

LE BIAIS DE RENOUVELLEMENT DE L'ÉCHANTILLON DANS LES ESTIMATIONS DE L'EPA¹

P.D. GHANGURDE²

L'auteur essaie d'évaluer l'impact d'un ajustement pour la non-réponse, par groupe de renouvellement, sur le biais de renouvellement dans les estimations de l'Enquête sur la population active du Canada. Des résultats sur le biais et sur des caractéristiques des non-répondants sont présentés et discutés. Un indice utilisé pour mesurer le biais de renouvellement est donné et des résultats d'une étude empirique sont analysés.

1. INTRODUCTION

Dans le plan d'échantillonnage de l'Enquête sur la population active du Canada (EPA), un sixième des ménages sont supprimés de l'échantillon à chaque mois et la même proportion d'autres ménages viennent les remplacer. L'échantillon est ainsi composé de six panels ou groupes de renouvellement. À un mois donné, les ménages à l'intérieur d'un groupe de renouvellement ont participé à l'enquête depuis un nombre de mois variant d'un à six, y compris le mois en question. Il est bien connu que, dans les enquêtes-ménages fondées sur un plan de renouvellement des unités de l'échantillon, les estimations relatives aux mêmes caractéristiques, qui sont calculées pour différents groupes de renouvellement, peuvent ne pas avoir la même valeur prévue. Ce phénomène, qu'on appelle le biais de renouvellement, a été étudié relativement à l'EPA et à d'autres enquêtes-ménages basées sur un plan de renouvellement des unités de l'échantillon (voir [1], [5], [7] et [8]).

¹ Texte présenté à la réunion de l'American Statistical Association, à Cincinnati, en août 1982.

² P.D. Ghangurde, Division des méthodes de recensement et d'enquêtes-ménages, Statistique Canada.

Le biais de renouvellement peut être dû à plusieurs facteurs. Dans l'EPA, on a observé des variations du taux de non-réponse au niveau des ménages en fonction du groupe de renouvellement, c'est-à-dire selon le nombre de mois de participation des ménages à l'enquête. Il a aussi été souligné que les caractéristiques des ménages de non-réponse ont tendance à être différentes de celles des ménages répondants. Ces deux facteurs peuvent influencer sur le biais. Le facteur de conditionnement du répondant ou sa familiarisation avec l'enquête au cours d'une période de six mois peut faire varier l'importance du biais de réponse dans les données recueillies d'un mois à l'autre. Les données obtenues lors des réinterviews de l'EPA offrent quelques indices de ce genre de biais variable pendant les six mois de participation à l'enquête. Toutefois, selon une hypothèse qui a été formulée dans les études sur le biais, le biais de renouvellement pourrait être causé par les écarts entre les probabilités de non-réponse des groupes de renouvellement [7]. Même si les probabilités individuelles sont inconnues, il est possible d'estimer leur moyenne à partir des taux de non-réponse.

Dans la présente étude, on tente d'évaluer l'effet des différences entre les taux de non-réponse des groupes de renouvellement et entre certaines caractéristiques des répondants et des non-répondants sur le biais de renouvellement. La section 2 présente quelques résultats concernant le biais et décrit l'incidence de ces résultats sur le biais dans les estimations calculées pour différents groupes de renouvellement. La section 3 fournit quelques données sur les taux de non-réponse dans l'EPA, sur des caractéristiques des répondants et des non-répondants selon le nombre de mois d'inclusion dans l'échantillon et sur l'influence de ces variables sur le biais attribuable au renouvellement. La section 4 explique l'ajustement des poids de l'EPA appliqué aux groupes de renouvellement pour compenser la non-réponse et l'effet de cette correction sur le biais de renouvellement. On y décrit aussi l'indice qui sert de mesure de ce biais. La section 5 contient une analyse des valeurs de cet indice calculées, à partir des enquêtes menées en 1981, pour les diverses catégories d'activité.

2. MODÈLE STATISTIQUE

Cette section expose un modèle qui contient des expressions représentant le biais introduit par les différences entre les taux de non-réponse et par les différences entre les caractéristiques des répondants et des non-répondants, ainsi que le biais de réponse de tout groupe dans l'échantillon auquel il est possible d'appliquer un ajustement de la pondération pour compenser la non-réponse. Les groupes de renouvellement peuvent être considérés comme un cas particulier de ce type de groupe.

Supposons qu'une population de taille N est divisée en "strates" de répondants et de non-répondants de taille N_1 et N_2 respectivement. Un échantillon aléatoire simple de taille n est prélevé et des réponses sont fournies par n_1 unités, mais non par $(n-n_1)$ unités.

