A Better Understanding of Weight Transformation Through a Measure of Change

Johane Dufour, François Gagnon, Yves Morin, Martin Renaud and Carl-Erik Särndal

Abstract

The literature on longitudinal surveys of households offers several approaches for creating a set of final weights for use in data analysis. Most of these approaches depend on various procedures for modifying weights. Initial weights are often transformed into a set of intermediate weights in order to compensate for nonresponse, and then into a set of final weights, through poststratification, in order to adjust the sample. The literature includes a great deal of information about this approach but none of the studies has really looked closely at an approach for measuring the relative importance of these two steps in measuring the effectiveness of the numerous existing alternatives for creating intermediate weights. The objective of this paper is to study and measure the change (from the initial to the final weight) which results from the procedure used to modify weights. A breakdown of the final weights is proposed in order to evaluate the relative impact of the nonresponse adjustment, the correction for poststratification and the interaction between these two adjustments. This measure of change is used as a tool for comparing the effectiveness of the various methods for adjusting for nonresponse, in particular the methods relying on the formation of Response Homogeneity Groups. The measure of change is examined through a simulation study, which uses data from a Statistics Canada longitudinal survey, the Survey of Labour and Income Dynamics. The measure of change is also applied to data obtained from a second longitudinal survey, the National Longitudinal Survey of Children and Youth.

Key Words: Nonresponse; Weighting; Calibration; Longitudinal survey; Measure of change.

1. Introduction

The literature contains many two-step approaches to transforming weights for household surveys. The first step involves an adjustment of the initial weights in order to compensate for nonresponse; the resulting weights are called intermediate weights. The second step produces the final weights through the process of poststratification, or more commonly through calibration (see Deville and Särndal 1992), in order to ensure that the final weights respect certain known population control totals. All of these weight modifications are designed to produce the “best possible set of final weights”.

At Statistics Canada, longitudinal surveys of households also use this two-step approach in weighting, and the research work undertaken by the Agency leans in this direction. The U.S. Bureau of the Census “Survey of Income and Program Participation (SIPP)” (see Rizzo, Kalton and Brick 1996) also uses this type of approach.

Several methods are recommended in the literature for adjusting weights to compensate for nonresponse. Rizzo et al. (1996) compared the estimates obtained through several of these methods to estimates from independent sources. However, not many authors have done simulations or proposed tools for comparing the relative effectiveness of the methods in terms of their ability to reduce the nonresponse bias.

The main objective of this document is to study and measure the change (between initial and final weights) resulting from the adoption of a two-step procedure for modifying weights. Thus, a measure of change involving four components is proposed in order to quantify the relative impact of the nonresponse adjustment, the correction for poststratification and the interaction between these two adjustments. The second objective is to use the measure of change to compare the effectiveness of the different nonresponse adjustment methods through a simulation study based on data from the Longitudinal Survey of Labour and Income Dynamics (SLID) and from the National Longitudinal Survey of Children and Youth (NLSCY). The longitudinal surveys are unique in that a great deal of information about respondents and non-respondents to the latest wave is available from respondents to the previous waves. Thus, more complex methods can be used to adjust for nonresponse.

A general framework for the weighting of longitudinal surveys of households is presented in section 2. Then, the measure of change which will be used to quantify the stages of transformation between the initial and the final weights is presented in section 3. Section 4 addresses the nonresponse adjustment strategies contained in the literature. This is followed by sections 5 and 6, which contain the results of the studies based on the SLID and NLSCY. The last section presents the conclusions of this study.

2. General framework for longitudinal weighting

In a longitudinal survey of households, individuals in the initial sample are followed over time, and are referred to as longitudinal individuals. This set of individuals is the one
which will be used in the studies presented in this document. They are referred to as the "reference unit". This section provides an overview of the steps followed in order to modify the initial weight for longitudinal individuals into a final weight.

2.1 Initial Weights

\[ U = \{1, \ldots, k, \ldots, N\} \] is a finite population. We are interested in variable \( y \) (the variable of interest), whose value for the \( k \)th unit is recorded as \( y_k \). The objective is to estimate the total \( Y = \sum_{k \in U} y_k \). Let \( w_{0k} \) be the initial weight for all \( k \in s \) units, where \( s \) is the longitudinal sample. In the absence of nonresponse, the set of initial weights \( \{w_{0k} : k \in s\} \) yields the \( \hat{Y} = \sum_{k \in s} w_{0k} y_k \) estimator for \( Y \). In this case we assume that the \( w_{0k} \) are normalized in order to ensure that \( \sum_{k \in s} w_{0k} = N \). Although \( \hat{Y} \) is unbiased for \( Y \), \( \hat{Y} \) has the drawback of not incorporating any ancillary information in the form of known control totals for poststrata.

