



N° 62F0026MIF au catalogue — N° 001
ISSN: 1708-8887
ISBN: 0-662-75349-6

Document de recherche

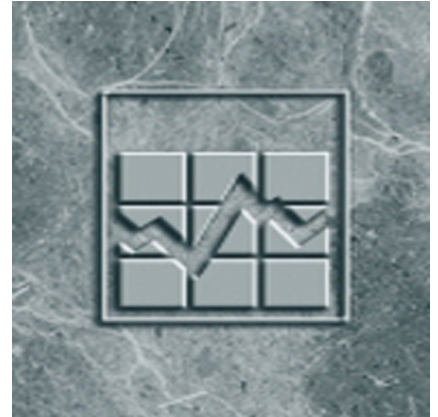
Série de documents de recherche sur les dépenses des ménages

Enquête sur les dépenses des ménages de 2001 - Indicateurs de qualité des données

2001

Division des méthodes d'enquêtes auprès des ménages
Immeuble R.H. Coats, Ottawa, K1A 0T6

Telephone: 613 951-7355



Statistique
Canada

Statistics
Canada

Canada

Comment obtenir d'autres renseignements

Veillez communiquer avec les Services à la clientèle, Division de la statistique du revenu, Statistique Canada, Ottawa, Ontario, K1A 0T6 ((613) 951-7355; (888) 297-7355: revenu@statcan.ca).

Pour obtenir des renseignements sur l'ensemble des données de Statistique Canada qui sont disponibles, veuillez composer l'un des numéros sans frais suivants. Vous pouvez également communiquer avec nous par courriel ou visiter notre site Web.

Service national de renseignements	1 800 263-1136
Service national d'appareils de télécommunications pour les malentendants	1 800 363-7629
Renseignements concernant le Programme des bibliothèques de dépôt	1 800 700-1033
Télécopieur pour le Programme des bibliothèques de dépôt	1 800 889-9734
Renseignements par courriel	infostats@statcan.ca
Site Web	www.statcan.ca

Renseignements sur les commandes et les abonnements

Le produit n° 62F0026MIF2003001 au catalogue est gratuit sur Internet. Les utilisateurs sont priés de se rendre à http://www.statcan.ca/cgi-bin/downpub/research_f.cgi.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois, et ce, dans la langue officielle de leur choix. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1 800 263-1136.



Statistique Canada
Division de la statistique du revenu

Série de documents de recherche sur les dépenses des ménages

Enquête sur les dépenses des ménages de 2001 - Indicateurs de qualité des données

2001

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2003

Tous droits réservés. Il est interdit de reproduire ou de transmettre le contenu de la présente publication, sous quelque forme ou par quelque moyen que ce soit, enregistrement sur support magnétique, reproduction électronique, mécanique, photographique, ou autre, ou de l'emmagasiner dans un système de recouvrement, sans l'autorisation écrite préalable des Services de concession des droits de licence, Division du marketing, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Novembre 2003

N° 62F0026MIF2003001 au catalogue

ISSN: 1708-8887

ISBN: 0-662-75349-6

Périodicité : Irrégulier

Ottawa

This publication is available in English upon request (Catalogue no. 62F0026MIE)

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Ce document a été préparé par les membres de l'unité de l'Enquête sur les dépenses des ménages de la Division des méthodes d'enquêtes auprès des ménages (DMEM) :

Sylvie DeBlois
José Gaudet
Bruno Lapierre
Jessica Law
Christian Nadeau
Sylvain Nadon

Les auteurs remercient l'équipe responsable de l'Enquête sur les dépenses des ménages de la Division de la statistique du revenu (DSR) pour leur collaboration dans la préparation de ce document.

Table des matières

Faits saillants	5
Introduction	7
1. Les erreurs d'échantillonnage.....	8
1.1 Les mesures de l'erreur d'échantillonnage	8
1.2 Les coefficients de variation.....	8
Tableau 1.1	
Coefficients de variation (%) par province et territoire ainsi qu'à l'échelle nationale pour l'estimation de la moyenne des dépenses par ménage pour quelques catégories de dépenses sommaires et pour l'estimation du revenu moyen.	9
Tableau 1.2	
Coefficients de variation (%) par province et territoire ainsi qu'à l'échelle nationale pour quelques catégories des caractéristiques des logements et l'équipement ménager	10
1.3 Modèle pour dériver une approximation du CV	10
1.4 Suppression des données non fiables dans les tableaux d'estimations.....	12
2. La non-réponse.....	12
2.1 Les taux de non-réponse et les taux de vacance	13
Tableau 2.1	
Taux de non-réponse (%) et taux de vacance (%) par province et territoire ainsi qu'à l'échelle nationale	14
2.2 La non-réponse selon le niveau d'urbanisation	15
Tableau 2.2	
Taux de non-réponse (%) et taux de vacance (%) par niveau d'urbanisation.....	15
2.3 La non-réponse selon les strates de revenu.....	16
Tableau 2.3	
Comparaison des taux (%) de non-réponse et de vacance des strates à revenu élevé et à faible revenu par rapport aux autres strates	16
2.4 L'ajustement pour la non-réponse.....	16
3. Erreurs de couverture.....	17
3.1 Le sous-dénombrement et le surdénombrement : les taux de glissement.....	18
Tableau 3.1	
Taux de glissement nationaux selon le groupe d'âge-sexe.....	19
Tableau 3.2	
Taux de glissement pour les provinces et les territoires selon le groupe d'âge-sexe	20
Tableau 3.3	
Taux de glissement pour les provinces et les territoires selon la taille de ménage.....	21
3.2 L'ajustement pour l'erreur de couverture de la population et du nombre de ménages ..	21
4. Les erreurs de réponse	23

Table des matières (fin)

5. Les erreurs de traitement	24
5.1 Proportion de ménages ou personnes nécessitant imputation, à l'échelle nationale, provinciale et territoriale	26
Tableau 5.1	
Ménages qui nécessitent l'imputation de dépenses par province et territoire	27
Tableau 5.2	
Personnes qui nécessitent l'imputation du revenu et personnes qui nécessitent l'imputation de dépenses en vêtements par province et territoire	28
Références	29
Annexe A Notation algébrique	30

Faits saillants

Les erreurs d'échantillonnage

- Les coefficients de variation (CV) de l'estimation de la moyenne des dépenses totales par ménage varient entre 1,2 % et 1,9 % selon les provinces, et il est de 0,7 % à l'échelle nationale. Les CV sont plus élevés dans les territoires, soit 5,8 % au Yukon, 4,7 % dans les Territoires du Nord-Ouest et 8,1 % au Nunavut.
- Les coefficients de variation (CV) de l'estimation de la moyenne pour les différentes catégories de dépenses sommaires sont inférieurs ou égaux à 2,4 % à l'échelle nationale, et généralement inférieurs à 6 % à l'échelle provinciale. Les résultats pour les différentes catégories de caractéristiques des logements et l'équipement ménager sont similaires. Les CV à l'échelle territoriale sont généralement plus élevés.