Supposons aussi que l'échantillon peut être divisé en K groupes, de sorte que les taux de non-réponse et les caractéristiques des répondants et des non-répondants varient d'un groupe à l'autre. Les méthodes de collecte des données utilisées pour ces groupes et le degré de conditionnement ou de familiarisation des répondants avec l'enquête peuvent être différents, entraînant ainsi des écarts entre les taux de non-réponse, entre les caractéristiques et peut-être aussi dans le biais de réponse. Par l'adaptation d'un résultat présenté dans [2] et [6] de façon à inclure une composante pour le biais de réponse, le biais de la moyenne \bar{y} de l'échantillon formé de n_1 unités (sans ajustement de la pondération pour compenser la non-réponse dans chaque groupe) correspond à la formule:

$$B(\bar{y}) = \frac{1}{\bar{R}} \sum_{i=1}^K P_i \bar{Y}_{1i} (R_i - \bar{R}) + \frac{1}{\bar{R}} \sum_{i=1}^K P_i (1-R_i) (\bar{Y}_{1i} - \bar{Y}_{2i}) + \frac{1}{\bar{R}} \sum_{i=1}^K P_i R_i \bar{\beta}_i, \quad (1)$$

où \bar{Y}_{1i} et \bar{Y}_{2i} sont les moyennes de la population de répondants et de non-répondants dans le i ème groupe, R_i correspond au taux de réponse du

i ème groupe, P_i est la proportion de l'ensemble de la population dans le i ème groupe, $\bar{\beta}_i$ est la moyenne du biais de réponse dans le i ème groupe et $\bar{R} = \sum_{i=1}^K P_i R_i$, le taux de réponse global.

L'équation (1) montre comment le biais se décompose en trois composantes. La première composante indique le biais produit par les différences entre les taux de réponse, la deuxième, le biais causé par les différences entre les caractéristiques des répondants et des non-répondants, tandis que la troisième correspond au biais de réponse. Pour simplifier l'analyse, la présente étude utilise des caractéristiques basées sur des attributs comme, par exemple, les proportions de personnes occupées et de chômeurs. L'estimation y_i , qui renferme un ajustement pour compenser la non-réponse appliqué à chaque groupe, est exprimée de la façon suivante:

$$\bar{y}_a = \frac{1}{n} \sum_{i=1}^K n_{.i} \bar{y}_i,$$

où $n_{.i}$ est la taille de l'échantillon dans le i ème groupe et \bar{y}_i est la moyenne des n_{1i} unités dans le i ème groupe. L'équation suivante donne le biais de \bar{y}_a :

$$B(\bar{y}_a) = \sum_{i=1}^K P_i (1-R_i) (\bar{Y}_{1i} - \bar{Y}_{2i}) + \sum_{i=1}^K P_i \bar{\beta}_i. \quad (2)$$

La première composante de la formule (1), le biais dû aux variations entre les taux de réponse d'un groupe à l'autre, est éliminée, la deuxième composante, le biais produit par la différences entre les caractéristiques des répondants et des non-répondants, reste identique, tandis que la troisième composante, le biais de réponse, peut ne pas avoir la même valeur que dans l'équation (1).

Au moyen d'un modèle qui distingue les erreurs de réponse et de non-réponse et qui utilise les probabilités de réponse au niveau des unités, on a décomposé le biais en une composante attribuable à la non-réponse et une autre causée par la réponse [3]. La décomposition du biais présentée plus haut n'inclut pas les probabilités de réponse des unités individuelles, mais elle est assez simple pour permettre une évaluation empirique des composantes.

Si les taux de réponse ne varient pas d'un groupe à l'autre, la première composante de l'équation (1) vaut zéro, de sorte que cette équation est identique à (2). Ainsi, un ajustement pour compenser la non-réponse n'entraîne pas une diminution du biais. La différence entre le biais de \bar{y} et celui de \bar{y}_a est calculée par l'équation (3).

$$B(\bar{y}) - B(\bar{y}_a) = \frac{1}{\bar{R}} \sum_{i=1}^K P_i (\bar{R}_i - \bar{R}) (\bar{Y}_{1i} + \bar{\beta}_i). \quad (3)$$

Donc, si les taux de réponse sont différents et si \bar{Y}_{1i} et $\bar{\beta}_i$ ne varient pas en fonction du groupe, il n'y a pas de changement dans le biais après l'application de l'ajustement pour compenser la non-réponse à l'intérieur des groupes. Si les moyennes \bar{Y}_{1i} et $\bar{\beta}_i$ varient d'un groupe à l'autre, le biais diminue si le membre de droite dans l'équation (3) est positif, mais il augmente si ce membre est négatif. Le changement du biais absolu de $|B(\bar{y})|$ à $|B(\bar{y}_a)|$ par suite de la correction de la non-réponse dépend donc du signe et de la grandeur du membre de droite de l'équation (3).