2.2 Nonresponse Adjustment and Intermediate Weights

Most surveys have to deal with nonresponse. Two approaches are often used to compensate for this: imputation and the correction of the initial weights of respondents through an adjustment factor. The latter is the one more commonly used in household surveys to compensate for total nonresponse, while imputation is often preferred when dealing with partial nonresponse. Total nonresponse reduces the size of the sample since the \( y_k \) value is only available for \( k \in r \), where \( r \subset s \) is the set of the \( m \) responding units. For this reduced set of data, the initial \( w_{0k} \) weights are, on average, too small and we have \( \sum_{k \in r} w_{0k} < N \). The estimator \( \hat{Y}^{\prime} = \sum_{k \in r} w_{0k} y_k \) is not admissible since it systematically underestimates \( Y \).

Weight adjustment is often chosen in order to compensate for total nonresponse in household surveys. A common method of adjusting weights involves constructing Response Homogeneity Groups (RHGs). These are designed so that each one is comprised of reference units having a similar probability of response. Then, within each RHG, an adjustment factor equal to the inverse of the RHG’s response rate (weighted or not) is calculated. For each respondent unit \( k \), the adjustment factor for nonresponse involves multiplying \( w_{0k} \) by the RHG’s adjustment factor. This operation results in a set of intermediate weights \( \{w_{r0k} : k \in r\} \), where \( \sum_{k \in r} w_{r0k} = N \). With these weights, we can construct the estimator \( \hat{Y}^{\prime \prime} = \sum_{k \in r} w_{r0k} y_k \), which eliminates the underestimation which is characteristic of \( \hat{Y}^{\prime} = \sum_{k \in r} w_{0k} y_k \). As in the case of the set of initial weights, the main drawback with this set is that it fails to incorporate the ancillary information available for poststrata.

2.3 Poststratification and Final Weights

A widely-used practice in household surveys involves modifying the intermediate weights through poststratification, or, more commonly, through calibration, so that the sum of the final weights on the set of respondents will correspond to the known population counts. Thus, poststratification produces a set of final weights \( \{w_{2k} : k \in r\} \), which incorporates the ancillary information and which is also consistent with the control totals for the poststrata. In this case, the final weights in each poststratum \( p \) confirm \( \sum_{r \in p} w_{2k} = N_p \), where \( N_p \) is the known element and \( r_p \) is the set of respondent units in the \( p \)th poststratum. It follows that \( \sum_{r \in p} w_{2k} = N \). Demographic and geographic variables are frequently used to define poststrata. The choice of poststrata, which must be sufficiently large, is limited by the availability of control totals. Several methods may be used to calibrate the intermediate weights to the selected control totals.

3. Measure of Change from Initial to Final Weights

In this section, a measure of the change between initial and final weights is presented so to better understand the effect of the weight modification procedure. The breakdown of this measure into four components makes it possible to quantify the effect of each of the weighting steps described in section 2. These components will be used in sections 5 and 6 in the comparison of various methods for adjusting weights for nonresponse.

If the initial weights are normalized so that \( \sum_{r \in p} w_{0k} = N \), and if \( r \subset s \), then the three sets of weights described in section 2 confirm the following relations:

\[
\sum_{r \in p} w_{0k} < N, \quad \sum_{r \in p} w_{1k} = N, \quad \sum_{r \in p} w_{2k} = N.
\]

Let

\[
\bar{\pi}_0 = \sum_{r \in p} \frac{w_{0k}}{w_{0k}} \quad \text{and} \quad \bar{\pi}_0 = \sum_{r \in p} \frac{w_{2k}}{w_{0k}}.
\]

The ratio \( \bar{\pi}_0 \) measures the average change in the intermediate weight set in relation to the initial weight set. As total nonresponse becomes more pronounced, \( \bar{\pi}_0 \) shifts further away from the value of 1, which is only obtained in the absence of nonresponse. The ratio \( \bar{\pi}_2 \) represents the average change in the set of final weights in relation to the set of initial weights.

The \( \bar{\pi}_0 \) and \( \bar{\pi}_2 \) ratios measure the average change in weight. To measure an individual change in weight, we define, for every \( k \in r \), \( \bar{\pi}_{0k} = w_{1k} / (w_{0k} \bar{\pi}_0) \), and \( \bar{\pi}_{2k} = w_{2k} / (w_{0k} \bar{\pi}_2) \). These quantities vary around 1. More specifically, their weighted averages equal 1:

\[
\sum_{r \in p} w_{0k} \bar{\pi}_{0k} = \sum_{r \in p} w_{0k} \bar{\pi}_{2k} = 1.
\]
The \( r_{01k} \) and \( r_{02k} \) quantities will be useful for measuring individual weight changes.