La non-réponse

- Le taux de non-réponse est de 23,8 %. La non-réponse est due aux refus (15,2 %), aux ménages qu'on n'a pas réussi à contacter (5,2 %) et aux ménages dont les données ont été considérées inutilisables (3,4 %).
- Le taux de non-réponse final a tendance à croître avec le niveau d'urbanisation. On observe un taux de non-réponse de 19,8% en milieu rural, et un taux de 30,6% dans les centres urbains d'un million d'habitants et plus. Cette tendance s'observe également dans les taux de non-réponse à la collecte.
- L'analyse des taux de réponse finaux dans les strates d'aires géographiques à revenu élevé et à faible revenu provenant du plan d'échantillonnage révèle que le taux de non-réponse dans les strates à revenu élevé (37,5 %) est plus élevé que ceux observés dans les strates régulières (23,5 %) et dans les strates à faible revenu (18,6 %). Le taux de refus des strates à revenu élevé est environ le double de celui des strates à faible revenu.

Les erreurs de couverture

- On observe un sous-dénombrement des ménages de 6 % à l'échelle nationale. On observe aussi un sous-dénombrement des ménages pour chacune des provinces et territoires, les taux variant de 4,1 % à 21,6 %. C'est aux Territoires du Nord-Ouest et au Nunavut que le sous-dénombrement des ménages est le plus élevé.
- On observe un sous-dénombrement des personnes de 7,3 % à l'échelle nationale. On observe aussi un sous-dénombrement des personnes pour chacune des provinces et territoires, les taux variant de 4,7 % à 21,3 %. C'est aux Territoires du Nord-Ouest et au Nunavut que le sous-dénombrement des personnes est le plus élevé.
- Les taux de glissement nationaux pour les enfants (0 à 6 ans et 7 à 17 ans) sont très différents de ceux des autres groupes d'âge. En effet, chez les enfants, on observe un surdénombrement ou un léger sous-dénombrement, alors que chez les adultes il

y a toujours sous-dénombrement. Le taux de sous-dénombrement pour l'ensemble des enfants est de 0,7 %, alors qu'il est de 9,2 % pour les adultes.

Les erreurs de réponse

- Les erreurs de réponse comprennent, entre autres, les erreurs de rappel, l'erreur télescopique et les erreurs dues aux réponses par procuration. Dans l'EDM, le fardeau imposé au répondant, étant donné la longueur de l'entrevue, peut entraîner la fatigue du répondant et avoir un impact sur la qualité des données. Le temps total d'entrevue varie selon les caractéristiques des ménages et peut être de plus de cinq heures pour certains ménages. Le temps moyen de l'entrevue était une heure et dix minutes.

Les erreurs de traitement reliées à l'imputation

i) Des variables de dépenses

- Il a fallu imputer au moins une variable de dépenses (excluant les dépenses vestimentaires) pour 7,2 % des ménages. Pour la majorité d'entre eux, seulement une ou deux variables parmi les 251 variables de dépenses ont été imputées.
- Environ 13 % des individus ont nécessité l'imputation de variables sur les dépenses vestimentaires mais pour la très grande majorité, le total était fourni par le répondant et seules les composantes ont été imputées.

ii) Des variables de revenu

- Il a fallu imputer au moins une variable de revenu pour moins de 4 % des individus. Pour environ 80 % d'entre eux, le revenu total était fourni par le répondant mais toutes les composantes (sources de revenu) ont été imputées.

Introduction

L'Enquête sur les dépenses des ménages (EDM) est une enquête annuelle qui collecte les dépenses et les revenus des ménages à l'aide d'entrevues personnelles. L'échantillon de l'EDM 2001 est composé de 25 740 logements¹ répartis dans les 10 provinces et les trois territoires. La collecte est effectuée au cours des mois de janvier, février et mars et les revenus et dépenses sont obtenus pour la période du 1^{er} janvier au 31 décembre de l'année précédente. Suite à un remaniement en 1997, cette enquête remplace l'Enquête périodique sur les dépenses des familles et l'Enquête sur l'équipement ménager (avec ajustements aux questionnaires et aux échantillons).

L'EDM, comme toutes les enquêtes, est sujette aux erreurs et ce malgré toutes les précautions prises aux différentes étapes de l'enquête pour les contrôler. Quoiqu'il n'existe pas de mesure exhaustive de la qualité des données d'une enquête, certaines mesures de qualité produites aux différentes étapes de l'enquête peuvent fournir à l'utilisateur de l'information pertinente pour une bonne interprétation des données.

Ce document vise donc à présenter les indicateurs de qualité produits pour l'enquête sur les dépenses des ménages de 2001. On y trouve les indicateurs de qualité usuels généralement utiles aux utilisateurs pour l'interprétation des données tels que les coefficients de variation, les taux de non-réponse, les taux de glissement et les taux d'imputation.

Les indicateurs de qualité ont été classés selon les principaux types d'erreur que l'on retrouve dans une enquête. Les erreurs d'échantillonnage, c'est-à-dire les erreurs dues au fait que les inférences tirées au sujet de l'ensemble de la population sont fondées sur des renseignements recueillis auprès d'un échantillon de la population et non pas auprès de toute la population, sont présentées dans la section 1. Les sections suivantes couvrent les erreurs non dues à l'échantillonnage. La non-réponse et les erreurs de couverture sont d'abord discutées dans les sections 2 et 3. Les erreurs de réponse et les erreurs de traitement sont traitées dans les sections 4 et 5 respectivement.

Ce document est axé sur la qualité des données. Pour une description détaillée de la méthodologie de l'enquête, on peut consulter la référence [1].

¹ De ces 25 740 logements, on doit identifier et enlever les logements inadmissibles (voir section 2.1) pour en arriver à 22 172 ménages auprès desquels on collecte les dépenses et les revenus.

1. Les erreurs d'échantillonnage

Les erreurs d'échantillonnage découlent du fait que les inférences que l'on tire au sujet de l'ensemble de la population à partir de l'enquête sont fondées sur l'information que l'on a recueillie auprès d'un échantillon de la population et non pas auprès de toute la population. Outre le plan de sondage et la méthode d'estimation qui ont été appliqués pour l'enquête sur les dépenses des ménages, la taille d'échantillon ainsi que la variabilité de chaque caractéristique sont des facteurs déterminants de l'erreur d'échantillonnage. Les caractéristiques qui sont rares ou qui sont distribuées de façon très différente dans la population auront une erreur d'échantillonnage plus grande que les caractéristiques qu'on observe plus fréquemment ou qui sont plus homogènes dans la population.

1.1 Les mesures de l'erreur d'échantillonnage

L'erreur-type est une mesure standard de l'erreur d'échantillonnage. L'erreur-type correspond au degré de variation de l'estimation considérant qu'un échantillon particulier a été choisi, plutôt qu'un autre, parmi tous les échantillons possibles de même taille sous le même plan de sondage. Comme l'EDM utilise un plan d'échantillonnage et une méthode d'estimation complexes, on estime l'erreur-type à partir d'une méthode de rééchantillonnage appelée le Jackknife. Pour plus de détails sur cette méthode, voir la référence [2].

Le coefficient de variation (CV) est également une mesure de fiabilité de l'estimation fréquemment utilisée. Il exprime tout simplement l'erreur-type en pourcentage de l'estimation. Ainsi, si on obtient une estimation Y pour une certaine caractéristique et que SE correspond à l'erreur-type estimée, le CV sera $(SE/Y) \times 100$.