On peut obtenir le biais de l'estimation de la moyenne pour le $i^{\text{ème}}$ groupe de renouvellement, avec et sans ajustement de la pondération pour compenser la non-réponse dans chaque groupe, à l'aide des équations (1) et (2) simplement en fixant $P_i = 1$ et en conservant les termes qui correspondent aux divers groupes de renouvellement. De plus, l'équation (3) permet de définir par la formule (4) la différence entre le biais des deux estimations calculées pour le $i^{\text{ème}}$ groupe de renouvellement:

$$B(\bar{y}_i) - B(\bar{y}_{ia}) = \frac{(\bar{R}_i - \bar{R})}{\bar{R}} (\bar{Y}_{1i} + \bar{\beta}_i), \quad (4)$$

où \bar{Y}_i et \bar{Y}_{ia} sont respectivement les estimations obtenues pour le $i^{\text{ème}}$ groupe de renouvellement avant et après l'ajustement. Supposons que $(\bar{Y}_{1i} + \bar{\beta}_i) > 0$ pour tous les groupes i ; si $\bar{R}_i < \bar{R}$, le biais associé au $i^{\text{ème}}$ groupe de renouvellement augmente après l'ajustement et si $\bar{R}_i > \bar{R}$, le biais diminue.

Comme la population de répondants dans divers groupes de renouvellement est identique pendant un mois d'enquête donné, on peut prétendre qu'il est

possible que les moyennes \bar{Y}_{1i} soient identiques parmi tous les groupes de renouvellement ou pour chaque nombre de mois d'inclusion dans l'échantillon. Toutefois, les différents degrés de familiarisation avec l'enquête ou de conditionnement chez les répondants peuvent produire des variations dans le biais de réponse, $\bar{\beta}_i$, d'un groupe de renouvellement à l'autre. Ainsi, la différence entre le biais de \bar{y} et de \bar{y}_a est le résultat de l'équation suivante.

$$B(\bar{y}) - B(\bar{y}_a) = \frac{1}{\bar{R}} \sum_{i=1}^K P_i (R_i - \bar{R}) \bar{\beta}_i. \quad (5)$$

Cependant, la différence entre le biais des estimations calculées pour le groupe de renouvellement i se trouve à l'aide de l'équation (4).

On peut aussi signaler que, si on suppose que \bar{Y}_{1i} et $\bar{\beta}_i$ sont constants pour tous les groupes i et que les taux de réponse varient d'un groupe à l'autre, la correction de la non-réponse dans les groupes de renouvellement ne produit aucun changement dans le biais des estimations calculées pour l'ensemble des groupes. Toutefois, le changement introduit dans le biais des estimations calculées pour un groupe de renouvellement donné s'explique par les différences entre les taux de réponse de chaque groupe.

Les résultats présentés ci-dessus sont utiles pour évaluer l'importance de divers facteurs dans le biais de renouvellement et pour étudier l'effet de la correction du poids appliquée à ce biais dans chaque groupe de renouvellement.

L'EPA est une enquête mensuelle menée à l'échelle nationale auprès d'un échantillon de 55,000 ménages. Chacune des dix provinces du Canada est divisée en régions économiques, c'est-à-dire en groupes de comtés qui possèdent une structure économique semblable. Ces régions sont décomposées en strates homogènes selon la répartition des personnes occupées dans diverses catégories de professions et d'activités économiques établie d'après les résultats du dernier recensement. On utilise un plan d'échantillonnage stratifié à plusieurs degrés qui prévoit deux degrés d'échantillonnage dans les unités

autoreprésentatives (UAR) urbaines et trois ou quatre degrés dans les unités non autoreprésentatives (UNAR) rurales. Aux premières étapes de l'échantillonnage, on applique une méthode de sélection avec probabilité proportionnelle à la taille de la population et, à l'étape finale, où les logements sont sélectionnés parmi les grappes, l'échantillonnage est systématique. À l'intérieur de chaque strate, six numéros de renouvellement sont attribués indépendamment aux grappes sélectionnées. Tous les mois, chaque sixième des ménages font partie de l'enquête depuis 1 à 6 mois. Ainsi, l'échantillon global est composé de six sous-échantillons de même taille qui sont aussi représentatifs les uns que les autres [4]. Il est possible de convertir le numéro de renouvellement des six groupes de renouvellement en un "nombre de mois d'inclusion dans l'échantillon" par une transformation simple.

L'ajustement du poids pour compenser la non-réponse est appliqué à la partie de l'échantillon contenue dans une unité de compensation en fonction du rapport entre le nombre de ménages dans cette unité et le nombre de ménages qui ont répondu à l'enquête. Dans les secteurs NAR, chaque unité primaire d'échantillonnage (UPÉ) est divisée en deux unités de compensation qui représentent une partie urbaine et une partie rurale. Dans les secteurs AR duplan d'enquête, les strates (désignées sous-unités) forment des unités de compensation. Le nombre d'unités de compensation dépasse donc 900 dans les UNAR et 800 dans les UAR.

Afin d'évaluer, avec et sans ajustement, le biais dû au renouvellement dans les estimations de l'EPA, on présente et on analyse à la section 3 les données recueillies, au cours des douze mois d'enquête de 1981, sur les taux de non-réponse ($1-\bar{R}_i$) et sur \bar{Y}_{1i} et \bar{Y}_{2i} , les proportions de répondants et de non-répondants classés comme "personnes occupées" et comme "chômeurs". Le "nombre de mois d'inclusion dans l'échantillon" correspond au nombre de mois (y compris le mois en cours) de participation d'un groupe de renouvellement à l'enquête. Aucune donnée sur le biais de réponse, $\bar{\beta}_i$, n'est présentée.

