula, the design effect due to unequal weighting, for a sample mean:

$$DEFF_{unequal\ weighting}\ (\bar{y}) = \frac{n \sum_j w_j^2}{(\sum_j w_j)^2} = 1 + cv_{wj}^2$$

where w_j is the final sample weights of the j^{th} unit in the sample and j=1,2…,n.

Kish's decomposition method received much attention by survey samplers. Gabler et al. (1999) provided a model-based justification for Kish's formula. Kalton et al., (2005), and Lehtonen and Pahkinen (2004) provided the analytical evaluation of design effects and examined separately the design effects resulting from proportionate and disproportionate stratification, clustering and unequal weighting adjustments. Park et al. (2003) extended Kish's decomposition and included the effect of stratification.

While decomposition of the design effect can provide an approximate measure of the impact of unequal weighting due to nonresponse adjustment, it is not a good approximation when weights are post-stratified or calibrated to known total (Lê et al., 2002). Moreover, the above equations are true under certain assumptions. In particular, the later equation depends on the assumption of equal strata means and constant within-stratum variances (Lê et al., 2002). Some of these assumptions do not hold in practice. Considering the limitation of decomposition method in practice, we empirically evaluate the design effect of the CCHS at different steps of the weighting process.

4. Evaluating the Design Effect at Different Steps of the Weighting Process

To derive design effects, variances under the CCHS design are calculated using the bootstrap method. An extended version of the proposed bootstrap method by Rao, Wu and Yue (1992) has been used by the CCHS since the survey's first occurrence in 2001. Yeo, Mantel and Liu (1999) described the adaption of the method to the particularity of Statistics Canada's National Population Health Survey, which is the predecessor to the CCHS and used a similar methodology. By using the bootstrap method, all of the design aspects that we are interested in, up to and including the weighting step of interest, are included in the design variance estimate. The estimates of design effects presented in this paper are produced using BOOTVAR, which is a set of SAS macros (also available for SPSS) developed at Statistics Canada to estimate variances using the bootstrap method.

The design effects of certain variables are calculated at the health region level, after each of the seven weight adjustments mentioned in section 2. A short description of these adjustments is given below.

The out-of-scope adjustment is used as the baseline in the evaluation process. During collection, some sampled dwellings are found to be out-of-scope. For example dwellings that are demolished, under construction or institutions are identified as being out-of-scope in the area frame while phone numbers that are out of order or for business are out-of-scope on the telephone frame. To adjust for out-of-scope, these units and their associated weights are removed from sample.

The second stage of the weighting process that is considered for this study is household nonresponse. Household nonresponse occurs when a household refuses to participate in the survey or cannot be reached for an interview. These units are removed from the sample and their weights are redistributed to responding households within response homogeneity groups. These groups are created using Eltinge and Yansaneh (1997) method which uses the results of logistic regression models to divide the sample into groups with similar response properties.

The third stage of interest for the study is the impact of the integration adjustment. The current integration approach takes into account the portion of the population covered by both frames as well as the under-coverage of the telephone frame. The households without a landline or without a listed telephone number are not covered by the telephone list frame. Not taking this into account could cause a bias if the households not covered by the telephone frame have different characteristics than the ones covered. To take the under-coverage into consideration, the CCHS integrates only the sampled households that are common to both frames. The weights of the households that only belong to the area frame (households without landline or without listed phone numbers) remain unchanged. This allows these households to represent other similar households in the population. For the common portion, the integration process applies a contribution to each frame. An adjustment factor α between 0 and 1 is applied to the weights; the weights of the area frame units that have telephone are multiplied by α and the weights of the telephone frame units are multiplied by 1- α. The term α was originally based on the overall sample size contribution of each frame to the common portion. Starting in 2008, a fixed α of 0.4 has been used for those units on both frames to ensure greater comparability of estimates across years. It should be mentioned that an adjustment in the area frame is also done before the integration to account for the under-coverage. The area frame has about 12 % households' under-coverage using the current LFS design. In order to deal with this frame defect, a post-stratification adjustment is applied at the HR level using the most up-to-date household counts.

