

Catalogue no. 11-522-XIE

**Statistics Canada International
Symposium Series - Proceedings**

**Symposium 2006 :
Methodological Issues in
Measuring Population Health**

2006



**Statistics
Canada**

**Statistique
Canada**

Canada

Methodological Challenges in Analyzing Nutrition Data from the Canadian Community Health Survey - Nutrition

François Verret¹

Abstract

Statistics Canada conducted the Canadian Community Health Survey – Nutrition in 2004. The survey's main objective was to estimate the distributions of Canadians' usual dietary intake at the provincial level for 15 age-sex groups. Such distributions are generally estimated with the SIDE application, but with the choices that were made concerning sample design and method of estimating sampling variability, obtaining those estimates is not a simple matter. This article describes the methodological challenges in estimating usual intake distributions from the survey data using SIDE.

KEYWORDS: Nutrition survey; usual intake; daily intake; measurement error models; bootstrapping.

1. Introduction

The Canadian Community Health Survey (CCHS) is a series of cross-sectional surveys consisting of two cycles: the “1” cycle surveys are designed to produce estimates of the general health of the population at the regional level, while the “2” cycles are intended to provide provincial-level estimates of the population's health in a specific part of the health field. During the development of the CCHS content, it was felt that nutrition was a very important subject to cover because of Canadians' concerns about excess weight and obesity. In addition, it had been several years since the last national survey of nutrition in 1972. It was therefore decided that cycle 2.2 of the CCHS would be about nutrition.

Specifically, the main objective of the CCHS – Nutrition was to produce estimates of the distributions of the daily nutritional intake of provincial populations in terms of nutrients, food and food groups at the provincial level for 15 age-sex groups throughout 2004. The age-sex groups covered were as follows: infants under the age of 1 (estimates required at the national level only); children aged 1 to 3 and children aged 4 to 8, both sexes combined; and the 9-13, 14-18, 19-30, 31-50, 51-70 and 71+ age groups broken down by gender. Persons living on Indian reserves and Crown land, institutional residents, full-time members of the Canadian Forces, and residents of selected remote areas were excluded from the target population (exclusions represent about 2% of the population of Canada's provinces).

A person's usual dietary intake can be expressed as the expectation of the theoretical distribution that generates his or her daily dietary intakes. A person's usual intake can also be regarded as the average of his or her daily intakes over a long period of time. Hence it is impossible to measure a person's usual intake directly. It is possible, however, to measure a person's daily intake by having him/her recall what he/she ate and drank in a 24-hour period. For the CCHS cycle 2.2, each recall covered the full 24 hours of the day preceding the interview (from midnight to midnight). Estimating each respondent's usual intake was not an option because of the number of recalls per respondent that would have been necessary, the response burden that would have been generated, and the collection cost per recall. Yet the survey's objective was to estimate the usual intake distribution in the population. That can be done with one recall per respondent, a second recall for a subset of those respondents and a sophisticated methodology based on the theory of measurement error models.

¹François Verret, Statistics Canada, 16th Floor, R.H. Coats Building, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A 0T6 (Francois.Verret@statcan.ca)

This article describes the issues encountered in applying the theory in the CCHS context. The next section contains a brief overview of some aspects of the sampling plan and the estimation process. Section 3 describes the methodology of the Software for Intake Distribution Estimation (SIDE), which is used to analyze nutritional data. Section 4 presents the issues, problems and solutions involved in estimating the sampling variance of the estimates produced by the SIDE.