3. ANALYSE DES DONNÉES DE L'EPA

Le tableau 1 indique les taux de non-réponse moyens, $(1-\bar{R}_i)$, selon le nombre de mois d'inclusion dans l'échantillon pour les douze mois civils de 1981. On

peut constater que ces taux varient beaucoup selon le type de région et, pour un type de région donné, selon le nombre de mois d'inclusion dans l'échantillon. Dans les UNAR, les UAR et au niveau du Canada, les taux de non-réponse sont élevés au premier mois, accusent une baisse considérable au deuxième mois et diminuent progressivement au cours des mois suivants. L'ampleur des taux de non-réponse au premier mois est due aux ménages "temporairement absents" et aux cas de "personne à la maison". Par la suite, ces taux diminuent au fur et à mesure que les interviewers déterminent quel est le meilleur moment de communiquer avec ces types de ménages. Les taux de non-réponse sont plus élevés dans les UAR, surtout dans les appartements (qui ne figurent pas au tableau 1), que dans les UNAR. À l'étape du traitement des données, les renseignements fournis par environ $\frac{1}{2}$ % des ménages au cours du mois précédent sont reportés à mois en cours. On obtient les taux de non-réponse présentés dans les tableaux en considérant ces ménages comme des répondants. Le tableau 1 permet également de voir que l'écart entre les taux de non-réponse et leur moyenne, $(R_i - \bar{R})$, est négatif au premier et, parfois, au deuxième mois d'inclusion dans l'échantillon, et positif dans les mois qui suivent. Le taux moyen \bar{R} est approximativement égal à \bar{R}_2 . Donc, d'après l'équation (4), le biais relatif au premier mois d'inclusion dans l'échantillon devrait augmenter si $(\bar{Y}_{1i} + \bar{\beta}_i)$ et la moyenne de la population \bar{Y}_i sont supposé invariables. Du troisième au sixième mois, le biais relatif devrait diminuer après la correction du poids pour compenser la non-réponse.

Au tableau 2 figurent les proportions estimées, $\hat{\bar{Y}}_{1i}$ et $\hat{\bar{Y}}_{2i}$, des chefs de ménage qui sont occupés ou en chômage, selon le nombre de mois d'inclusion dans l'échantillon, dans le cas des ménages répondants et non répondants. Les estimations ont été calculées à partir des fichiers longitudinaux de l'EPA pour la période de mars à août 1976 et elles sont basées sur des chiffres non pondérés. Les données sur les non-répondants, qui ont participé à l'enquête au moins une fois au cours de la période de six mois, proviennent des mois où ils ont effectivement répondu. Les ménages de non-réponse ont tendance à avoir une plus forte proportion de chefs occupés et une plus faible proportion de chefs en chômage que les ménages répondants. Il est connu que la différence entre les proportions de répondants et de non-répondants classés comme personnes occupées a tendance à être 0.10 et, dans le cas des chômeurs, environ 0.005, le signe de ces différences demeurant identique. D'un nombre

de mois à l'autre, on ne distingue aucune tendance particulière dans l'évolution des proportions de chefs occupés ou en chômage parmi les ménages répondants et non répondants.

L'incidence du premier mois dans la première composante est négative pour tous les mois civils, tant dans le cas des personnes occupées que dans le cas des chômeurs, ce qui indique que le biais du premier mois d'inclusion dans l'échantillon devrait augmenter après l'application de l'ajustement pour compenser la non-réponse.

L'analyse présentée aux sections 2 et 3 désigne les groupes de renouvellement comme ceux auxquels serait appliqué l'ajustement de la non-réponse. Il se peut qu'avec des données réelles on ne puisse pas observer des changements relatifs, à cause de l'effet produit par des différences entre les taux de non-réponse des autres groupes et parce que les valeurs de \bar{Y}_{1i} et de $\bar{\beta}_i$ peuvent varier pendant la période de six mois. À la section 5, nous analysons l'effet de l'ajustement pour compenser la non-réponse appliqué aux groupes de renouvellement sur le biais de renouvellement, et on tente d'expliquer les résultats à l'aide du modèle décrit plus haut.

Il convient de souligner que l'ajustement de la non-réponse dans le système de pondération actuel de l'EPA se fait au niveau d'unités de compensation qui sont beaucoup plus petites que les UNAR et les UAR délimitées à l'intérieur d'une province. Les estimations du biais de renouvellement qui sont basées sur le système actuel de pondération et de correction de la non-réponse renferment donc un ajustement des différences entre les taux de non-réponse des deux types de région, mais non de celles entre les taux des groupes de renouvellement.