Finalement, l'erreur-type ou le coefficient de variation peuvent être utilisés pour dériver une autre mesure de précision des estimations, l'intervalle de confiance. Cette mesure indique le niveau de confiance selon lequel la valeur réelle dans la population pour une caractéristique qu'on a observée se trouve à l'intérieur de certaines limites. Un intervalle avec un niveau de confiance de 95 % correspond à l'estimation obtenue à partir de l'échantillon ± 2 erreur-type : $(Y \pm 2 SE)$.² Cela signifie que si on répétait l'échantillonnage un grand nombre de fois, chaque échantillon fournirait un intervalle différent et 95 % des intervalles contiendraient la vraie valeur de la caractéristique. De façon similaire, si on répétait l'échantillonnage, l'intervalle $Y \pm SE$ contiendrait la vraie valeur dans 68 % des cas.

1.2 Les coefficients de variation

Des estimations des coefficients de variation sont calculées pour les estimations de plusieurs caractéristiques collectées dans l'EDM. Les CV pour les estimations de la moyenne des dépenses par ménage (pour les catégories de dépenses détaillées), ainsi que pour les caractéristiques des logements et l'équipement ménager, sont disponibles à l'échelle nationale, provinciale et territoriale dans la publication *Guide de l'utilisateur — Enquête sur les dépenses des ménages* (référence [3]).

² L'intervalle de confiance se calcule directement à partir du CV de façon similaire, soit $Y \pm 2 (CV \times Y)/100$.

1,5 % et 2,2 % du total des dépenses (données non présentées). De plus, si on fait exception de ces catégories, les CV sont généralement inférieurs ou égaux à 6 % à l'échelle provinciale.

Le tableau 1.2 donne un aperçu des CV pour les caractéristiques des logements et l'équipement ménager à l'échelle provinciale et territoriale ainsi qu'à l'échelle nationale pour l'estimation de quelques-unes des catégories.

Tableau 1.2
Coefficients de variation (%) par province et territoire ainsi qu'à l'échelle nationale pour quelques catégories des caractéristiques des logements et l'équipement ménager

CATÉGORIES	Can.	T.-N.-L.	Î.-P.-É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sask.	Alb.	C.-B.	Yn	T. N.-O.	Nt
Propriétaire	0,9	1,6	3,1	1,7	1,7	1,9	1,8	1,9	1,8	1,8	1,6	8,6	13,9	16,0
Locataire	1,6	5,1	7,2	4,5	4,5	2,7	3,3	4,7	4,6	4,5	2,9	15,0	13,5	4,1
Machine à laver	0,6	0,9	2,0	1,3	1,1	1,0	1,5	1,7	1,2	1,3	1,4	3,0	4,3	5,4
Sécheuse	0,7	1,1	2,2	1,4	1,3	1,1	1,5	1,7	1,1	1,4	1,3	3,5	4,2	5,7
Lave-vaisselle	1,1	4,3	5,1	3,1	3,2	2,1	2,5	2,8	2,9	2,2	2,0	11,9	13,1	33,6
Congélateur	1,1	1,4	2,5	1,9	2,0	2,5	2,3	1,7	1,7	1,8	2,1	5,8	5,1	9,4
Four à micro-ondes	0,4	0,9	1,2	0,7	0,7	0,9	0,7	0,7	0,9	0,6	0,9	2,9	2,9	6,7
Téléphone cellulaire	1,2	3,9	5,5	3,4	4,0	2,9	2,1	2,9	2,9	2,1	2,5	17,2	15,1	24,4
Lecteur de disques compacts	0,7	1,7	3,3	1,6	1,6	1,6	1,4	1,6	1,5	1,1	1,3	4,3	3,1	3,5
Télédistribution	1,0	2,1	5,3	2,3	2,4	2,4	2,0	2,3	2,8	1,8	1,6	10,3	11,0	3,5
Ordinateur personnel	0,9	2,6	4,3	2,3	2,7	2,2	1,7	2,3	2,6	1,7	1,7	8,3	5,5	14,0
Connexion tél. régulière à un ordinateur (Modem)	1,9	5,1	7,0	4,6	4,0	3,8	3,4	5,4	5,6	4,9	4,4	12,4	11,3	19,2
Utilisation d'Internet (maison)	1,1	3,2	5,4	2,8	3,4	2,5	2,0	2,6	3,3	2,2	2,2	8,5	5,4	17,7
Véhicules possédés (un)	1,5	3,1	4,7	3,2	3,4	2,9	3,2	3,2	3,1	3,2	2,6	8,3	6,3	14,7
Véhicules possédés (2 ou plus)	1,4	4,2	4,6	3,4	3,7	3,3	3,0	3,2	2,6	2,4	2,5	9,1	16,4	37,2

Les coefficients de variation pour les caractéristiques des logements et l'équipement ménager sont généralement inférieurs à 5 % à l'échelle provinciale, avec quelques exceptions dans la catégorie *Connexion téléphonique régulière à un ordinateur (Modem)*. Les CV sont plus élevés dans les territoires où l'on retrouve une proportion moindre des équipements.

Les CV pour les caractéristiques des logements et l'équipement ménager à l'échelle nationale sont inférieurs ou égaux à 1,9 %.

1.3 Modèle pour dériver une approximation du CV

Des estimations pour différents domaines d'intérêt (par ex., par quintile de revenu) sont disponibles pour les catégories de dépenses sommaires dans la publication *Les habitudes de dépenses au Canada* (référence [4]). Des estimations pour différents domaines d'intérêt pour les catégories de dépenses détaillées sont également disponibles sur demande auprès de la Division de la statistique du revenu. (Pour plus de

détails sur les tableaux disponibles auprès de la Division de la statistique du revenu, voir les références [3] ou [4].) Pour des raisons opérationnelles, il n'est pas possible de produire les CV pour tous les différents niveaux d'agrégation qui pourraient intéresser les utilisateurs.

1.3.1 Approximation du CV pour les estimations des domaines

Il est toutefois possible de calculer une approximation du CV en utilisant une relation entre le nombre de ménages de l'échantillon qui ont déclaré des dépenses pour une catégorie et le CV à un niveau agrégé. Cette relation, basée sur la tendance du CV à croître proportionnellement à une diminution de la racine carrée du nombre de ménages déclarant une dépense, est illustrée ci-après.

Formule d'approximation du CV pour un domaine (un sous-groupe de la population)

Si $CV(Y)$ représente le CV pour l'estimation de la moyenne par ménage d'une certaine caractéristique pour toute la population, alors on peut calculer une approximation du CV de l'estimation de cette caractéristique pour un domaine (que l'on peut considérer comme un sous-groupe de la population tel qu'un type de ménage, un quintile de revenu, un niveau d'urbanisation, ...) à partir de l'équation suivante :

$$CV(Y_d) = CV(Y) \times \sqrt{\frac{nP}{n_d P_d}}$$

Où

n : le nombre de ménages dans l'échantillon

P : l'estimation de la proportion des ménages déclarant une valeur > 0 pour cette caractéristique dans la population

n_d : le nombre de ménages de l'échantillon dans le domaine d

P_d : l'estimation de la proportion des ménages déclarant une valeur > 0 pour cette caractéristique dans le domaine d

Généralement on utilise le CV, la taille n et la proportion P à l'échelle nationale pour calculer les approximations pour les différents domaines. Dans le cas où on cherche à calculer une approximation du CV pour un domaine entièrement contenu dans une seule province (par ex. région métropolitaine), il est préférable d'utiliser ces valeurs à l'échelle provinciale puisque les CV provinciaux sont publiés pour l'EDM de 2001 (référence [3]). Il est important de se rappeler que la valeur du CV obtenue par cette approche n'est qu'une approximation du CV.