2. Canadian Community Health Survey - Nutrition

It was determined that a sample of 35,000 respondents (and therefore 35,000 first recalls) was needed to ensure that the estimators had the desired precision. That is a very large sample for a nutrition survey. We had to use several frames to generate the sample, which was spread evenly across the four quarters of 2004 to eliminate seasonal bias in the data. A subsample of 10,000 second recalls was also collected to control for day-to-day variation in respondents' dietary habits. For more details on the sampling plan see Béland *et al.* (2003) and Junkins and Vigneault (2003). The plan provides an address for the interviewer to visit in order to contact the resident household. One person from the household was selected by computer to take a computer-assisted interview. The first recall was done in person at the respondent's home, while the second recall was done by telephone. The data from the first recall are considered more reliable than the data from the second recall. Respondents might have a tendency to bias their responses in the second recall either because they had tried to eat better following the first interview or to speed up the second interview.

The sample weighting process is complex because of the sample design that had to be used to achieve the survey's objectives. For more information on the CCHS's standard weighting procedures, see Brisebois and Thivierge (2001). The preferred method of estimating sampling variability with the CCHS data is bootstrapping. For cycle 2.2, it was determined that bootstrapping was the most appropriate technique because of the sampling plan's complexity, the time required to produce the estimates and the desirable properties of the variance estimators it produces (Rao and Wu, 1988). For each of the 500 bootstrap replicates, we sampled n_h-1 clusters with replacement from the n_h clusters in stratum h . The kind of bootstrapping we used assumes that the first-stage sampling fractions are negligible.

3. Analyzing Nutrition Data with the Software for Intake Distribution Estimation (SIDE)

3.1 Software for Intake Distribution Estimation

The Software for Intake Distribution Estimation (SIDE) is used by most analysts when the survey data are in the form of 24-hour recalls. It is highly popular because it is very thorough: it applies the measurement error model required to estimate the usual intake distribution in a population, and it makes sophisticated adjustments to the data before and after. The SIDE methodology is described in detail in Nusser *et al.* (1996) and summarized in the subsection 3.2.

3.2 The steps executed by SIDE

Step 1 involves adjusting the daily intake data to improve their quality and simplify the calculations made in subsequent steps. First, SIDE forces the mean and variance of the second recalls to be equal to the mean and variance of the first recalls. This adjustment is based on the assumption that the first recall data are of better quality than the second recall data. A ratio adjustment can also be made to eliminate the effects of detrimental variables (continuous or discrete). Step 1 ends with a smoothing process that produces an equal-weights sample, which simplifies the calculations in subsequent steps. We compute the empirical distribution function of the daily intakes

using the formula $\hat{F}_Y(a) = \sum_{i=1}^n w_i \sum_{j=1}^{k_i} I_{Y_{ij}}(a)$, where Y_{ij} is the measurement of the daily intake (adjusted as described above) of recall j of respondent i and $I_{Y_{ij}}(a)$ is the indicator function, which is equal to 1 if Y_{ij} is less than or equal to a . The function is made continuous by connecting the mid-points of the steps of $\hat{F}_Y(a)$ with straight lines. The

resulting function is \tilde{F}_Y . A sample with equal weights Z_{ij} is created using the formula $Z_{ij} = \tilde{F}_Y^{-1} \left[\left(\sum_{i=1}^n k_i \right)^{-1} (s_{ij} - 0.5) \right]$ for $i=1, 2, \dots, n$ and $j=1, 2, \dots, k_i$, where s_{ij} is the rank of Y_{ij} .

The measurement error model used by SIDE requires normality in the data. For this reason, Step 2 is a complex normality transformation. First, a power transformation is applied to the data. Then the data are put in order and matched with the percentiles in the standardized normal distribution. Lastly, the data are divided into equal intervals and regressions are used to find the best cubic transformations on the intervals. The result is X_{ij} data that are zero-centred, unlike the Z_{ij} and the Y_{ij} , which are centred on the means of the untransformed first-recall daily intakes.