4. CORRECTION DU POIDS AU NIVEAU DES GROUPES DE RENOUVELLEMENT

Le poids final de l'EPA est composé de cinq facteurs: 1) un facteur de pondération mathématique, 2) un facteur rural-urbain, 3) un facteur de sous-pondération de grappe, 4) un facteur de compensation et 5) un facteur âge-sexe. Le facteur de pondération mathématique appliqué à un ménage est égal à l'inverse de la fraction de sondage déterminée par le plan de sondage. Dans chaque province, le poids attribué aux strates urbaines (AR) et rurales (NAR)

est identique, sauf pour quelques exceptions; il y a ainsi vingt régions à l'échelle du Canada qui ont le même facteur de pondération mathématique. Le facteur de sous-pondération de grappe correspond à la fraction de sondage inverse de chaque grappe. Le facteur de compensation corrige le poids pour tenir compte de la non-réponse, et le facteur âge-sexe est une estimation par quotient calculée à partir de projections du nombre total de personnes classées dans les diverses catégories d'âge et de sexe dans chacune des provinces.

Comme il a été expliqué à la section 2, la correction du poids pour tenir compte de la non-réponse est effectuée, pour l'ensemble des ménages dans l'échantillon, à l'intérieur d'unités de compensation. Afin d'évaluer l'effet de l'ajustement du poids par groupe de renouvellement, on a décidé d'utiliser des régions de plus en plus petites (comme unités de compensation), en commençant par les groupes de renouvellement au niveau des provinces. Pour ajuster le poids final dans chacun des groupes de renouvellement dans ces régions, on a multiplié le poids par des facteurs d'ajustement:

$$R_H(i) = \frac{\text{ménages répondants dans l'échantillon}}{\text{ménages répondants dans le groupe de renouvellement (i)}}$$

$$R_p(i) = \frac{\text{personnes répondantes dans l'échantillon}}{\text{personnes répondantes dans le groupe de renouvellement (i)}}$$

Le premier de ces facteurs pondère l'estimation relative aux ménages d'un groupe de renouvellement jusqu'au niveau de l'échantillon des ménages répondants. Le facteur de compensation pondère l'estimation jusqu'au niveau de l'échantillon des ménages dans l'unité de compensation. Le deuxième facteur, calculé à partir du nombre de répondants individuels, pondère l'estimation jusqu'au niveau de l'ensemble de l'échantillon des personnes qui ont répondu à l'enquête, compensant ainsi les différences entre la taille des ménages ou les différences entre le nombre de personnes prises en compte à l'intérieur des ménages. Il est connu que la taille des ménages non répondants a tendance à être plus petite que celle des ménages répondants. Les variations des taux de non-réponse d'un groupe de renouvellement à l'autre peuvent provenir des différences dans la taille moyenne des ménages.

Si $\hat{Y}(i)$ est un total estimé pour le $i^{\text{ième}}$ groupe de renouvellement et si $Y(i)$ représente la vraie valeur du total du $i^{\text{ième}}$ groupe, l'estimation du biais relatif du total estimé pour le $i^{\text{ième}}$ groupe de renouvellement peut être calculée par l'équation suivante:

$$B_y(i) = \frac{\hat{Y}(i) - Y(i)}{Y(i)} ; i = 1, 2, \dots, 6. \quad (6)$$

Étant donné que les $Y(i)$ sont inconnus et qu'on peut supposer qu'ils sont approximativement égaux (parce que la valeur prévue de la taille des groupes de renouvellement est égale d'un groupe à l'autre dans le cas de grandes régions), $\hat{Y}(\cdot)$, la moyenne des estimations d'un total calculées pour les six groupes de renouvellement, peut remplacer $Y(i)$.

L'indice du biais de renouvellement pour le $i^{\text{ième}}$ groupe de renouvellement est défini par l'équation suivante:

$$I_y(i) = \frac{\hat{Y}(i)}{\hat{Y}(\cdot)} \cdot 100 = 1 + \beta_y(i) \cdot 100 \quad (7)$$

On peut souligner que, puisque la moyenne des estimations d'un total calculées pour les six groupes de renouvellement est utilisée à la place des vraies valeurs, $I_y(i)$ peut être biaisé, mais il est utile comme mesure d'évaluation de la différence entre le biais relatif d'un groupe de renouvellement à l'autre pour divers sous-groupes de la population, de même que pour ajuster le poids en fonction des chiffres concernant les ménages et les personnes. De même, il est possible de définir $P_y(i)$, le biais de renouvellement dans les estimations calculées pour la population, pour chacun des groupes de renouvellement.

Lorsque la valeur de l'indice $I_y(i)$ est supérieure à 100, le biais relatif est positif, et quand cette valeur est inférieure à 100, le biais relatif est négatif. On peut interpréter l'indice $I_p(i)$ de la même manière.