The total weight change, from the set of initial to final weights, going through the set of intermediate weights, can be calculated by a measure of change, also called distance. Here, \( D \) is the following measure of change:

\[
D = \frac{\sum_r w_{0k} \left( \frac{w_{2k}}{w_{0k}} - 1 \right)^2}{\sum_r w_{0k}}.
\]

In fact, \( D \) is a weighted average of the following individual weight change factors:

\[
\left( \frac{w_{2k}}{w_{0k}} - 1 \right)^2 = \left( \frac{w_{2k}}{w_{1k}} \frac{w_{1k}}{w_{0k}} - 1 \right)^2.
\]

The measure of change \( D \) breaks down into four components, as set out in the following equation:

\[
D = R_{01} + R_{12} + R_{\text{int}} + G
\]

where

\[
R_{01} = \bar{m}_{02}^2 \frac{\sum_r w_{0k} (r_{01k} - 1)^2}{\sum_r w_{0k}},
\]

\[
R_{12} = \bar{m}_{02}^2 \frac{\sum_r w_{0k} (r_{02k} - r_{01k})^2}{\sum_r w_{0k}},
\]

\[
R_{\text{int}} = 2 \bar{m}_{02}^2 \frac{\sum_r w_{0k} (r_{01k} - 1)(r_{02k} - r_{01k})}{\sum_r w_{0k}},
\]

and

\[
G = (\bar{m}_{02} - 1)^2.
\]

It should be noted that the measure of change \( D \) is always positive, equality being at zero when the two following conditions are met:

(i) absence of nonresponse \((r = s \text{ and } w_{1k} = w_{0k} \text{ for all } k)\),

(ii) absence of poststratification effect on the intermediate weights \((w_{2k} = w_{1k} \text{ for all } k)\).

A high nonresponse rate would tend to increase the value of the measure of change \( D \) since in such a case, \( w_{1k} \) is generally much larger than \( w_{0k} \).

\( R_{01} \) measures the individual weight changes which result from going from the initial to the intermediate set. Later, we will see that the component \( R_{01} \) is somehow associated with the quality of the nonresponse model and that a large \( R_{01} \) value is preferable. \( R_{12} \) measures the individual weight changes which result from going from the intermediate to the final set. \( R_{\text{int}} \) measures the interaction between the two types of change and \( G \) measures the change in average weight between the initial and final sets.

In addition to its interpretation as a distance, the measure of change \( D \) can also be interpreted as a mean square error of changes \( w_{2k} / w_{0k} \) in relation to 1, and in relation to the distribution defined by all the \( w_{0k} \). From this perspective, the component \( G \) corresponds to the bias squared (or the square of the difference between the \( w_{0k} \) average of \( w_{2k} / w_{0k} \) and 1), while the sum of the other three components corresponds to the variance. In the simplest case, where a nonresponse adjustment is calculated using a single RHG, and where no poststratification is applied, we have \( w_{0k} = N / n \) for all \( k \in s \) (in the case of a size \( n \) simple random selection) and \( w_{1k} / w_{2k} = N / m \) for all \( k \in r \), (where the nonresponse adjustment factor is \( n/m \), i.e., the inverse of the response rate). We then have \( D = G = \{(n/m)-1\}^2 \) and \( R_{01} = R_{12} = R_{\text{int}} = 0 \).

Some significant conclusions may be drawn from looking at the relative importance of \( R_{01} \), \( R_{12} \) and \( R_{\text{int}} \) if \( R_{01} \) is high at the same time that \( R_{12} \) is not very high, the survey is one in which the nonresponse adjustment creates significant individual changes in weights, while poststratification only results in a slight change in individual weights. However, when \( R_{12} \) is high, poststratification brings about very large individual changes. The results presented in sections 5 and 6 will show that \( R_{01} \) can be used to compare the effectiveness of various nonresponse adjustment methods. As well, the sign of \( R_{\text{int}} \) indicates whether the two types of individual change are moving in the same direction (\( R_{\text{int}} > 0 \)) or in opposite directions (\( R_{\text{int}} < 0 \)). In reality, we expect \( R_{\text{int}} \) to be very small, if not negligible.