The fourth stage to be studied is the derivation of the person-level weight. At this step, the household-level weights are adjusted using the inverse of the person-level selection probabilities to calculate person-level weights. The person-level selection in the CCHS is done with unequal probabilities of selection based on the household size and the age of the household members. At this stage, the concept of the variables examined in the study is changed since the statistical unit is the person selected in a household. The same concepts were preserved but estimates will now reflect characteristics in terms of people instead of households.

The fifth step of the weighting adjustments to examine is the person-level nonresponse adjustment. The CCHS interview has two parts: first the interviewer completes the list of the household's members and then one person is selected for the interview. It is possible that the household roster is obtained (household response) but the selected person refuses to be interviewed or cannot be reached for some reasons. This causes person-level nonresponse. The same treatment used in household nonresponse is used at this stage with response homogeneity groups being created based on person-level characteristics.
The sixth stage to be examined is Winsorization. The weighting process may cause some units to have extreme weights that can have a large impact on the variance. The weights of these units are adjusted downward using a "Winsorization" trimming approach.

The last step to study is calibration. This adjustment is done to make sure that the sum of the final weights corresponds to the population estimates defined at the HR level for each of the 10 age-sex groups. The five age groups are 12-19, 20-29, 30-44, 45-64 and 65+, for both males and females. At the same time, weights are adjusted to ensure that each collection period is equally represented in the population. The truncated linear method is used and the calibration is done using CALMAR (Sautory, 2003) with the most up to date population counts and the most up to date geographic boundaries.

In order to evaluate the design effect at these steps, we need to have variables that are available at the household level. We choose variables with different prevalence rates to get more exhaustive results. The variables of interest are:

- Households with at least one "child" (national prevalence at 20%)
- Households with one or two "kids" (national prevalence at 10%)
- Households with five members (national prevalence at 5%)

It should be mentioned that "child" is defined as a person who is less than 12 years old, and "kid" refers to a person living in a household whose age is less than 6 years.

These variables are only available for the respondent households, as the CCHS data files only contain information of the responding households. Therefore, data for the nonrespondent households are imputed based on the prevalence rate of the observed units at the CCHS stratum level.

The design effects are calculated at health region level (112 health regions). In other words, for each variable 112 design effects are estimated at each step of the adjustment. Chatrchi et al. (2015) studied the changes of the design effect estimates after each steps of the weight adjustment and provided more explanations on the reason and nature of each adjustment. Figure 2 is extracted from that paper. It presents the medians of design effects estimated at health region level.