5. ANALYSE DES DONNÉES SUR L'INDICE DU BIAIS DE RENOUVELLEMENT

Cette section présente et analyse des tableaux où figurent des données sur l'indice du biais de renouvellement pour la population selon les catégories d'activité sur le marché du travail et en fonction du type de région et des groupes d'âge et de sexe. On obtient les valeurs présentées dans les tableaux en utilisant les poids finals et en effectuant la même correction de la non-réponse à partir des chiffres sur les ménages et des chiffres sur les personnes. On compare les valeurs des indices calculées avec et sans l'ajustement des poids dans le but d'évaluer l'effet de cette correction sur les estimations du biais de renouvellement. L'ajustement du poids des groupes de renouvellement à partir des chiffres sur les ménages a été fait au niveau de chaque province. Les poids finals des ménages dans les six groupes de renouvellement ont donc été multipliés par les facteurs d'ajustement $R_H(i)$; $i = 1, 2, \dots, 6$. De même, l'ajustement en fonction des chiffres sur les personnes a été effectué pour chaque province au moyen des facteurs $R_P(i)$; $i = 1, 2, \dots, 6$. Pour mesurer l'incidence de ces ajustements sur les estimations calculées pour la population, nous présentons le tableau 3, qui montre l'indice du biais de renouvellement, pour les estimations relatives à la population, selon le type de région et le nombre de mois d'inclusion dans l'échantillon, pour les douze enquêtes menées en 1981. Les valeurs de l'indice calculées sans l'ajustement du poids indiquent un sous-dénombrement relatif des personnes, au premier et au sixième mois, dans les régions AR et NAR. Les valeurs de l'indice obtenues pour les estimations calculées avec l'ajustement des chiffres sur les ménages montrent une amélioration du dénombrement. Toutefois, cet ajustement repose sur l'hypothèse de l'égalité de la taille des ménages d'un groupe de renouvellement à l'autre. Les valeurs de l'indice basées sur l'ajustement par les chiffres sur les répondants se rapprochent plus de 100.0 dans les deux types de région que les valeurs produites par l'ajustement au niveau des ménages. L'ajustement fait à partir des chiffres sur les personnes semble donc mieux corriger les différences dans le dénombrement qui agissent sur les estimations que l'ajustement fondé sur les chiffres concernant les ménages. Le fait que les valeurs de l'indice sont élevées dans les premiers mois et relativement faibles dans les derniers mois pourrait être dû à des variations de la taille des ménages de non-réponse en fonction du nombre de mois d'inclusion dans l'échantillon.

Les tableaux 4 et 5 présentent des données sur les valeurs moyennes de l'indice selon le type de région et le groupe d'âge et sexe, pour les douze enquêtes menées en 1981. Les valeurs obtenues par type de région sans ajustement du poids indiquent que le biais relatif des estimations concernant les chômeurs a tendance à être positif dans les deux premiers mois et à diminuer dans les derniers mois. Le biais relatif des estimations concernant les personnes occupées et les personnes actives a tendance à être négatif au premier mois et positif dans les mois subséquents. Les données sur les valeurs de l'indice calculées pour les groupes d'âge et les deux sexes affichent des tendances semblables à celles observées pour les types de région.

L'ajustement du poids pour compenser la non-réponse à l'aide des chiffres sur les ménages a tendance à augmenter les valeurs de l'indice au premier ainsi qu'au cinquième et au sixième mois. Les valeurs de l'indice obtenues pour les autres mois ont tendance à diminuer. Cette observation s'applique aux valeurs de l'indice calculées pour les catégories d'activité en fonction du type de région et du groupe d'âge et du sexe. On peut attribuer l'augmentation des valeurs de l'indice, au premier mois, à des taux de réponse en-dessous de la moyenne et la diminution des valeurs dans les mois subséquents, à des taux de réponse au-dessus de la moyenne. La baisse observée aux deux derniers mois ne peut pas être imputable à des taux de réponse au-dessus de la moyenne, si on suppose que $(\bar{Y}_{1i} + \bar{\beta}_i)$ est constant.

L'ajustement du poids pour compenser la non-réponse en fonction des chiffres sur les personnes a tendance à augmenter les valeurs de l'indice au premier mois et à les diminuer, du troisième au sixième mois. Les valeurs ainsi calculées pour le premier mois sont généralement plus élevées que celles obtenues à partir de l'ajustement au niveau des ménages. L'ajustement fait avec les chiffres sur les personnes semble corriger les effets que les différences dans les taux de réponse et le nombre de personnes dénombrées dans les groupes de renouvellement ont sur les estimations. Les taux de réponse et le nombre de personnes prises en compte sont faibles au premier mois, ce qui produit une augmentation du biais relatif après l'ajustement. Il semble que la diminution du biais relatif, du troisième au sixième mois, soit due à des taux de réponse

en-dessous de la moyenne, au niveau des ménages, qui sont corrigés pour compenser les variations du nombre de personnes dénombrées d'un groupe de renouvellement à l'autre.

6. RÉSUMÉ ET CONCLUSION

La présente étude examine un modèle qui divise l'ensemble du biais en trois composantes, nommément l'effet des différences observées d'un groupe à l'autre de l'échantillon entre les taux de réponse, entre le biais de réponse et entre les caractéristiques des répondants et des non-répondants. Ce modèle indique aussi l'incidence de divers autres facteurs sur le biais de renouvellement. Il est possible de considérer les groupes de renouvellement comme un cas particulier de ce genre de groupes où l'ajustement du poids pour compenser la non-réponse peut être fait séparément.