1.3.2 Approximation du CV à partir du fichier de microdonnées

Les utilisateurs du fichier de microdonnées peuvent se servir d'une autre approche pour dériver une approximation du CV des estimations, qui sera généralement plus performante que celle présentée dans la section précédente pour les CV des catégories de dépenses détaillées. Cette approche est décrite en détails dans la documentation accompagnant le fichier de microdonnées de 2001. Elle est utilisable seulement à partir

du fichier de microdonnées puisqu'il est nécessaire d'avoir les données et les poids de chaque ménage pour calculer cette approximation.

Le document sur la qualité des données de l'enquête de 1997 renferme des résultats de l'évaluation de la performance de ces deux méthodes d'approximation du CV.

1.4 Suppression des données non fiables dans les tableaux d'estimations

Comme le coefficient de variation est un indicateur de la fiabilité des données, on aimerait l'utiliser pour déterminer si les estimations devraient être publiées ou non. On considère que les estimations dont le CV est estimé à plus de 33 % ne sont pas suffisamment fiables pour être publiées. Toutefois les estimations des CV ne sont pas calculées pour une grande partie des estimations publiées, la règle de suppression pour les estimations de dépenses est donc basée sur le nombre de ménages qui déclarent une valeur supérieure à zéro.³

On peut démontrer que les CV sont en général inférieurs à 33 % si le nombre de ménages qui déclarent une dépense est supérieur à 30. Comme il s'agit d'une règle approximative, certaines estimations peuvent être publiées même si le CV est supérieur à 33 % et certaines estimations ne seront pas publiées malgré un CV inférieur à 33 %. Le document sur la qualité des données de l'EDM 1997 donne les résultats de l'évaluation des risques d'erreurs de l'utilisation de la règle de suppression.

2. La non-réponse

Les erreurs dues à la non-réponse découlent du fait que certains répondants potentiels ne fournissent pas l'information nécessaire ou que cette information s'avère inutilisable. Lorsque le répondant a omis de répondre à certaines questions seulement, on parle alors de non-réponse partielle. Dans ce cas, les données manquantes sont imputées. Les erreurs liées à l'imputation sont présentées dans la section 5 portant sur les erreurs de traitement. Dans la présente section, la non-réponse englobe la non-réponse à la collecte, due principalement à l'impossibilité de contacter le ménage ou au refus des membres du ménage de participer à l'enquête, que ce soit partiellement ou complètement, ainsi que les ménages pour lesquels les données ont été collectées mais sont inutilisables.

Le principal impact de la non-réponse sur la qualité des données est qu'elle peut induire un biais dans les estimations si les caractéristiques des répondants et des non-répondants diffèrent et que cette différence a un impact sur les caractéristiques étudiées. Les taux de non-réponse peuvent être calculés facilement mais ils n'ont qu'une valeur indicative à l'égard de la qualité des données car ils ne permettent pas de mesurer l'importance du biais associé aux estimations. L'ampleur de la non-réponse peut être considérée comme une évaluation des risques de biais dans les estimations.

³ En pratique, on utilise l'estimation de la proportion des ménages déclarant une dépense que l'on multiplie par la taille d'échantillon.

2.1 Les taux de non-réponse et les taux de vacance

Dans l'EDM, comme les unités sélectionnées sont des logements, les intervieweurs doivent d'abord identifier les logements inadmissibles, c'est-à-dire les logements occupés par des personnes ne faisant pas partie de la population cible, les logements qui n'existent plus (démoli, maison mobile déplacée ou logement converti en entreprise) et les logements vacants (inoccupés, saisonniers ou en construction).

Parmi les logements admissibles, on évalue ensuite la proportion des ménages qui n'ont pas répondu à l'enquête, qu'on appelle le taux de non-réponse à la collecte. Ces derniers comprennent les ménages qui ont refusé de participer à l'enquête et les ménages où il a été impossible d'établir un contact avec les répondants parce qu'ils étaient absents ou encore à cause de circonstances spéciales (problème de langue, maladie ou décès).

Toujours parmi les logements admissibles, on détermine également le taux de données inutilisables. Les données inutilisables correspondent au nombre de ménages pour lesquels le questionnaire était au moins partiellement complet mais qui ont été rejetés lors du traitement des données. Il existe deux causes principales de rejet. D'abord lorsqu'une partie importante des questions sur le revenu ou des questions sur les dépenses a été laissée sans réponse, le questionnaire est classé incomplet et n'est pas utilisé. L'autre source de rejet correspond aux questionnaires pour lesquels la différence entre les entrées (revenu et autres sources d'argent du ménage) et les déboursés (dépenses et variation nette des actifs et passifs) est supérieure à 20 %. Ces questionnaires sont également exclus de l'estimation et considérés comme de la non-réponse.

Le tableau 2.1 présente le taux de non-réponse final, le taux de non-réponse à la collecte, ventilé selon les refus et les non-contacts ainsi que le taux de données inutilisables, ventilé selon les incomplets et les non balancés. On y trouve également le taux de vacance. Ces taux sont fournis tant à l'échelle nationale que provinciale et territoriale.

Tableau 2.1
Taux de non-réponse (%) et taux de vacance (%) par province et territoire ainsi qu'à l'échelle nationale

Province ou Territoire	Taux de vacance	Taux de non-réponse À la collecte			Taux de données inutilisables			Taux de non-réponse final (à l'estimation)
		TOTAL	Pas de contact	Refus	TOTAL	Incomplet	Non balancé	
Canada	11,5	20,4	5,2	15,2	3,4	1,9	1,5	23,8
Terre-Neuve-et-Labrador	18,4	16,1	5,4	10,8	2,1	1,1	1,0	18,3
Île-du-Prince-Édouard	19,3	15,3	4,6	10,7	0,8	0,4	0,4	16,0
Nouvelle-Écosse	13,5	21,6	6,7	14,9	4,3	3,0	1,3	25,9
Nouveau-Brunswick	15,9	18,4	4,5	14,0	2,8	1,4	1,4	21,2
Québec	6,9	21,1	6,5	14,6	1,5	0,9	0,6	22,7
Ontario	7,7	28,1	7,3	20,8	7,7	4,8	2,9	35,8
Manitoba	13,8	15,4	3,3	12,1	1,8	0,6	1,2	17,2
Saskatchewan	10,8	17,5	4,1	13,4	2,9	1,2	1,7	20,4
Alberta	5,3	19,3	3,2	16,1	3,5	1,5	2,0	22,8
Colombie-Britannique	7,7	22,3	3,8	18,5	3,1	1,8	1,3	25,4
Yukon	31,4	24,1	3,8	20,3	3,2	2,3	0,9	27,2
Territoires du Nord-Ouest	13,3	11,7	6,1	5,6	1,5	0,7	0,7	13,1
Nunavut	12,6	15,0	7,7	7,3	0,4	0,4	0,0	15,4

Le taux de non-réponse final au Canada est de 23,8 %. Il est dû aux refus (15,2 %), aux ménages qu'il a été impossible de rejoindre (5,2 %), et finalement aux ménages dont les données étaient inutilisables (3,4 %). Pour presque toutes les provinces, les refus représentent la cause majeure de non-réponse, suivis des non-contacts, puis des données inutilisables. Les seules exceptions sont l'Ontario et l'Alberta pour lesquelles la proportion de données inutilisables est légèrement supérieure à la proportion de non-contacts.