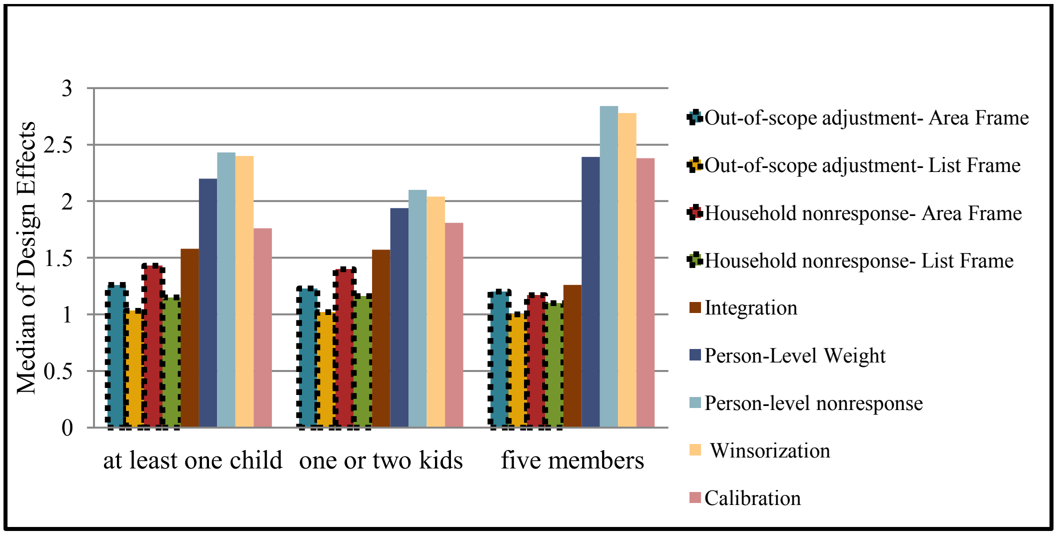


Figure 2: Median (P50) of design effects at the health region level

The first four bars (dashed outlined) represent the design effects of step 1 and step 2 for the area frame and the list frame separately (before integration) while the other bars show the design effects after integration (combined frame). The effect of each step can be determined by comparing the design effect for the step of interest to the design effect for the previous step

The median of design effects after the out-of-scope adjustment in the area frame is around 1.2. This measure is around 1 for the telephone frame. These results are expected since the telephone frame sampling design is a simple random sampling process within each stratum while the sampling design of the area frame is a multi-stage stratified cluster design.

After the household nonresponse adjustment, design effects go up slightly for both frames. This rise is expected since the nonresponse adjustment increases the variability of weights and therefore the variance of estimates in favor of reducing potential nonresponse bias. Also, the differences between the increases seen with the nonresponse adjustment on the two frames might be explained partly by the fact that there is more nonresponse on the telephone frame.

There is a significant jump in design effects after joining the two frames. The median design effect within the health regions was between 1.2 and 1.4 for the area frame and between 1.1 and 1.2 for the telephone frame before integration. This amount increased to around 1.3 to 1.6 after integration for the combined sample. The increase can be explained by the fact that the CCHS integration process is not optimal. As mentioned earlier, the integration factor (α) has been set to a fixed value to facilitate the comparability of results from one year to another. This adds a lot of variability to the weights and consequently increases the design effects.

The median of the design effect within health regions varies between 1.9 and 2.4 after the person-level adjustment. This rise in the design effect is mainly due to the unequal probability of selection used by the CCHS. Depending on the household composition, the person-level selection adjustment can be as low as 1 and as high as 20 in some cases.

The median design effect at the health region level fluctuates between 2.1 and 2.8 while the median of the design effect in the previous step was between 1.9 and 2.4. This is similar in nature to the household nonresponse adjustment. The nonresponse adjustment increases the variability of the weights.

After the winsorization step, the median design effect varies between 2 and 2.8. Comparing this with the results of the previous step, it can be concluded that winsorization does not have a significant effect on the median of the design effect mainly because very few units are winsorized.

The calibration step significantly decreases the design effect. The median of the design effect after calibration is about 1.8 to 2.4, compared to 2 to 2.8 at the previous step.

Overall, the bar chart suggests that the person-level-weight and the nonresponse adjustments have the most negative impact on the design effect while calibration decreases the design effect. The integration step has also a negative impact that is non-negligible.

The trend of the design effects throughout the weighting process is similar for all variables. However, we can notice some differences for the variable "household with five members" compared to the others. This is most likely due to how the design interacts with the variable of interest. Households with many members have different landline rates than those with fewer members. Also, the person-level adjustment (unequal probabilities of selection) will have a larger impact on these types of households. This is also the variable with the lowest prevalence rate and large weights can have a greater impact on the variance and the design effects in this situation.

5. Person-level Weights and Household-level Weights

The CCHS also provides household weights for household analyses. The weighting process of the household weights is similar to the process described in Section 2. However, the person-level adjustment is excluded, as the final product represents households rather than individuals.

As shown in Figure 2, there is a significant rise after the person-level adjustment. Hence, it is expected that the estimated design effects be lower when household weights are used.