### 4. Nonresponse Adjustment Strategies

The literature contains several methods for adjusting weights (including the method described in section 2.2) to compensate for nonresponse. Another method, which is frequently used in longitudinal surveys, involves adjusting weights in accordance with the inverse of the predicted probability of response obtained through a logistic regression. We also find methods of adjustment based on calibration, which use marginal distributions of the initial sample or of the population. Singh, Wu and Boyer (1995) used this approach in order to derive a method of adjustment capable of producing coherent estimates in longitudinal surveys from one wave to the next. Deville (1998) recommended a method of correction for nonresponse by calibration or balanced sampling. For a review of nonresponse adjustment methods, refer to Kalton and Kasprzyk (1986), Platek, Singh and Tremblay (1978), Chapman, Bailey and Kasprzyk (1986) and to Little (1986). In this document, only methods relying on the creation of RHGs are considered.
4.1 Formation of RHGs

In most surveys, aside from a few stratification variables from the sample frame, very little information is available about non-respondents. Therefore, the choice of RHGs is very limited and the strata are often used as RHGs. In these cases, the assumption is that the probability of response is the same for all units in a given stratum. However, in longitudinal surveys, a great deal of information about respondents and non-respondents in the current wave is available from the responses provided in the previous waves. This information can then be used to create RHGs within which the assumption of a uniform response mechanism is plausible. This leads to a better nonresponse adjustment and, therefore, a reduction in the risk of introducing a nonresponse bias into the estimates.

4.1.1 Method for the Selection of Variables for the Formation of RHGs

By definition, an RHG is formed from a set of variables capable of predicting the propensity to respond. If the set of variables which is defined at the outset is too large, univariate tests may be used to isolate the most important variables to distinguish the characteristics of respondents from those of nonrespondents. With this set of important variables, a selection method may then be applied for retaining the best variables for explaining the propensity to respond. Two of the current variable selection methods are: the Logistic Regression Model (LR) and the Segmentation Model (SM).

4.1.1.1 Logistic Regression

Under the LR method, the combined use of the “fact of having responded to the survey or not” as a dependent variable, standardized weights and the “stepwise” procedure result in a list of the most significant dichotomous variables for explaining the propensity to respond. As a general rule, RHGs are created according to $2^q$ possible combinations, based on a set of $q$ explanatory variables used. The LR is often referred to as the symmetrical approach. However, if certain additional constraints are applied when the RHGs are created, this could reduce their numbers. For instance, we could require a minimum number of reference units ($n$) and a response rate (RR) (weighted or not weighted) greater than a certain level in each of the RHGs. Kalton and Kasprzyk (1986) encourage the use of such constraints in order to avoid increasing the variance associated with extreme weights. However, these constraints may reduce the effectiveness of the nonresponse adjustment and result in an increase in the bias. When an RHG does not meet one of these constraints, it has to be combined with another RHG. The combination of RHGs continues until all of the RHGs meet the additional constraints imposed. This leads to $2^q - J$ valid combinations, where $J$ represents the reduction resulting from the combination of RHGs.

4.1.1.2 Segmentation Model

The SM method, which is referred to as non-symmetrical, is based on the CHAID (Chi-square Automatic Interaction Detection) algorithm developed by Kass (1980). It divides the sample into sub-groups according to the response rate of the explanatory variables by using a Chi-square test. The segmentation process continues until a significant explanatory variable is found. The final sub-groups created through the SM become the RHGs, for which the nonresponse adjustments are calculated. As in the case of the LR, additional constraints may be imposed.

In Figure 1 we see that the SM method divided the sample into several RHGs based on the different explanatory variables. The RHGs are once again represented by the shaded boxes. The segmentation continues until it is no longer possible to find explanatory variables.

4.1.2 Nonresponse Adjustment Factor

Whether the RHGs are formed by relying on the LR or the SM, a uniform response mechanism is assumed within each RHG. Thus, the nonresponse adjustment factor is given by the inverse of the response rate (weighted by $w_{0k}$ or not weighted) for the RHG.

5. Empirical Study Based on the Survey of Labour and Income Dynamics (SLID)

Data from the SLID were used for an empirical study designed to compare the effectiveness of the LR and SM. The SLID is a longitudinal survey of households that started in 1993; one of its objectives is to provide information on the economic well-being of Canadian society (see Lavigne and Michaud 1998).

These two methods were tested through a simulation by analyzing some variables of interest and various domains. The components of the measure of change, the absolute and relative biases and the variances were studied.

5.1 Description of the Empirical Study

The first step in the empirical study was to estimate the probability of response to the first wave of the survey for each of the units in the longitudinal sample. Variables which could potentially explain the propensity to respond (based on a preliminary interview) were used to form a very large number of RHGs. All of the individuals in the sample were assigned to an RHG on the basis of the values of the explanatory variables. A probability of response was then estimated for each RHG on the basis of the weighted response rate. Then, only the respondents and their
The probability of response were retained in the reference sample for the simulation. Nonresponse was then generated for the reference sample through Poisson sampling. This procedure, illustrated in Figure 2, was independently repeated 100 times, thus creating 100 sets of respondents and nonrespondents. The average response rate for each repetition was around 90%, which was the rate observed in the first wave of the SLID.