Les taux de non-réponse finaux varient selon les provinces. C'est aux Territoires du Nord-Ouest que l'on observe le taux de non-réponse le plus bas, soit de 13,1 %. C'est également pour ce territoire que la proportion de refus est la plus faible (5,6 %). Les taux de non-réponse de Terre-Neuve-et-Labrador, de l'Île-du-Prince-Édouard, du Manitoba et du Nunavut sont inférieurs à 20 %, tandis que les taux sont supérieurs à 20 % pour les autres provinces et territoires. Le taux de non-réponse de l'Ontario est particulièrement élevé atteignant presque 36 %. C'est aussi dans cette province que l'on trouve le plus haut taux de refus (20,8 %).

Le taux de vacance est présenté au tableau 2.1, mais on doit considérer que les logements vacants ne contribuent pas au biais de l'échantillon dans la mesure où ils sont identifiés correctement. En analysant les taux de vacance, on décèle les problèmes d'identification des logements liés à la collecte. Le taux de vacance de l'EDM de 2001 est de 11,5 %. Le taux de vacance de l'EDM est légèrement plus bas que celui de l'Enquête sur la population active (EPA) pour la même période. Notons que le taux de vacance plus élevé (31,4 %) au Yukon est dû dans une certaine mesure à une répartition de l'échantillon différente en 2001 pour ce territoire et dans une moindre mesure aux changements apportés par l'EPA quant à la couverture au Yukon.

2.2 La non-réponse selon le niveau d'urbanisation

La non-réponse varie selon le niveau d'urbanisation. Les divers taux à l'échelle nationale sont présentés par niveau d'urbanisation dans le tableau 2.2.⁴

Tableau 2.2
Taux de non-réponse (%) et taux de vacance (%) par niveau d'urbanisation

Catégorie d'urbanisation	Taux de vacance	Taux de non-réponse à la collecte			Taux de données inutilisables			Taux de non-réponse final (à l'estimation)
		TOTAL	Pas de contact	Refus	TOTAL	Incomplet	Non balancé	
URBAIN								
1 000 000 ou plus	4,1	26,8	7,8	19,0	3,7	2,1	1,6	30,6
500 000 à 999 999	3,9	18,7	3,3	15,3	2,7	1,0	1,7	21,4
250 000 à 499 999	6,3	26,0	6,9	19,2	5,9	3,8	2,2	32,0
100 000 à 249 999	7,9	22,0	5,0	17,0	3,3	2,1	1,1	25,3
30 000 à 99 999	5,3	19,1	3,7	15,4	3,3	2,1	1,1	22,3
Moins de 30 000	12,5	17,1	4,6	12,5	3,0	1,6	1,4	20,1
RURAL	23,4	16,6	4,6	12,0	3,2	1,8	1,4	19,8
TOTAL	11,5	20,4	5,2	15,2	3,4	1,9	1,5	23,8

Le taux de non-réponse final augmente généralement avec le niveau d'urbanisation. Selon les résultats du tableau 2.2, il n'y a que le groupe des 500 000 à 999 999 et, dans une moindre mesure, le groupe des 1 000 000 ou plus, qui ne suivent pas cette tendance.

Le taux de non-réponse à la collecte a lui aussi tendance à croître avec le niveau d'urbanisation. Il y a un écart important de près de 10% entre les catégories d'urbanisation «Moins de 30 000» et «1 000 000 ou plus». Les refus comptent pour plus de 60 % de la non-réponse totale à tous les niveaux d'urbanisation. Bien que la catégorie d'urbanisation « 500 000 à 999 999 » possède le plus faible taux de non-réponse totale pour les régions urbaines de 30 000 habitants ou plus, c'est pour celle-ci que la proportion de non-réponse due au refus des ménages de participer à l'enquête est la plus élevée (71,5 %).

En examinant les taux de vacance par niveau d'urbanisation, il ressort que le taux de vacance est près de deux fois plus élevé en région rurale (23,4 %) que dans les régions urbaines à faible population (12,5 %). Ces dernières régions ont aussi un taux de vacance supérieur à celui des régions urbaines à plus forte population. Ce phénomène s'observe également dans l'EPA et s'explique sans doute par un plus grand nombre de logements saisonniers en milieu rural. Comme l'échantillon de l'EDM est plus concentré dans les régions urbaines à forte population que celui de l'EPA, on peut s'attendre à observer un taux de vacance national légèrement inférieur.

⁴ Des tableaux portant sur les taux de non-réponse selon le niveau d'urbanisation par province sont disponibles sur demande auprès de la Division des méthodes d'enquêtes auprès des ménages.

2.3 La non-réponse selon les strates de revenu

Il est impossible de comparer le taux de réponse selon le revenu car cette information n'est pas accessible pour les non-répondants. Toutefois, le plan d'échantillonnage de l'EPA, utilisé pour l'EDM, a été conçu de sorte à former, dans neuf grandes villes, des strates d'aires géographiques où le revenu moyen des ménages excède 100 000 \$, et dans sept grandes villes, des strates composées d'appartements habités par des ménages ayant un revenu moyen inférieur à 20 000 \$. Même si le nombre de telles strates est petit et qu'il compte pour seulement un petit nombre de logements dans l'échantillon de l'EDM (environ 580 pour les strates de revenus élevés et 160 pour les strates de revenus faibles, soit moins de 3 % de l'échantillon), la comparaison des taux de non-réponse dans ces deux groupes par rapport à l'ensemble des autres strates est révélatrice. Le tableau 2.3 présente ces résultats.

Tableau 2.3
Comparaison des taux (%) de non-réponse et de vacance des strates à revenu élevé et à faible revenu par rapport aux autres strates

Type de strate en fonction du revenu	Taux de vacance	Taux de non-réponse à la collecte			Taux de données inutilisables			Taux de non-réponse final (à l'estimation)
		TOTAL	Pas de contact	Refus	TOTAL	Incomplet	Non balancé	
Revenu élevé	2,9	31,5	8,8	22,7	5,9	1,8	4,1	37,5
Régulières	11,7	20,1	5,0	15,1	3,3	1,9	1,4	23,5
Revenu faible	5,1	17,9	6,2	11,7	0,7	0,7	0,0	18,6
TOTAL	11,5	20,4	5,2	15,2	3,4	1,9	1,5	23,8

Dans les strates à revenu élevé, le taux de non-réponse final (37,5 %) est de beaucoup supérieur à celui dans les strates régulières. Le taux de refus des strates à revenu élevé est de plus de 20%, ce qui est environ le double de celui des strates à faible revenu.

Les ménages des strates à faible revenu ont un taux de non-réponse à la collecte similaire à celui des strates régulières. Toutefois, les strates régulières ont un taux de non-réponse final de 26% supérieur à celui des strates à faible revenu. Ceci s'explique en grande partie par le fait qu'il n'y a presque pas de données inutilisables pour les strates à faible revenu.

Comme pour les enquêtes de 1997 à 2000, le taux de vacance est plus élevé pour les strates régulières que pour chacun des deux autres types de strate. Dans les strates à revenu élevé et à revenu faible, les taux de vacance sont semblables.

2.4 L'ajustement pour la non-réponse

Pour compenser la non-réponse, les poids de l'EDM sont gonflés par l'inverse du taux de réponse pondéré à l'intérieur de certains groupes définis en fonction des différents niveaux d'urbanisation dans chaque province. Les taux pondérés diffèrent des taux présentés dans cette section puisqu'ils tiennent compte du poids de sondage de chaque ménage. Une description algébrique de l'ajustement pour la non-réponse est présentée à l'annexe A.