Figure 3 shows the design effect of the households with at least one child for the last three steps of the weighting process using person weights and household weights (excluding the person-level adjustment). The right side of the graph illustrates design effects at health region level using person weights and the left sides shows design effect computed by household weights. The first box plots in both sides (design effect after applying nonresponse adjustment) confirm that at this stage household weights provide more efficient estimates. Using household weights, the median of design effects across 112 health regions is 1.76. However, using person weights this measure is 2.43.

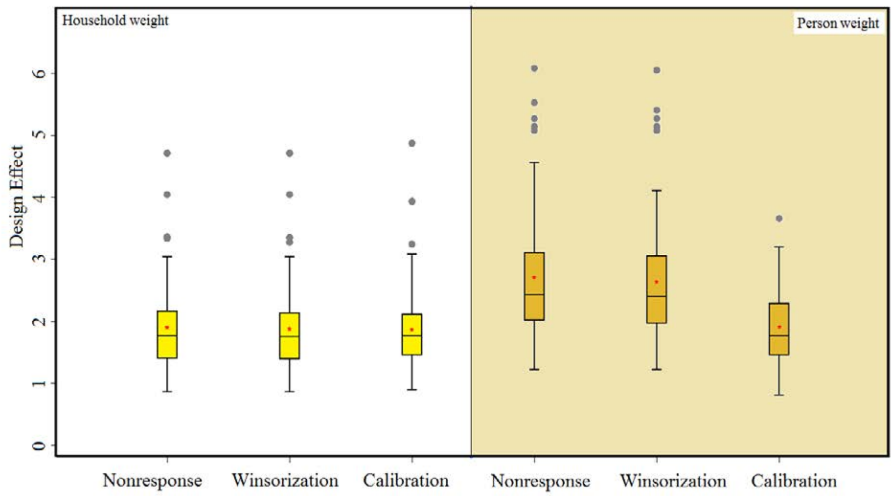


Figure 3: Design effect at health region level after nonresponse, Winsorization and Calibration adjustments using person and household weights

The second box plots suggest that Winsorization does not have a significant effect on the median of the design effects mainly because very few units are winsorized.

The third box plot in each side of the plot demonstrates the design effect distribution at the health region level, after the calibration adjustment. Calibration significantly decreases the design effect produced by person weighs, while it has almost no impact on the household design effects. More precisely, calibration works in favor of the estimates using person weights, but has no influence on the estimates produced by household weights. This could be due to the former post-stratification adjustment which was done to compensate for the under-coverage in the area frame. As mentioned earlier in Section 2, a post-stratification adjustment is applied at the HR level using the most up-to-date household counts. Therefore, calibrating the weights to household counts is not a big adjustment as weights were (partially) adjusted to match the household counts in an earlier step. In addition, this could be due to the choice of the calibration variables used in two procedures. Person weights are calibrated to health regions and ten age-sex groups. On the other hands, household weights are calibrated to the total household counts at the province by household size. In particular, the calibration is done to ensure that the sum of the final household weights corresponds to the average of the monthly household counts at the province by household size.