In step 3, SIDE fits the following measurement error model:

$$\begin{aligned} X_{ij} &= x_i + u_{ij}, \\ u_{ij} &= \sigma_i e_{ij}, \end{aligned} \quad i = 1, \dots, n, \quad j = 1, \dots, k_i, \quad (1)$$

where X_{ij} is the j^{th} measurement of the transformed daily intake of respondent i , x_i is the usual intake of respondent i , and u_{ij} is the measurement error. We also assume the following distributions: $x_i \sim NI(\mu_x, \sigma_x^2)$, $e_{ij} \sim NI(0,1)$ and $\sigma_i^2 \sim (\mu_A, \sigma_A^2)$. As mentioned earlier, a precise estimate of x_i cannot be produced with the survey data. Instead, the parameters of interest are μ_x and σ_x^2 because they completely define the usual intake distribution in the population on the normal scale. Let n be the number of respondents (or the number of first recalls), N the total number of recalls, k_i the number of recalls of respondent i (equal to 1 or 2), $\bar{X}_{i\cdot}$ the mean of the recalls of respondent i and $n_0 = N - N^{-1}(4N - 3n)$. The centrality parameter μ_x is estimated by $n^{-1} \sum \bar{X}_{i\cdot}$, and σ_x^2 is estimated by the method of moments (see Table 1) as follows:

$$\hat{\sigma}_x^2 = \frac{1}{n_0} \left[\sum_{i=1}^n k_i (\bar{X}_{i\cdot} - \hat{\mu}_x)^2 - \frac{n-1}{N-n-1} \sum_{i=1}^n \sum_{j=1}^{k_i} (X_{ij} - \bar{X}_{i\cdot})^2 \right]. \quad (2)$$

Table 1
SIDE analysis of variance (ANOVA) table

Source	Df	SS	E[MS]
Individual	$n-1$	$\sum_{i=1}^n k_i (\bar{X}_{i\cdot} - \hat{\mu}_x)^2$	$\frac{n_0}{n-1} \sigma_x^2 + \mu_A$
Residual	$N-n-1$	$\sum_{i=1}^n \sum_{j=1}^{k_i} (X_{ij} - \bar{X}_{i\cdot})^2$	μ_A
Total	$N-2$		

The final step involves transforming the estimated usual intake distribution from the normal scale back to the original scale. This is a very important step because consumption patterns are defined in the original scale. The transformation is given by the expectation of daily intake Y in the original scale when we know the usual intake value in the normal scale \ddot{x}_i : $E[Y|x = \ddot{x}_i] = E[g^{-1}(x + u)|x = \ddot{x}_i] = h(\ddot{x}_i)$, where g is the step 2 transformation to normality, u is the measurement error, and h is the desired transformation.

Once the distribution of the usual intake in the population has been estimated, the analyst is generally interested in the following statistics: mean usual intake, usual intake percentiles and the proportion of the population above or below a given consumption limit. However, it is important to tailor the way SIDE is used to the survey's sampling plan and the method used to estimate the sampling variability. Because of the number and complexity of the steps

executed by SIDE, this is not a simple exercise. An important aspect of this adjustment process is described in the next section.

4. Problem in Estimating the Usual Intake Variance

4.1 The problem

Algebraically, the usual intake variance estimator $\hat{\sigma}_x^2$ given in equation (2) can take negative values. When it is negative, the estimated usual intake distribution is not defined. When that happens, SIDE stops executing, and the user is left with no estimate. The negative values occur because the method of moments is used to estimate σ_x^2 . One might suppose that a better approach would be to produce an estimator by another method that yields only positive values, but developing such an estimator is a laborious task with the measurement error model in (1).

The probability of producing a negative estimate of the variance parameter when the method of moments is used is not zero; for more details, see subsection 4.3. We need to bear this in mind when doing the point estimation of the usual intake distribution. Similarly, when bootstrap replication is used to measure the sampling variability, it is possible to obtain negative variance estimates for the individual replicates. In theory, some proportion of the replicates should have a negative variance estimate. Hence, when we estimate the usual intake distribution for a population, it is important to keep in mind that the variance parameter for the point estimate and the sampling variance estimate can take negative values.