For each of the 100 repetitions, a nonresponse adjustment was done using the LR method to create the RHGs. Similarly, a nonresponse adjustment was done using the SM to create RHGs for each of the first 20 repetitions. With the SM approach, the number of repetitions was limited to 20, given the stability of the results and since several manual interventions and the use of a specific software package (in our case: Knowledge Seeker - ANGOSS Software 1995) were required.

Several variants of the variable selection method were studied:

a) LR_i, where i represents, out of the 100 repetitions, the approximate average of the number of RHGs generated through the LR method. In this study, i = 4, 16, 40, 60. For instance, for LR_40, the q = 6 most important explanatory variables for the propensity to respond were first identified. The RHGs were then formed using the \(2^q - J\) valid combinations of these q = 6 explanatory variables. The imposition of additional constraints \(n > 30\) and \(RR > 50\%) in each RHG led to the re-grouping of some RHGs. On average, out of 100 repetitions, 24 RHGs had to be regrouped \(J = 24\) and a total \(2^q - J = 2^6 - 24 = 40\) RHGs were formed, hence the LR_40 designation. In the simulation study, LR_i, where i = 4, 16, 40, 60 RHGs corresponds, respectively to \(q = 2, 4, 6, 8\) explanatory variables.

b) SM_i, where i indicates the approximate average in the first 20 repetitions of the number of RHGs generated through the SM method. In this study, i = 16, 25, 40. For example, for SM_16, one SM was used with a significance level \(p\) of 0.0001. After the imposition of the same additional constraints as for the LR, an average 16 RHGs were created. SM_i, where i = 16, 25, 40 RHGs corresponds, respectively, to the significance levels of 0.0001; 0.0005; 0.0025. The higher the level used, the easier it is to identify the significant differences, which makes it possible to achieve a more detailed segmentation and, hence, a greater number of RHGs.

c) A method with a single RHG (1_RHG) was also used for comparison purposes. This method involves defining the entire sample as a single RHG for each of the 100 repetitions. It should be noted that this method is only effective if the response mechanism is uniform within the entire sample, which is rarely the case. At first, the initial weights were normalized so that, in order to eliminate the effect of undercoverage and to better isolate the effect of nonresponse. Thus, G will only measure the average change caused by the nonresponse adjustment.

At first, the initial weights were normalized so that \(\Sigma_i w_{0i} = N\), in order to eliminate the effect of undercoverage and to better isolate the effect of nonresponse. Thus, G will only measure the average change caused by the nonresponse adjustment.

---

Figure 1. Depiction of the Formation of RHGs by Method.
Once the initial weights were normalized, each set of final weights was then the result of a two step process: a nonresponse adjustment (based on one of the eight methods mentioned: 1_GRH, LR_i, where i = 4, 16, 40, 60 and SM_i, where i = 6, 25, 40) and a same poststratification (14 age-sex groups by province).

5.2 Analysis of the Results of the Empirical Study

For each of the methods discussed in the previous section, the components of the measure of change D were studied. Also, the average, absolute and relative nonresponse bias and the average variance of the estimates were analyzed.

5.2.1 Measure of Change (D)

Table 1 presents the average value of D and its components for each of the M repetitions (where M = 100 for the LR and M = 20 for the SM) as well as the percentage contribution of each element to the average value of D. We observe, in the first place, that for the 1_GRH method, $R_{01}$ is nil since one single nonresponse adjustment was made to the set of respondents. Thus, $w_{ik} = \alpha w_{0k}$, where $\alpha$ is a constant, so $R_{01} = 1$ for every $k \in r$ and $R_{01} = 0$. We also observe that $D$ increases as the number of RHGs increases, irrespective of whether the LR or SM method is used. Thus, the more RHGs there are to compensate for nonresponse, the greater the total change to which the weights are subjected. In addition, the values of $D$ are higher for the SM than for the LR.

For the LR and the SM, the contribution of $R_{01}$ to the measure of change increases as the number of RHGs increases, since nonresponse is more readily targeted as the number of RHGs increases. Consequently, the nonresponse adjustment often becomes more important and, thereby, the weights vary more and more. In addition, the contribution of $R_{1r}$ to the measure of change is much more important with the SM than with the LR. This indicates that the SM seems to be better at modeling nonresponse and isolating the specific trends of the LR.