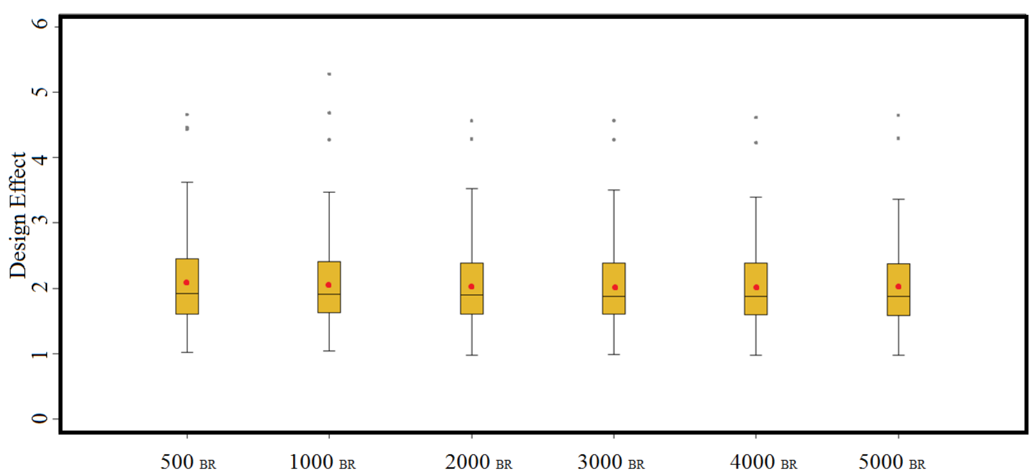
The potential gain in precision of an estimate due to calibration depends on the correlation between calibration variables and the variables of interests. The correlation between calibration variables used at person weights (HR, age and sex) and the variables of interest (health related variables) provides more efficient estimates. However at the household level, the calibration does not improve the efficiency of estimates as the relationship between calibration variables at household level and variables of interest is not very strong. This highlights the importance of calibration variables, and the fact that calibration should be done using auxiliary variables that are correlated with the variables of the interest.

6. Role of Bootstrap Method

The bootstrap method is an important part of this study, as design effects of different steps of weighting process are all derived from the bootstrap estimate of variances under the CCHS design. As mentioned earlier the Rao, Wu and Yue bootstrap method was used to estimate the design effects, and the number of bootstrap replicates was set to 500.

A key question would be how the Rao, Wu and Yue bootstrap method and number of replicates affect the results of this study. Mach et al. (2007) and Girard (2009) studied the Rao, Wu and Yue bootstrap method under different scenarios including different sampling method and nonresponse adjustment. In particular, Girard (2009) showed that even though the Rao-Wu bootstrap method considers the nonresponse mechanism as a deterministic process and does not fully capture the variance due to nonresponse, it provides reasonable estimates when the sampling fraction is small. Conceivably, the effect of the post-stratification and calibration on the estimate of variances would be ignorable if the number of the post-strata is not too large. Regarding the number of bootstrap replicates (BR), most of the literature suggests choosing sufficiently large number (Pattengale et al., 2010). We performed a simulation study and examined how the estimates of design effect vary using different number of replicates.

Figure 4 presents the design effect after calibration adjustment across 112 health regions with different number of bootstrap replicates for one of the variables of interest, household with one or two kids.



Median	1.92	1.91	1.90	1.88	1.88	1.88

Figure 4: Design effect at health region level with different number of bootstrap replicates

Even though the overall display of design effect across 112 health regions does not change much and the median of the design effects fluctuates slightly, evaluating the design effect estimates for each health region shows that increasing the number of bootstrap replicates affects the estimates of design effect noticeably. In particular, the estimates of design effect for each health region vary 6% in average as the number of bootstrap replicates increases from 500 to 5000.

To address this issue, we estimated the design effect at different steps of weighting process for each health region, using different number of bootstrap replicates ranging from 500 to 5000. Comparing the estimates of design effect for each health region suggests that the estimates of design effect are not stable as the number of bootstrap replicates changes. For instance, Figure 5 shows the changes of design effect estimates for one of the health regions, HR 3565 (Waterloo Health Unit), after integration, person-level weighing, person-level nonresponse, winsorization and calibration adjustments