4.2 The solutions

4.2.1 Case 1: The point estimate of the usual intake variance is negative

Option 1: Use a variance estimator truncated to zero

The acceptable values for the usual intake variance σ_x^2 are zero or greater than zero, while the $\hat{\sigma}_x^2$ estimator can take negative values. The first solution to the negative variance estimate problem is to use $\hat{\sigma}_x^{2*} = \max(\hat{\sigma}_x^2, 0)$ instead of $\hat{\sigma}_x^2$ as the σ_x^2 estimator. For the bootstrap replicates, we will use $\hat{\sigma}_x^{2*(b)} = \max(\hat{\sigma}_x^{2(b)}, 0)$. Note that $\hat{\sigma}_x^{2*}$ is a positive-biased estimator since $\hat{\sigma}_x^2$ is unbiased. When $\hat{\sigma}_x^2$ is negative, $\hat{\sigma}_x^{2*}$ is zero. This means that usual intake does not vary from person to person and therefore that the entire population concerned has exactly the same usual intake μ_x . As a result, the estimated usual intake distribution is discrete, with the entire mass of probability concentrated at the mean. With that distribution, it is possible to estimate the mean usual intake and the proportion of population that is above or below a particular consumption threshold (the estimated probability is either 0 or 1). However, the percentiles are not defined for such a distribution, which prevents the analyst from producing a percentile estimate from the estimated distribution.

Option 2: Set a value for the usual intake variance based on an external source

It may make more sense for an analyst to use a more reliable, more stable external source to estimate the variance of the usual intake of the population being studied. A negative or zero variance estimate for the point estimate might be an indication that the data are not sufficiently precise to estimate the variance parameter. To justify this approach, we have to assume that the usual intake variance for the domain concerned and the variance for the external source are equal. This approach is easy to implement since we generally have external sources for which the assumption is reasonable and because SIDE allows us to preset a value for σ_x^2 through the input parameters.

To estimate the sampling variance, we must use the same method as for the point estimate. We therefore have to set σ_x^2 values for the individual replicates. If we set the same value for all the replicates, the sampling variability measured by the bootstrap replication will be solely due to the variation from replicate to replicate in the estimation of the mean and the transformations. This amounts to assuming that the value based on the external source is known exactly, with no sampling error, which is assuming a great deal. If possible, we will not make that assumption, and we will address the problem by varying the values from replicate to replicate to imitate the sampling variation due to the usual intake variance estimate based on the source. It is relatively simple to follow this method when the source data are from the survey. For example, in the estimation process for an Atlantic province such as Prince Edward Island, we could set variance values for the point estimate and the 500 bootstrap estimates based on the Atlantic sample as a whole instead of using the more unstable values from the provincial sample.

4.2.2 Case 2: The point estimate of the usual intake variance is positive (but the variance of some bootstrap replicates is negative)

Option 1: Remove problematic replicates from the sampling variance computation.

When the main estimate yields a positive usual intake variance estimate and some replicates have a negative estimate, it is perfectly natural to want to remove those replicates from the sampling variance calculations. This approach biases the sampling variance estimate, since some combinations of primary sampling units are eliminated from the variance calculation. That bias will be small if the number of replicates eliminated is small. Consequently, we should adopt this approach if the number of replicates to be removed from the sampling variance calculation is small.

Option 2: Use a variance estimator truncated to zero

An alternative that allows the use of information from every replicate is to use the usual intake variance estimator $\hat{\sigma}_x^{2*(b)} = \max(\hat{\sigma}_x^{2(b)}, 0)$. This estimator is positive-biased. On the other hand, the sampling variance estimator derived from it is less biased than in the previous solution because all combinations of primary sampling units are included in the calculation.