Table 1

<table>
<thead>
<tr>
<th>Method</th>
<th>$D$</th>
<th>$R_{01}(x10^{-3})$</th>
<th>$R_{01}/D(%)$</th>
<th>$R_{12}(x10^{-3})$</th>
<th>$R_{12}/D(%)$</th>
<th>$R_{int}(x10^{-5})$</th>
<th>$R_{int}/D(%)$</th>
<th>$G(x10^{-2})$</th>
<th>$G/D(%)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_RHG</td>
<td>0.012135</td>
<td>0.00</td>
<td>0.00</td>
<td>1.17</td>
<td>9.66</td>
<td>0.00</td>
<td>0.00</td>
<td>1.11</td>
<td>90.34</td>
</tr>
<tr>
<td>LR_4</td>
<td>0.012952</td>
<td>0.78</td>
<td>6.04</td>
<td>1.10</td>
<td>8.49</td>
<td>0.06</td>
<td>0.01</td>
<td>1.11</td>
<td>85.46</td>
</tr>
<tr>
<td>LR_16</td>
<td>0.013809</td>
<td>1.66</td>
<td>11.97</td>
<td>1.00</td>
<td>7.31</td>
<td>3.76</td>
<td>0.54</td>
<td>1.11</td>
<td>80.19</td>
</tr>
<tr>
<td>LR_40</td>
<td>0.014426</td>
<td>2.32</td>
<td>16.02</td>
<td>0.96</td>
<td>6.66</td>
<td>4.02</td>
<td>0.55</td>
<td>1.11</td>
<td>76.77</td>
</tr>
<tr>
<td>LR_60</td>
<td>0.01498</td>
<td>2.85</td>
<td>19.00</td>
<td>0.95</td>
<td>6.35</td>
<td>3.75</td>
<td>0.49</td>
<td>1.11</td>
<td>74.15</td>
</tr>
<tr>
<td>SM_16</td>
<td>0.015712</td>
<td>3.42</td>
<td>21.33</td>
<td>0.97</td>
<td>6.19</td>
<td>3.40</td>
<td>0.43</td>
<td>1.11</td>
<td>72.05</td>
</tr>
<tr>
<td>SM_25</td>
<td>0.016713</td>
<td>4.44</td>
<td>26.02</td>
<td>0.95</td>
<td>5.73</td>
<td>2.95</td>
<td>0.36</td>
<td>1.11</td>
<td>67.89</td>
</tr>
<tr>
<td>SM_40</td>
<td>0.018202</td>
<td>5.97</td>
<td>32.37</td>
<td>0.95</td>
<td>5.23</td>
<td>1.20</td>
<td>0.14</td>
<td>1.11</td>
<td>62.26</td>
</tr>
</tbody>
</table>
As for $R_{12}$, it is almost constant, regardless of which method and number of RHGs are used. However, despite the fact that it changes very little, its contribution to the measure of change diminishes as the number of RHGs increases. This is due to the fact that there is more variation in the weights with a nonresponse adjustment, and the modifications which poststratification creates in the weights are less and less important as the number of RHGs increases.

In the case of $R_{int}$, its value is negligible and its contribution to the measure of change is very small. This means that the interaction between the nonresponse adjustment and poststratification is practically nil.

Finally, $G$ remains constant, irrespective of which method and how many RHGs are used. As with $R_{12}$, the contribution of $G$ to the measure of change diminishes as the number of RHGs increases. A larger number of RHGs is better at targeting nonresponse, thereby causing more variations in the set of intermediate weights.

Since, with all of these methods, $G$ is constant, $R_{int}$ is close to zero and $R_{12}$ is nearly constant, it is clear that the variations in $D$ are mostly influenced by the variations in $R_{01}$.

Graph 1 shows the average contribution in percentage of $R_{01}$ and $R_{12}$ to the measure of change. For LR and SM, the contribution of $R_{01}$ increases with the number of RHGs while that of $R_{12}$ diminishes. Also, the contribution of $R_{01}$ is greater for SM than for LR, while that of $R_{12}$ is less for SM than for LR. The profile of the contribution of $R_{01}$ is the same as the profile of $D$ (Table 1). This confirms that the variations in the measure of change are mainly due to the variations in $R_{01}$.

Graph 2 shows the comparison between the LR and SM in terms of the average percentage contribution in percentage of $R_{01}$ to $D$. For a given number of RHGs, $R_{01}$ contributes to a larger percentage of $D$ through the SM method than through the LR method. This means that individual changes in the weights between the initial and intermediate sets are greater for SM than for LR.
5.2.2 Relative and Absolute Biases

The Relative Bias (RB) and the Absolute Bias (AB) were used to compare the performance of LR relative to SM in reducing the nonresponse bias:

\[ RB_i = 100 \left( \frac{\hat{Y}_i - Y}{Y} \right) \quad \text{and} \quad AB_i = \hat{Y}_i - Y; \]

where \( \hat{Y}_i \) is the estimate of the variable of interest obtained for the \( i \)-th repetition, \( i = 1, 2, \ldots, M \), \( M = 100 \) for the LR, \( M = 20 \) for the SM and \( Y \) is the total for the variable of interest obtained from the reference sample.