Si les taux de réponse varient d'un groupe de renouvellement à l'autre et si la proportion de répondants et de non-répondants qui possèdent une caractéristique et le biais de réponse correspondant sont identiques pour tous les groupes de renouvellement, la correction de la non-réponse appliquée à chaque groupe de renouvellement ne change pas le biais des estimations. Toutefois, le biais de renouvellement peut augmenter ou diminuer, si le taux de réponse est respectivement en dessous ou au-dessus du taux de réponse moyen. Cette conclusion est confirmée par les données sur les valeurs de l'indice obtenues avant et après l'ajustement du poids en fonction des chiffres sur les personnes.

On entend faire des analyses des valeurs de l'indice pour des estimations relatives à l'activité sur le marché du travail et d'autres caractéristiques calculées à partir d'ensembles de données plus grands. On étudiera l'effet des différences entre la taille moyenne des ménages dans les groupes de renouvellement et parmi les ménages répondants et non répondants sur les estimations du biais de renouvellement. L'analyse portera également sur le rôle des différences entre les taux de réponse et des variations du biais de réponse, après la correction de la non-réponse, dans chaque groupe de renouvellement.

REMERCIEMENTS

L'auteur tient à remercier R. Vettore et R. Barnes pour l'élaboration de programmes informatiques, de même que les arbitres pour leurs remarques pertinentes.

BIBLIOGRAPHIE

- [1] Bailar, B.A., (1975), The Effect of Rotation Group Bias on Estimates from Panel Surveys, JASA vol. 70, pp. 23-30.
- [2] Bailar, B.A., Bailey, L. et Corby, C., (1978), A Comparison of Some Adjustment and Weighting Procedures for Survey Data. Survey Sampling and Measurement, Academic Press, pp.175-198.
- [3] Platek, R., Singh, M.P. et Tremblay, V., (1978), Adjustment for Non-Response in Surveys, Techniques d'enquête, vol. 3. n° 1, pp. 1-24.
- [4] Statistique Canada (1976), Méthodologie de l'enquête sur la population active du Canada, n° 71-526 au catalogue, hors série.
- [5] Tessier, R. et Tremblay, V., (1976), Findings on Rotation Group Biases, document interne, Statistique Canada.
- [6] Thomsen, I., (1973), A Note on the efficiency of Sub-class Means to Reduce the Effects of Non-response when Analyzing Survey Data, Statistical Review, publié par le National Central Bureau of Statistics, Stockholm, Suède, vol. 11, n° 4.
- [7] Williams, W.H. et Mallows, C.L., (1970), Systematic Biases in Panel Surveys Due to Differential Non-Response, JASA, vol. 65, n° 331.
- [8] Woltman, H. et Bushery, J., (1975), A Panel Bias Study in the National Crime Survey, document présenté à la réunion annuelle de l'ASA.

TABLEAU 1. % Taux de non-réponse des ménages, en pourcentage, selon le nombre de mois d'inclusion dans l'échantillon et le type de région (1981)

Nombre de mois	Type de région		
	UNAR	AR	Canada†
1	6.6	7.9	7.3
2	4.0	4.6	4.4
3	3.5	4.4	3.9
4	3.5	4.1	3.8
5	3.2	3.8	3.6
6	3.1	3.6	3.4
Nombre moyen de ménages	26,707	28,645	55,352

† à l'exclusion des secteurs spéciaux

TABLEAU 2. Proportions estimées des chefs de ménage occupés et en chômage chez les ménages répondants et non répondants

Nombre de mois	Répondants \hat{Y}_{1i}		Non-Répondants \hat{Y}_{2i}		$\hat{Y}_{1i} - \hat{Y}_{2i}$	
	Occupés	Chômeurs	Occupés	Chômeurs	Occupés	Chômeurs
1	0.6893	0.0383	0.7839	0.0335	-0.0946	0.0048
2	0.6962	0.0344	0.7841	0.0321	-0.0879	0.0023
3	0.7006	0.0311	0.7851	0.0300	-0.0845	0.0011
4	0.7006	0.0364	0.7877	0.0281	-0.0871	0.0083
5	0.6972	0.0317	0.7821	0.0317	-0.0849	0.0000
6	0.6927	0.0331	0.7767	0.0320	-0.0840	0.0011
Moyenne	0.6961	0.0342	0.7833	0.0311	-0.0872	0.0031

TABLEAU 3. Indice du biais de renouvellement pour la population par type de région

Poids	Type de région	Nombre de mois d'inclusion dans l'échantillon					
		1	2	3	4	5	6
Non ajusté	AR	97.0	101.1	101.2	100.6	100.2	99.7
	UNAR	97.7	101.0	100.8	100.9	100.2	99.4
Ajusté pour les ménages	AR	98.7	98.7	99.4	100.0	101.1	102.1
	UNAR	99.3	98.6	99.0	100.3	101.1	101.8
Ajusté pour la population	AR	100.4	100.5	100.2	99.7	99.6	99.5
	UNAR	100.9	100.3	99.8	99.9	99.6	99.4