L'ajustement des poids pour la non-réponse permet de tenir compte des différences au niveau de la non-réponse par niveau d'urbanisation mentionnées dans la section 2.2. Il permettra de réduire le biais dans la mesure où les caractéristiques des répondants et des non-répondants sont similaires pour un même niveau d'urbanisation.

3. Erreurs de couverture

Lors de la conception de l'enquête, on a défini la population visée, qu'on appelle la population cible. Rappelons d'abord la définition de cette population cible pour l'EDM puisqu'une bonne compréhension de la population cible est nécessaire à une bonne interprétation des données de l'enquête. Il est important de préciser que l'EDM utilise la base de sondage de l'enquête sur la population active (EPA).

La population cible

La population cible correspond aux individus vivant dans les ménages privés. Les pensionnaires d'établissement institutionnel tel que les prisons, les hôpitaux pour malades chroniques, les résidences pour personnes âgées ainsi que les membres d'ordres religieux et d'autres groupes vivant en communauté, les membres des Forces Armées vivant dans les camps militaires et les individus vivant de façon permanente dans les hôtels ou les maisons de chambres sont donc exclus. On exclut aussi les représentants officiels de pays étrangers qui vivent au Canada et leurs familles ainsi que les résidents des réserves indiennes et des terres publiques. L'enquête couvre donc près de 98 % de la population dans les 10 provinces. Au Yukon, la couverture de la population est passé de 81 % en 1999 à 88 % en 2001. Cette hausse est due aux modifications apportées à la couverture de ce territoire pour l'EPA. Ainsi, certaines communautés et régions non organisées qui étaient exclues sont dorénavant incluses. La couverture des Territoires du Nord-Ouest correspond à 92 % de la population et la couverture du Nunavut est de 89 %.⁵

Les erreurs de couverture proviennent d'une représentation inadéquate de la population cible à partir des unités de la base de sondage. Certaines unités de la population cible peuvent être omises de la base de sondage, on parle alors d'un sous-dénombrement. D'autres unités qui ne sont pas dans la population cible peuvent être incluses par erreur ou certaines unités peuvent être incluses plus d'une fois, ces unités sont responsables du surdénombrement.

⁵ Pour ce qui est de la proportion de ménages, la couverture du Yukon, des Territoires du Nord-Ouest et du Nunavut correspondent respectivement à 88 %, 93 % et 90 % des ménages.

3.1 Le sous-dénombrement et le surdénombrement : les taux de glissement

Dans l'EDM, la sélection de l'échantillon est effectuée en utilisant une liste de logements dans chaque grappe sélectionnée. L'omission de logements lors de la création de la liste, les nouveaux logements qui s'ajoutent entre la création de la liste et la visite des intervieweurs (principalement dans les secteurs en développement) ainsi que la classification erronée de logements vacants contribuent au sous-dénombrement. Le fait d'inclure des logements qui ne seraient pas dans les limites de la grappe est une source de surdénombrement. De façon similaire, des erreurs peuvent se glisser lors de la collecte des données, lors de l'identification des personnes qui sont membres du ménage sélectionné. Ces erreurs contribuent également au sous-dénombrement et au surdénombrement.

Une bonne représentation de la population cible est indispensable à la production d'estimations de dépenses réalistes. Le nombre de personnes par ménage est également une caractéristique importante dans l'estimation des dépenses moyennes des ménages. Il est donc nécessaire que l'échantillon, non seulement représente adéquatement les individus de la population cible, mais également la distribution des ménages selon leur taille.

En 1999, une nouvelle méthode de pondération a été introduite utilisant de nouveaux groupes de contrôle. Cette nouvelle méthode nous permet de mieux corriger la représentativité de la population visée en ciblant des groupes d'âge et de sexe mieux délimités que ceux utilisés précédemment et pour lesquels la couverture varie d'un groupe à l'autre.

On observe généralement un sous-dénombrement net du nombre de personnes et du nombre de ménages dans l'EDM, que l'on rectifie par un ajustement des poids en utilisant des données auxiliaires, qui sont basées sur des estimations démographiques post-censitaires. Le taux de glissement (voir annexe A) est une mesure du pourcentage d'écart entre les estimations provenant de ces données auxiliaires et les estimations de l'enquête calculées en utilisant des poids non ajustés avec ces données.⁶ Les taux de glissement par groupe d'âge-sexe à l'échelle nationale, provinciale et territoriale sont présentés dans les tableaux 3.1 et 3.2 alors que les taux de glissement selon les tailles de ménages que l'on utilise lors de l'ajustement des poids se trouvent dans le tableau 3.3. Un taux positif correspond à un surdénombrement du nombre de personnes ou de ménages dans l'enquête.

⁶ On utilise les sous-poids, c'est-à-dire les poids de sondage ajustés pour la non-réponse (voir Annexe A).

Tableau 3.1
Taux de glissement nationaux selon le groupe d'âge-sexe

Taux de glissement national par groupe d'âge-sexe				
	Âge	Sexe		Total
		Hommes	Femmes	
Canada ¹	0-6 ans	0,5	5,2	2,8
	7-17 ans	-3,5	-1,6	-2,6
	18-24 ans	-6,5	-3,2	-4,9
	25-34 ans	-13,4	-10,2	-11,8
	35-54 ans	-13,6	-7,9	-10,7
	55-59 ans	-7,3	-15,9	-11,7
	60-64 ans	-12,4	-4,7	-8,4
	65-69 ans	-6,4	-0,8	-3,5
	70 ans et +	-9,4	-2,5	-5,4
	Total	-9,2	-5,4	-7,3

¹ Exclut les territoires.

Pour l'EDM de 2001, le taux de sous-dénombrement de la population était 7,3 %. En analysant le tableau 3.1 au niveau des groupes d'âge, on constate qu'à l'échelle nationale, les taux de glissement pour les enfants (0 à 6 ans et 7 à 17 ans) sont très différents de ceux des autres groupes d'âge. En effet, chez les enfants, on observe un surdénombrement ou un léger sous-dénombrement, alors que chez les adultes il y a toujours sous-dénombrement. Le taux de sous-dénombrement pour l'ensemble des enfants est de 0,7 %, alors qu'il est de 9,2 % pour les adultes (données non présentées).

Les taux les plus élevés à l'échelle nationale sont observés chez les femmes de 55 à 59 ans, ainsi que chez les hommes de 25 à 34 ans et de 35 à 54 ans. À l'exception de la catégorie des 55 à 59 ans, le taux de sous-dénombrement chez les femmes est toujours inférieur à celui des hommes.

Tel que mentionné précédemment, l'EDM utilise la base de sondage de l'enquête sur la population active (EPA). Pour une même période, le taux de sous-dénombrement de l'EPA à l'échelle nationale est de 9,8 % (référence [5]), ce qui est légèrement supérieur au taux de 8,8 % de l'EDM observé lorsqu'on se restreint aux personnes de 15 ans et plus (données non présentées).