The Average Relative Bias (ARB) and the Average Absolute Bias (AAB) are calculated by taking, respectively, the average of the RB and the AB for all repetitions:

\[ \text{ARB} = \frac{1}{M} \sum_{i=1}^{M} RB_i \quad \text{and} \quad \text{AAB} = \frac{1}{M} \sum_{i=1}^{M} AB_i \]

where \( M = 100 \) in the case of the LR and \( M = 20 \) in the case of the SM.

For the 100 repetitions, national estimates were produced for the following three variables: “person living, or not, in a family whose revenue is less than the Low Income Cutoff (LICO)”, “Individual Total Income (TI)” and “Individual Wages and Salaries (WS)”. The ARB for each estimate was calculated for the eight methods under study. Given the large sample size, the low nonresponse rate (10%) and the fact that a large number of control totals was used for poststratification, the ARB is very small (see Table 2) for each of the methods used.

In Table 2 we see that, for each of the three variables, the ARB is more or less constant for the SM, irrespective of how many RHGs are used. Also, for the LR, the ARB for the TI and SW is more or less constant not withstanding the number RHGs used. On the other hand, for the LICO, the ARB for method LR_4 is much smaller than the ARB for the other three LR methods. This could be due to the fact that the LICO is a variable derived from several other variables, unlike the TI and the SW, which are observed variables. The ARB for the three variables for method 1_RHG is much larger than the ARB produced by the SM and the LR, except for the LICO, since in this case the ARB is more or less equivalent to the ARB of the LR. Thus, it appears that method 1_RHG does not perform as well as the SM and the LR. In the best case, it is more or less equivalent to LR. Unlike SM, we observe that the progression of ARB is not strictly downwards for the LR, as the number of RHGs increases.

Despite the fact that the ARB is minimal for the variables studied for Canada, it can increase rapidly for small domains. In this study, other domains were also reviewed. Although some variances were observed in several of these cases, it seems that the ARB for the SM is generally smaller than the ARB for the LR and the method 1_RHG. A more detailed study of a larger number of interest and domain variables would be beneficial for corroborating these conclusions.

As previously indicated, the individual changes in the weights caused by the nonresponse adjustments are greater for the SM than for the LR (see Graph 2). This would suggest that the SM is more effective in reducing the nonresponse AB for a fixed number of RHGs. Graph 3 confirms this observation, showing that the AAB for the LICO is smaller through the SM than through the LR method.

5.2.3 Variance Estimates

Variance estimates were produced for the three variables of interest through the Jackknife method. For LICO (Graph 4), the average variance of estimates is approximately the same, regardless of the method used. However, there is a slight decrease when the number of RHGs increases, for both the LR and the SM. Also, based on the empirical study, average variance estimates for the SM are slightly smaller than for the LR. Therefore, the larger dispersion in the weight (a higher value for \( D \)) does not entail an increase in variance.

6. Application to the National Longitudinal Survey of Children and Youth (NLSCY) Data

In this section, most of the analyses done with the help of the LR and SM in the empirical study with data from the SLID are reproduced with the information obtained from the NLSCY. Just like the SLID, the NLSCY is a longitudinal survey of households. It started in 1994 and is designed to collect information for analyzing policies and developing programs addressing critical factors affecting the development of children in Canada (see Michaud, Morin, Clermont and Lafleure 1998).

Table 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>1_RHG</th>
<th>LR_4</th>
<th>LR_16</th>
<th>LR_40</th>
<th>LR_60</th>
<th>SM_16</th>
<th>SM_25</th>
<th>SM_40</th>
</tr>
</thead>
<tbody>
<tr>
<td>LICO</td>
<td>0.37</td>
<td>0.15</td>
<td>0.43</td>
<td>0.37</td>
<td>0.31</td>
<td>0.14</td>
<td>0.12</td>
<td>0.08</td>
</tr>
<tr>
<td>TI</td>
<td>-0.32</td>
<td>-0.09</td>
<td>-0.06</td>
<td>-0.05</td>
<td>-0.06</td>
<td>-0.006</td>
<td>-0.005</td>
<td>0.002</td>
</tr>
<tr>
<td>WS</td>
<td>-0.44</td>
<td>-0.13</td>
<td>-0.15</td>
<td>-0.19</td>
<td>-0.14</td>
<td>-0.10</td>
<td>-0.09</td>
<td>-0.09</td>
</tr>
</tbody>
</table>
6.1 Description and Analysis of the Results of the Application

The following methods were used for this study: LR\(_i\), where \(i = 4, 14, 41, 70\) with, respectively \(q = 2, 4, 6, 8\) variables, and SM\(_i\), where \(i = 19, 36\) with significance levels of 0.001 and 0.005, respectively. The same two constraints imposed for the SLID were re-applied when the RHGs were created. The same poststratification was used (22 age-sex groups by province) for each of the methods under study.