One consequence of this approach is the fact that the shape of the point-estimated usual intake distribution will be continuous, while the shape of the distributions based on replicates for which $\hat{\sigma}_x^{2*(b)}$ is zero will be discrete, with all of the probability mass concentrated in the mean. When the statistics based on those distributions are the mean usual intake or the proportion of the population above or below a particular threshold, the sampling variability is easy to calculate because the statistics are well defined for both the point estimate and all the replicates. However, when the statistic of interest is a percentile, it is not defined for replicates that have a variance estimate of zero. One solution to the problem is to use the Woodruff method (1952) adapted for the bootstrap context. The Woodruff method is applicable here because, even though it is impossible to produce percentile estimates for all the replicates, it is possible to estimate probabilities (areas under the curve) for each replicate.

The process of constructing the confidence interval is based on the application of the Woodruff method (1952) described in Lohr (1999), section 9.5.2. Let F_x be the usual intake distribution function, \hat{F}_x the point estimation of that distribution and the percentile of interest $\theta_q = F_x^{-1}(q)$. The level $1 - \alpha$ bootstrap confidence interval for $F_x(y)$

is given by $\hat{F}_x(y) \pm z_{\alpha/2} \sqrt{\hat{V}_B[\hat{F}_x(y)]}$, where index B denotes the variance estimator obtained by bootstrap replication and $z_{\alpha/2}$ is the $100 \times \alpha / 2$ -percentile of the standardized normal distribution. We can calculate the bootstrap variance by including information from all the replicates because it is an interval for the proportion of the population whose usual intake is below threshold y . We can therefore construct a confidence interval for $F_x(\theta_q)$, from which the confidence interval for θ_q can be derived. If the sample distribution of $\hat{F}_x(\theta_q)$ is approximately normal, since F_x and \hat{F}_x are continuous, we have:

$$\begin{aligned}
0.95 &\approx P\left(\hat{F}_x(\theta_q) - z_{\alpha/2} \sqrt{\hat{V}_B[\hat{F}_x(\theta_q)]} \leq q \leq \hat{F}_x(\theta_q) + z_{\alpha/2} \sqrt{\hat{V}_B[\hat{F}_x(\theta_q)]}\right) \\
&= P\left(\hat{F}_x^{-1}\left\{q - z_{\alpha/2} \sqrt{\hat{V}_B[\hat{F}_x(\theta_q)]}\right\} \leq \theta_q \leq \hat{F}_x^{-1}\left\{q + z_{\alpha/2} \sqrt{\hat{V}_B[\hat{F}_x(\theta_q)]}\right\}\right).
\end{aligned}$$

Hence, the approximate level $1 - \alpha$ interval for quantile θ_q is given by:

$$\left[\hat{F}_x^{-1}\left\{q - z_{\alpha/2} \sqrt{\hat{V}_B[\hat{F}_x(\hat{\theta}_q)]}\right\}, \hat{F}_x^{-1}\left\{q + z_{\alpha/2} \sqrt{\hat{V}_B[\hat{F}_x(\hat{\theta}_q)]}\right\} \right]. \quad (3)$$

The bootstrap variance computed here is based on calculations of probabilities with replicates rather than percentile calculations, which is possible for every replicate, even when the usual intake variance estimate is zero. In practice, the following steps will be carried out in applying the method:

1. Generate a point estimate $(\hat{\theta}_q)$ of percentile θ_q .
2. Estimate the proportion of the population with a usual intake below the value computed in step 1 for each bootstrap replicate $(\hat{F}_x^{(1)}(\hat{\theta}_q), \hat{F}_x^{(2)}(\hat{\theta}_q), \dots, \hat{F}_x^{(500)}(\hat{\theta}_q))$.
3. Calculate the bootstrap variance estimator of $\hat{F}_x(\hat{\theta}_q)$ and construct the corresponding confidence interval

$$\left[\max\left\{0, q - z_{\alpha/2} \sqrt{\hat{V}_B[\hat{F}_x(\hat{\theta}_q)]}\right\}, \min\left\{1, q + z_{\alpha/2} \sqrt{\hat{V}_B[\hat{F}_x(\hat{\theta}_q)]}\right\} \right]. \quad (4)$$

4. Generate a point estimate of the confidence interval percentiles that correspond to the interval limits calculated in step 3. The interval generated corresponds to the one given by equation (4).