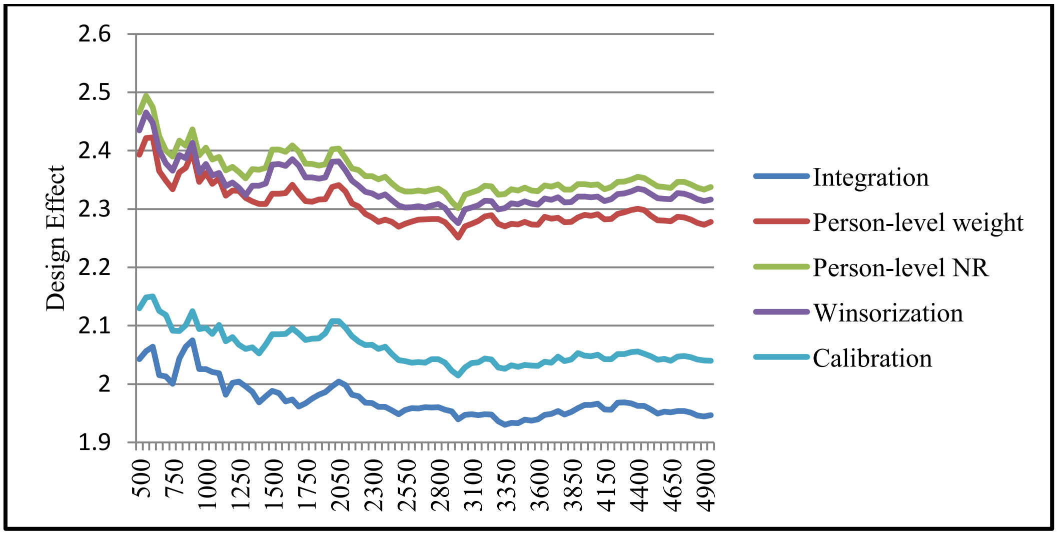


Figure 5: Design effect of health region 3565 with different number of bootstrap replicates

The fact that estimates of design effects are unsteady for different number of bootstrap replications does not contradict the results presented in previous sections, as the impact of the different steps of weighting adjustment on the design effect is still detectable and appears constant in Figure 5. In other words, the instability of design effect does not invalidate the results of this study and we can still conclude that nonresponse and person-level-weight adjustments have the most negative impact on the design effect and calibration decreases the design effect.

7. Concluding Remarks

In this paper, we empirically examined the impact of different steps of the weighting process on the efficiency of the Canadian Community Health Survey design, and showed which of the survey steps cause more variability in the estimates and cause the estimates to be less efficient compared to a simple random sample. The comparison results suggest that using unequal person-selection probabilities and the nonresponse adjustment considerably increase the variability of design. Similarly, using multiple frames under the current integration method has a negative effect on the design effect. On the other hand, winsorization and calibration using population counts decrease the design effect and improve the precision of the estimates. Although these results were expected, the analysis helped quantify the changes. The role of the calibration step and the choice of the auxiliary information used in the calibration procedure were highlighted. In addition, the effect of the bootstrap method (Rao, Wu and Yue) and number of replicates on the results was discussed.

Considering the above points, using only one frame which consists of a list of individuals instead of a list of dwellings would lower the design effect and reduce the variation. Moreover, the calibration variables should be chosen with respect to their correlation with the variables of the interest. The calibration adjustment would improve the precision of the estimates when the relationship between y variables and the auxiliary information is strong.

While results discussed here are valid in the framework of the CCHS they may be helpful for other multi-stage household surveys with similar complexities.

Acknowledgements

The authors would like to thank Steven Thomas and Marie-Claude Duval for their comments and suggestions that helped to improve the content of this study and ultimately of this document.

References

Cox, B. G., Binder, D. A., Chinnappa, B. N., Christianson, A., Colledge, M. J., & Kott, P. S. (Eds.).(2011). *Business survey methods* (Vol. 214). John Wiley & Sons.

Chatrchi, G., Duval, M.-C., Brisebois, F., and Thomas, S. (2015). The Impact of Typical Survey Weighting Adjustments on the Design Effect: A Case Study. Survey Methods: Insights from the Field, Retrieved from http://surveyinsights.org/?p=491

Eltinge J. L and Yansaneh, I. S. (1997), Diagnostics for formation of non-response adjustment cells with an application to income non-response in the U.S Consumer Expenditure Survey. *Survey Methodology*, 23, pp. 33-40.