Unlike the empirical study based on the SLID, only the data collected in the first two waves of the NLSCY were used. There was no simulation and the initial weights were not normalized (\(\sum_{i} w_{ik} = \hat{N} < N\)). It should be noted that the undercoverage of the NLSCY is around 13% and its nonresponse is around 8%.

The conclusions drawn from the results presented in Table 3 are similar to those obtained in the simulation (Table 1). However, we observe that the relative contribution by \(R_{i1}\) to the measure of change is weaker for the NLSCY than for the SLID. This result indicates that the nonresponse adjustment of the SLID produces larger individual changes in the weights, thereby resulting in a larger contribution by \(R_{i0}\). Therefore, the nonresponse adjustment in the case of the NLSCY had no significant effect on the individual changes in the weights, contrary to what was observed in the case of the SLID.

The relative contribution by \(R_{i2}\) to the measure of change is higher for the NLSCY than for the SLID. This result indicates that the more refined poststratification of the NLSCY results in greater individual changes in the weights, which translates into a greater contribution of \(R_{i2}\). Therefore, the NLSCY benefits a great deal from poststratification, which is less important for the SLID.
With respect to $R_{int}$, as with the SLID, its contribution to the measure of change is negligible. Contrary to the SLID, the sign of $R_{int}$ is negative, which means that the interaction between $R_{01}$ and $R_{12}$ is negative.

With respect to $G$, as in the case of the SLID, it is the key source of contribution to the measure of change. In the case of the NLSCY, $G$ not only includes the average change in weight resulting from the nonresponse adjustment, but also the average change in weight resulting from the correction for undercoverage through poststratification.

When all of these results are compared, it becomes evident that the two surveys are very similar since $R_{int} = 0$ and the sum of the contributions to the measure of change of $R_{01}$ and $R_{12}$ is around 35% in both cases. However, the NLSCY is also very different from the SLID since $R_{int}$ predominates in the former one, while $R_{01}$ predominates in the latter.

Just as with the SLID, $D$ increases with the number of RHGs and this measure is greater for the SM than for the LR. In fact, the value of $D$ is greater for the NLSCY than for the SLID, mainly because of the NLSCY undercoverage, which results in an increase in $G$ and, therefore, in $D$.

The average contribution of $R_{01}$ for the LR and the SM increases with the number of RHGs, whereas that of $R_{12}$ diminishes (Graph 5). The contribution of $R_{int}$ is also greater for the SM than for the LR, unlike the contribution of $R_{12}$, which is smaller for the SM than for the LR.

As was observed with the empirical study, the profile of the contribution of $R_{01}$ to the measure of change is the same as that of the measure itself. This shows that the variations in $D$ depend directly on $R_{01}$.

Graph 6 enables us to compare the LR and the SM, presenting the average contribution of $R_{01}$ to the measure of change for the methods with an essentially equivalent number of RHGs. As with the SLID, the results indicate that nonresponse seems to be better targeted with the SM than with the LR method.

Unlike the SLID simulation study, the bias was not evaluated since no external source of data was available for evaluation purposes.
7. Conclusion

This document highlights the fact that the choice of RHGs and method for defining them depends on the: i) availability of ancillary information, ii) need to reduce the nonresponse bias for all estimates, and iii) time and operational constraints. The empirical study, as well as the NLSCY data, showed that the SM method appears to be better than the LR one in reducing the nonresponse bias. The results also demonstrated that the proposed measure of change can be a very useful tool for comparing different weighting strategies.

In particular, it would appear that, as the value of $R_{ij}$ increases, the reduction of the bias obtained from using RHGs increases. Given the difficulty in obtaining a reliable estimate of the nonresponse bias in a survey, the relationship identified between the size of $R_{ij}$ and the decrease in the bias suggests that $R_{ij}$ should be used as a tool for evaluating nonresponse adjustment methods. This requires that $R_{ij}$ first be determined for different RHG sets. Then, the set with the highest $R_{ij}$ value is likely to be more effective than the other alternatives in reducing the nonresponse bias for most of the variables of interest.

The measure of change presented could also be used to compare the different calibration strategies. In this case, the nonresponse adjustment could remain the same for all of the poststratification methods under study. A detailed study of the behaviour of $R_{ij}$ could be done and would no doubt lead to certain conclusions, as this study did about $R_{ij}$. This type of study would not necessarily have to be restricted to the longitudinal context but could quite readily be done with a cross-sectional study. Also, the measure of change could be useful in evaluating different nonresponse adjustment methods in cross-sectional surveys.

Acknowledgments

The authors would like to thank M. Hladky, M. Latouche, C. Nadeau and N. Tremblay for their important contributions to this project.

References