The confidence interval given in (4) may be wide if the data are not sufficiently precise. In particular, the lower limit may be 0 and the upper limit may be 1. This will result in the loss of one or both limits of the confidence interval constructed in step 4. This “loss” of information is due to a lack of precision in the data and not to the method used. In that case, the data are too imprecise to provide an exact value for the confidence interval over a proportion given in (4).

Option 3: Set a value for the usual intake variance based on an external source

To make the results more precise than in option 2, we need to make additional assumptions about the usual intake distribution that we want to estimate. One possible solution is to assume that the variance of the usual intakes of the population concerned is equal to the variance for another population (for which we have greater precision in the variance estimate). This approach also has the advantage of correcting for negative variance estimates in the bootstrap replicates. We will use the values of the usual intake variance for the more precise population sample when we estimate the usual intake distribution for the population of interest. This takes us back to option 2 in subsection 4.2.1. It is important to note that in this case, we will set the variance values not only for the bootstrap replicates whose initial variance estimate is negative, but also for the other bootstrap replicates and the main estimate. Indeed, it makes sense to take full advantage of the precision afforded by the additional assumption.

4.3 Estimating the magnitude of the problem

It may be helpful to estimate the problem’s magnitude in terms of the probability that it will arise in a number of situations. First, computing that probability may be useful in constructing the sampling plan at the sample size determination stage. For given values of the parameters in the model described in (1), we can calculate the sample sizes needed to ensure that the probability of having a negative value of $\hat{\sigma}_x^2$ is reasonably small. The larger the numbers of first and second recalls, the smaller the probability will be. That is the approach taken in this survey (Junkins and Vigneault, 2003). We can also compute the probability once the sample is collected to determine how frequently the problem will arise for different values of the model’s parameters. Thus, it can be used to characterize a sample. Finally, the situation in which it is of greatest value to compute the probability in the context of this article is the bootstrap replication process. We can calculate the probability that a replicate will produce a negative variance estimate on the basis of the estimated point values of the parameters of the measurement error model. This shows us

the number of replicates that can be expected to “fail” before we make the calculations for each replicate. This process is worthwhile because the calculations for the entire set of replicates take a significant amount of time to run in SIDE and because it helps us determine in advance which solution we will choose for the “failed” replicates. In particular, we will be able to decide ahead of time whether we will need to enhance the sample’s precision by making an additional assumption about the usual intake variance parameter. Furthermore, we may select one of the options in subsection 4.2 on the basis of the probability of having a negative point estimate and negative bootstrap estimates instead of the fact that the point estimate is negative or non-negative. To compute the probability, we will use the expected values of the usual intake variance parameters.

Because of the number and complexity of the steps in SIDE, it is not easy to compute the probability of obtaining a negative variance estimate. The difficulty is compounded by the complex survey design. Accordingly, we will make several assumptions before calculating the probability. First, we will assume a simple random sampling design, which will allow us to skip the weight equalization step. Second, we will take it for granted that no initial adjustment is required and that normality transformation g and reverse transformation h are simply linear. Third, we will simplify the measurement error model in (1) by assuming that the between-individuals variance is constant from person to person, i.e., that parameter σ_A^2 is zero. The latter assumption changes the measurement error model into a conventional model in which the X_{ij} are normal. With these conditions, we can use Fisher’s distribution to compute the probability of interest:

$$\begin{aligned}
 P(\hat{\sigma}_x^2 < 0) &= P\left(\frac{1}{n_0} \left[\sum_{i=1}^n k_i (\bar{X}_{i\cdot} - \hat{\mu}_x)^2 - \frac{n-1}{N-n-1} \sum_{i=1}^n \sum_{j=1}^{k_i} (X_{ij} - \bar{X}_{i\cdot})^2 \right] < 0\right) \\
 &= P\left(\frac{\frac{1}{n-1} \sum_{i=1}^n k_i (\bar{X}_{i\cdot} - \hat{\mu}_x)^2}{\frac{n_0}{n-1} \sigma_x^2 + \mu_A} \bigg/ \frac{\frac{1}{N-n-1} \sum_{i=1}^n \sum_{j=1}^{k_i} (X_{ij} - \bar{X}_{i\cdot})^2}{\mu_A} < \mu_A \bigg/ \left(\frac{n_0}{n-1} \sigma_x^2 + \mu_A \right)\right) \\
 &= P\left(F_{N-n-1}^{n-1} < \mu_A \bigg/ \left(\frac{n_0}{n-1} \sigma_x^2 + \mu_A \right)\right).
 \end{aligned} \tag{5}$$

If n is large, the right side of the inequality in the last line of equation (5) is approximately equal to $1/[1 + (1 + \gamma)\sigma_x^2 / \mu_A]$ (a constant smaller than 1), where $\gamma = (N - n)/n$ is the number of second recalls divided by the number of first recalls. On the other side of the inequality, Fisher’s statistic has a distribution that becomes more concentrated around 1 as sample sizes increase. Consequently, the probability shrinks if n is fixed and γ increases toward 1 (if the number of second recalls increases toward n) or if γ is fixed and n increases (the number of respondents increases while the proportion of second recalls remains constant). Furthermore, estimator $\hat{\sigma}_x^2$ in (2) is formed from a quadratic combination of Horvitz-Thompson-type estimators. Thus, since the sampling fractions are negligible, most of the estimator’s variability will be due to sampling from a finite population and not to the mechanism that generates the population (Binder and Roberts, 2003). Bootstrap replication in this case provides an estimate of the sampling variability and the total variability. Formula (5) represents the expected proportion of replicates that will produce a negative variance estimate. In practice, however, we do not have values for the parameters. We will therefore replace those values in the formula with the corresponding point estimates (or their expected values) to estimate the probability.

In the general case where the design is complex and all the steps in SIDE must be carried out, formula (5) must be adapted to produce the exact probability. In the applications described above, however, we want a measurement of the problem’s magnitude rather than an exact figure. For that purpose, we can use the probability in (5) in the general case. In practice, the value produced by the formula appears to be very close to the observed proportion of “failed” replicates.

References

- Béland, Y., Dufour, J., MacNabb, L., and Pierre, F. (2003), "Sample Design of the 2004 Canadian Nutrition Survey", *Proceedings of the Survey Methods Section, Statistical Society of Canada*, pp. 91-96.
- Binder, D. A., and Roberts, G. R. (2003), "Design-based and Model-based Methods for Estimating Model Parameters", in R. L. Chambers et C. J. Skinner (eds.) *Analysis of survey data*, New York: Wiley, pp. 29-48.
- Brisebois, F., and Thivierge, S. (2001), "The Weighting Strategy of the Canadian Community Health Survey", *Proceedings of the Survey Research Methods Section, American Statistical Association*.
- Junkins, B., and Vigneault, M. (2003), "Number of Repeat Recalls for CCHS Nutrition Focus Survey", unpublished report, Ottawa, Canada: Health Canada.
- Lohr, S.L. (1999), *Sampling: Design and Analysis*, Duxbury Press, New York.
- Nusser S. M., Carriquiry A. L., Dodd K. W., and Fuller W. A. (1996), "A Semiparametric Transformation Approach to Estimating Usual Daily Intake Distributions", *Journal of American Statistical Association*, 91, pp. 1440-1449.
- Rao, J. N. K., and Wu, C. F. J. (1988), "Resampling Inference with Complex Survey Data", *Journal of the American Statistical Association*, 83, pp. 231-241.
- Woodruff R. S. (1952), "Confidence Intervals for Medians and Other Position Measures", *Journal of American Statistical Association*, 47, pp. 635-646.