Gabler, S., Haeder, S., & Lahiri, P. (1999). A model Based Justification of Kish's Formula for Design Effects for Weighting and Clustering. *Survey Methodology*, 25, pp. 105-106.

Gambino, J. (2009). Design Effect Caveats. *The American Statistician*, 63, 141-146.

Girard, C. (2009). The Rao-Wu rescaling bootstrap: from theory to practice. In *Proceedings of the Federal Committee on Statistical Methodology Research Conference,* (pp. 2-4).Washington, DC: Federal Committee on Statistical Methodology.

Kalton, G., Brick J. M., & Lê, T. (2005). Chapter IV: Estimating Components of design effects for use in sample design. *Household Surveys in Developing and Transition Countries*. United Nations Publications.

Kish, L. (1965). *Survey Sampling.* New York: John Wiley & Sons.

Kish, L. (1987). Weighting in Deft2. *The survey statistician*, pp. 26-30.

Kish, L. (1992). Weighting for unequal Pi. *Journal of Official Statistics,* 8, pp. 183-200.

Lê, T., Brick J. M., & Kalton G. (2002). Decomposing Design Effects. In *Proceedings of the American Statistical Association Joint Statistical Meetings, Section on Survey Research Methods*. New York: American Statistical Association.

Mach, L., Saïdi, A. & Pettapiece, R. (2007). Study of the Properties of the Rao-Wu bootstrapvariance estimator: What happens when assumptions do not hold?. In *Proceedings of the Survey Methods Section, SSC Annual Meeting.* St. John's: Statistical Society of Canada.

Lehtonen, R., & Pahkinen, E. (2004). *Practical Methods for Design and Analysis of Complex Surveys*. Chichester: John Wiley & Sons.

Park, I., Winglee, M., Clark, J., Rust, K., Sedlak, A., & Morganstein D. (2003). Design Effect and Survey Planning. In *Proceedings of the American Statistical Association Joint Statistical Meetings, Section on Survey Research Methods.* San Francisco: American Statistical Association.

Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R., Moret, B. M., & Stamatakis, A. (2010). How many bootstrap replicates are necessary?. *Journal of Computational Biology*, *17*(3), 337-354.

Rao, J.N.K., Wu. C.J.F., & Yue, K. (1992). Some Recent Work on Resampling Methods for Complex Surveys. Survey Methodology, 18, 209-217.

Sarafin, C., Simard, M. & Thomas, S. (2007). A Review of the Weighting Strategy for the Canadian Community Health Survey. In *Proceedings of the Survey Methods Section, SSC Annual Meeting.* St. John's: Statistical Society of Canada.

Särndal, C. E., Swensson, B., & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.

Sautory, O. (2003) CALMAR 2: A New Version of the CALMAR Calibration Adjustment Program. In *Proceedings of Statistics Canada International Symposium, Challenges in survey taking for the next decade.* Ottawa: Statistics Canada.

Statistics Canada (2011). Canadian Community Health Survey (CCHS) Annual component; User guide, 2010 and 2009-2010 Microdata files. *Statistics Canada.*

Statistics Canada (2005), *User Guide (BOOTVAR 3.1 SAS Version*), Statistics Canada,

Available from: http://www.statcan.gc.ca/rdc-cdr/bootvar_sas-eng.htm

Wilder, K., & Thomas, S. (2010). Dual frame Weighting and Estimation Challenges for the Canadian Community Health Survey. In *Proceedings of Statistics Canada International Symposium, Social Statistics: The Interplay among Censuses, Surveys and Administrative Data.* Ottawa: Statistics Canada.

Yeo, D., Mantel, H., & Liu, T. P. (1999). Bootstrap variance estimation for the national population health survey. In *Proceedings of the American Statistical Association Joint Statistical Meetings*, *Section on Survey Research Methods.* Baltimore: American Statistical Association.