TABLEAU 4. Indice du biais de renouvellement par type de région (1981)

Poids	Caractéristique	Type de région	Nombre de mois d'inclusion dans l'échantillon					
			1	2	3	4	5	6
Non ajusté	Personnes occupées	AR	99.9	101.0	101.3	100.7	100.4	99.8
		UNAR	96.8	100.9	100.6	101.2	100.7	99.9
	Chômeurs	AR	99.1	102.6	101.3	100.4	97.7	98.9
		UNAR	103.3	101.5	101.4	99.8	96.5	97.6
	Personnes actives	AR	97.0	101.1	101.3	100.7	100.2	99.7
		UNAR	97.3	100.9	100.7	101.1	100.3	99.7
Ajusté pour les ménages	Personnes occupées	AR	98.6	98.5	99.5	100.1	101.2	102.1
		UNAR	98.3	98.4	98.7	100.6	101.6	102.4
	Chômeurs	AR	100.8	100.3	99.5	99.8	98.5	101.1
		UNAR	104.9	99.2	99.6	99.2	97.3	99.8
	Personnes actives	AR	98.7	98.7	99.5	100.1	101.0	102.0
		UNAR	98.9	98.4	98.8	100.5	101.2	102.1
Ajusté pour la population	Personnes occupées	AR	100.2	100.3	100.4	99.7	99.7	99.6
		UNAR	100.0	100.2	99.6	100.2	100.1	99.9
	Chômeurs	AR	102.4	102.1	100.4	99.4	97.1	98.6
		UNAR	106.4	100.8	100.5	98.9	96.0	97.4
	Personnes actives	AR	100.4	100.5	100.4	99.7	99.5	99.5
		UNAR	100.6	100.2	99.7	100.1	99.8	99.6

TABEAU 5. Indice du biais de renouvellement par groupe d'âge et sexe (1981)

Poids	Caractéristique	Groupe âge-sexe	Nombre de mois d'inclusion dans l'échantillon					
			1	2	3	4	5	6
Non ajusté	Personnes occupées	M 15-24	96.5	99.7	100.7	101.0	101.1	101.1
		F 15-24	96.0	99.7	101.1	100.9	101.2	101.1
		M 25+	97.0	101.4	101.3	100.7	100.1	99.5
		F 25+	96.9	101.2	101.2	101.0	100.5	99.3
	Chômeurs	M 15-24	100.9	102.3	101.1	100.7	96.9	98.2
		F 15-24	102.4	102.7	97.7	98.9	100.0	98.2
		M 25+	98.0	102.1	101.6	100.3	98.0	100.1
		F 25+	100.1	102.3	104.5	100.4	95.3	97.4
	Personnes actives	M 15-24	97.2	100.1	100.8	100.9	100.4	100.6
		F 15-24	96.8	100.0	100.7	100.7	101.0	100.8
		M 25+	97.1	101.4	101.3	100.7	100.0	99.5
		F 25+	97.1	101.3	101.4	101.0	100.1	99.1
Ajusté pour les ménages	Personnes occupées	M 15-24	98.2	97.2	98.8	100.3	101.9	103.5
		F 15-24	97.6	97.2	99.3	100.3	102.1	103.5
		M 25+	98.7	98.9	99.4	100.1	101.0	101.8
		F 25+	98.6	98.8	99.3	100.3	101.3	101.6
	Chômeurs	M 15-24	102.6	100.0	99.3	100.1	97.7	100.3
		F 15-24	104.2	100.3	96.0	98.3	100.8	100.4
		M 25+	99.6	99.8	99.8	99.7	98.8	102.3
		F 25+	101.8	100.0	102.6	99.8	96.1	99.6
	Personnes actives	M 15-24	98.9	97.6	98.9	100.3	101.3	103.0
		F 15-24	98.5	97.6	98.8	100.0	101.9	103.2
		M 25+	98.8	99.0	99.5	100.1	100.8	101.0
		F 25+	98.8	98.8	99.6	100.3	101.0	101.5
Ajusté pour la population	Personnes occupées	M 15-24	99.9	99.0	99.7	100.0	100.5	101.0
		F 15-24	99.3	99.1	100.2	100.0	100.6	101.0
		M 25+	100.4	100.7	100.3	99.7	99.5	99.3
		F 25+	100.3	100.6	100.2	100.0	99.9	99.1
	Chômeurs	M 15-24	104.2	101.7	100.1	99.7	96.3	98.0
		F 15-24	105.8	102.0	96.8	98.0	99.4	98.0
		M 25+	101.2	101.6	100.7	99.4	97.4	99.8
		F 25+	103.4	101.7	103.5	99.5	94.8	97.1
	Personnes actives	M 15-24	100.5	99.4	99.8	99.9	99.8	100.5
		F 15-24	100.2	99.4	99.7	99.7	100.4	100.6
		M 25+	100.4	100.8	100.3	99.7	99.4	99.3
		F 25+	100.5	100.6	100.4	99.9	99.5	98.9