Tableau 3.2
Taux de glissement pour les provinces et les territoires selon le groupe d'âge-
sexe

Taux de glissement par groupe d'âge-sexe																	
Sexe	Âge	Terre-Neuve-et-Labrador	Île-du-Prince-Édouard	Nouvelle-Écosse	Nouveau-Brunswick	Québec	Ontario	Manitoba	Saskatchewan	Alberta	Colombie-Britannique	Yukon	Territoires du Nord-Ouest	Nunavut			
Homme	0-6	-0,1	-21,1	-2,0	-8,1	12,0	0,2	-14,8	6,4	-8,3	-4,7	7,4	-18,3	-19,4			
	7-17	-17,4	5,1	-3,3	-7,4	-10,9	7,3	-6,8	1,6	-8,1	-18,6						
	18-24	-27,1	-31,6	-18,0	-23,0	-1,5	1,9	-13,9	-31,6	-21,9	-9,0						
	25-34	-43,3	-13,8	-23,8	-25,9	-7,8	-9,6	-17,2	-12,7	-13,7	-25,6						
	35-54	-14,0	-15,5	-11,9	-15,5	-12,5	-11,9	-11,3	-17,5	-13,7	-19,9						
	55-59	-18,6	-11,3	-15,9	-6,0	-1,3	-8,8	-11,8	-24,3	-1,5	-10,2				-18,1	-27,4	-20,3
	60-64	-0,3	-6,5	-12,1	-38,1	-7,7	-15,7	-16,4	-28,5	-11,5	-4,8						
	65-69	15,5	-15,5	15,2	4,5	-5,4	-11,7	-10,8	-13,1	2,5	-5,1						
	70 +	9,2	-3,9	-11,6	-4,6	-12,7	-12,7	-0,3	2,0	-1,9	-7,3						
	Total	-16,1	-12,6	-11,2	-14,8	-7,5	-6,3	-11,3	-11,9	-11,4	-15,5				-11,5	-24,6	-19,9
Femme	0-6	-21,7	39,3	-1,6	0,8	14,5	3,8	7,9	8,6	4,9	-4,3	10,3	-15,8	-22,8			
	7-17	-9,9	-1,9	0,4	-4,5	-4,2	8,3	-10,4	-2,6	-10,8	-16,2						
	18-24	-30,4	-34,8	-18,5	-16,3	14,7	-5,2	-14,9	-24,6	5,6	-17,7						
	25-34	-27,3	-8,1	-15,9	-11,8	-2,1	-11,7	-2,0	0,4	-11,9	-18,4						
	35-54	-11,6	-12,0	-6,3	-5,9	-9,9	-3,0	-7,5	-7,4	-8,7	-17,8						
	55-59	-8,0	4,5	-10,1	-10,5	-2,1	-23,3	-27,6	-28,2	-17,6	-20,7				-10,9	-19,0	-13,3
	60-64	12,4	-9,9	15,9	2,2	-7,0	-7,3	-12,8	-5,6	-1,0	-1,0						
	65-69	-14,3	-23,1	-9,7	6,9	2,8	-4,4	14,7	-9,9	13,8	-5,0						
	70 +	1,9	6,2	-10,1	-12,1	-8,2	1,6	-2,4	-1,0	3,5	-4,3						
	Total	-13,7	-6,0	-7,2	-7,0	-2,8	-3,2	-6,5	-6,2	-5,3	-14,3				-5,7	-17,9	-17,4
Total	-14,9	-9,3	-9,1	-10,9	-5,2	-4,7	-8,9	-9,1	-8,4	-14,9	-8,6	-21,3	-18,7				

On observe un sous-dénombrement net pour chacune des provinces et territoires, les taux variant de 4,7 % à 21,3 %. L'ampleur du sous-dénombrement est plus importante aux Territoires du Nord-Ouest ainsi qu'au Nunavut. On doit mentionner qu'un faible taux de sous-dénombrement global n'est toutefois pas garant d'une meilleure couverture. Par exemple, le taux de glissement général observé à l'Île-du-Prince-Édouard (-9,3 %) cache le pire cas de surdénombrement à l'échelle provinciale des groupes d'âge-sexe (39,3 % chez les filles de 0 à 6 ans) et les troisième et quatrième pires cas de sous-dénombrement (respectivement de 34,8 % et 31,6% chez les femmes et les hommes de 18 à 24 ans). En revanche, bien que le taux de glissement général observé en Colombie-Britannique soit de -14,9%, cette province n'affiche que le treizième pire cas de sous-dénombrement (25,6% chez les hommes de 25 à 34 ans). On remarque également que le schéma de variation des taux de glissement pour les groupes d'âge-sexe diffère sensiblement d'une province à l'autre.

Tableau 3.3
Taux de glissement pour les provinces et les territoires selon la taille de ménage

Province ou territoire	Taux de glissement			
	Ménages	Ménages d'une personne	Ménages de deux personnes	Ménages de trois personnes et plus
Canada	-6,0	-4,0	-6,1	-7,3
Terre-Neuve-et-Labrador	-10,1	0,8	-0,3	-20,3
Île-du-Prince-Édouard	-6,3	-4,6	6,2	-16,2
Nouvelle-Écosse	-7,4	-2,8	-5,8	-11,4
Nouveau-Brunswick	-6,7	0,2	1,0	-16,4
Québec	-5,3	-10,9	0,0	-5,4
Ontario	-4,9	0,0	-12,9	-2,1
Manitoba	-4,4	4,8	-6,0	-9,6
Saskatchewan	-7,0	-4,5	-2,0	-13,1
Alberta	-4,1	-0,7	2,6	-10,9
Colombie-Britannique	-11,0	-4,1	-9,5	-16,8
Yukon	-12,2	-32,1	-3,9	-4,7
Territoires du Nord-Ouest	-21,6	-17,9	-5,0	-29,6
Nunavut	-17,7 ⁷

On observe une sous-estimation du nombre de ménages de 6 % à l'échelle nationale. Ce taux de glissement est comparable au taux de glissement pour ce qui est des personnes qui est de -7,3 %.

À l'échelle provinciale et territoriale, il y a une variation importante du taux de glissement pour les ménages d'une personne. Le taux varie de -32,1 % pour le Yukon jusqu'à 4,8 % pour le Manitoba. Le même phénomène est observable pour les ménages de deux personnes et de trois personnes et plus mais avec une variation moindre. Pour les ménages de deux personnes, les taux de glissement varient de -12,9 % pour l'Ontario à 6,2 % pour l'Île-du-Prince-Édouard. Pour les ménages de trois personnes et plus, les taux de glissement varient de -29,6 % pour les Territoires du Nord-Ouest à -2,1% pour l'Ontario. Notons que l'on observe toujours un sous-dénombrement pour la catégorie des ménages de trois personnes et plus. À l'exception du Québec, de l'Ontario et du Yukon, le sous-dénombrement le plus élevé est observé pour les ménages de trois personnes et plus.

3.2 L'ajustement pour l'erreur de couverture de la population et du nombre de ménages

Pour corriger le problème de couverture illustré dans les tableaux 3.1 et 3.2 et réduire le biais qui en découle, on ajuste les données de l'enquête lors de la pondération en utilisant les estimations démographiques pour les groupes d'âge-sexe définis dans ces tableaux, et ce pour chaque province. Pour plus de détails sur la méthodologie de l'ajustement, voir la référence [1]. Cet ajustement réduit considérablement le biais engendré par les erreurs de couverture mais il n'élimine pas complètement le biais si les

⁷ Pour le Nunavut, seul le nombre total de ménages est utilisé.

caractéristiques des individus que l'on a omis de l'enquête diffèrent de celles des individus inclus pour un même groupe d'âge dans une province.

On notera également que l'efficacité de l'ajustement de couverture à partir des estimations démographiques dépend beaucoup de la qualité de ces estimations démographiques et de leur exactitude à bien représenter la population cible de l'enquête. Les estimations démographiques ne sont pas exemptes d'erreurs. Ce sont des estimations post-censitaires basées sur les comptes de population du recensement de 1996, ajustées pour le sous-dénombrement net et qui tiennent compte des récentes statistiques sur la migration, les naissances, la mortalité, etc. Ces estimations démographiques sont ajustées pour tenir compte de certaines exclusions spécifiques aux enquêtes auprès des ménages comme les personnes vivant en institutions. Conceptuellement, elles diffèrent légèrement de la population cible de l'EDM en incluant les personnes habitant les ménages collectifs qui ne sont pas des institutions, par exemple les membres de groupes vivant en communauté et les individus vivant de façon permanente dans les hôtels ou les maisons de chambres. Cependant, on considère cette différence négligeable puisque ces individus représentent moins de 0,4 % de la population canadienne.

Pour corriger le problème de représentativité de l'échantillon par rapport au nombre de ménages selon leur taille, illustré dans le tableau 3.3, on ajuste les données de l'enquête à l'aide de données auxiliaires. En ajustant les poids de l'EDM pour refléter les estimations post-censitaires du nombre de ménages selon la taille, on vise à compenser le biais qui découle d'une représentation inadéquate des ménages. Toutefois on n'éliminera pas nécessairement ce biais si les caractéristiques des ménages non interviewés (omis ou non-répondants) diffèrent de celles des ménages répondants pour une même taille de ménage. Comme dans le cas des estimations démographiques de population, l'efficacité de l'ajustement dépend de la qualité des données auxiliaires sur le nombre de ménages. À ce titre, on notera que certains changements ont été apportés à la méthodologie utilisée pour la production des données auxiliaires sur le nombre de ménages pour les Territoires du Nord-Ouest et le Nunavut. L'impact de ces changements sur les résultats de tests statistiques visant la comparaison des estimations de l'EDM 2001 avec celles de l'EDM 1999 pour ces deux territoires n'a toutefois pas été étudié.

En plus des estimations démographiques quant aux groupes âge-sexe par province, trois autres ensembles de données auxiliaires sont utilisés pour ajuster les données d'enquêtes lors de la pondération en vue d'en améliorer la représentativité. Le premier ensemble de données est utilisé pour contrôler le nombre d'enfants et d'adultes dans certaines grandes villes. Le second est conçu pour contrôler le nombre de ménages monoparentaux et de couples avec enfants par province. Finalement, des comptes par grandes catégories de revenu en salaires et traitements sont utilisés lors de l'ajustement des poids afin d'assurer une certaine cohérence entre les distributions de revenu de l'EDM et celles provenant de sources externes.

4. Les erreurs de réponse

Les erreurs de réponse correspondent au manque d'exactitude des réponses aux questions. Elles peuvent être attribuables à différents facteurs parmi lesquels se retrouvent un questionnaire qui nécessite des améliorations, une mauvaise interprétation des questions par les intervieweurs ou les répondants ainsi que des erreurs dans les déclarations des répondants.

Dans l'EDM, les erreurs dans les déclarations des répondants peuvent survenir pour différentes raisons. Premièrement, il y a les erreurs de rappel qui se produisent lorsque le répondant oublie les dépenses effectuées durant la période de l'enquête (qui correspond à l'année civile) ou fournit une valeur erronée à cause de l'intervalle de temps écoulé entre le moment de l'achat et la date de l'entrevue. Les erreurs de rappel sont probablement la source d'erreur de réponse la plus importante de l'enquête étant donné que la période de référence est longue (12 mois) et qu'une grande variété d'information est demandée.

Une des principales mesures utilisées pour minimiser l'erreur de rappel dans l'EDM est de calculer, pour chaque ménage, la différence entre les entrées d'argent (revenus et autres montants reçus par le ménage) et les déboursés (dépenses plus variation nette de l'actif et du passif). Lorsque la différence excède 10 % des entrées d'argent ou des déboursés, le montant le plus élevé étant retenu, les répondants sont contactés à nouveau pour obtenir des renseignements supplémentaires et tenter d'identifier les erreurs ou les omissions. On encourage également le répondant à consulter divers documents (factures, relevés bancaires,...) pour fournir des données plus exactes. Pour déterminer les dépenses pour de petits articles achetés à intervalles réguliers, les intervieweurs suggèrent généralement aux répondants d'estimer la fréquence des achats et le prix généralement payé pour dériver les dépenses sur une période de 12 mois.

Une seconde source d'erreur dans la déclaration des répondants est l'erreur télescopique qui consiste à inclure dans la période de référence des événements qui se sont produits avant ou après celle-ci. Dans l'EDM, on considère que le fait d'utiliser l'année civile est un bon repère pour le début de la période de référence. De plus, comme la période de référence est longue, l'impact de l'erreur télescopique est moins important.

Les réponses par procuration peuvent également contribuer à l'erreur de réponse. Le membre du ménage qui a effectué une dépense est généralement la personne qui peut la rapporter avec la meilleure précision. C'est sûrement le cas par exemple pour les achats personnels. Les dépenses rapportées par personne interposée sont plus susceptibles d'être entachées d'une erreur de réponse et ce type d'erreur risque d'affecter plus sérieusement certaines catégories de dépenses.

Parmi d'autres sources d'erreur de réponse, il ne faut pas oublier le degré de coopération du répondant. Il se peut que le répondant décide pour des raisons personnelles de ne pas mentionner certaines dépenses ou de déformer les faits.

Dans l'EDM, on considère également que le fardeau imposé au répondant, étant donné la longueur de l'entrevue et la grande variété d'articles à rapporter, ainsi que le rythme

de l'entrevue peuvent entraîner la fatigue du répondant et avoir un impact sur la qualité des réponses obtenues. Le temps d'entrevue varie beaucoup d'un ménage à l'autre selon la taille du ménage, le revenu et plusieurs autres caractéristiques. Pour certains ménages, l'entrevue peut prendre plus de cinq heures.

Il convient de signaler que l'on a ajouté des questions à l'EDM de 2001. Entre autres, pour 2001 seulement, des questions supplémentaires à l'EDM ont été ajoutées pour que les données de l'enquête puissent servir à la pondération de l'Indice des prix à la consommation. Ce changement peut nuire aux comparaisons historiques à l'égard de certaines variables. Par exemple, on a ajouté des questions à la rubrique « Produits de soins personnels » afin de recueillir des renseignements précis sur les produits de soins, le maquillage, les parfums, les désodorisants et les produits pour l'hygiène buccale. Ainsi, il se peut que les répondants aient donné un nombre accru de renseignements précis et que l'estimation à la hausse pour les « Produits de soins personnels » soit attribuable, du moins en partie, à la capacité accrue qu'ont les répondants de se souvenir de certains détails. L'effet des questions supplémentaires sur les estimations est difficile à quantifier.

Quoique l'on considère que les erreurs de réponses sont une source importante d'erreur dans une entrevue rétrospective, elles demeurent l'aspect lié à la qualité des données le plus difficile à mesurer. Il faut généralement entreprendre des études spéciales très coûteuses pour tenter de les mesurer. On tente de les réduire en implantant des techniques de sondage développées à cet effet.

5. Les erreurs de traitement

Des erreurs peuvent se produire lors de toutes les manipulations des données. Les principales étapes du traitement des données sont le codage des réponses, la saisie des données, la vérification, l'imputation des non-réponses partielles et la pondération. Dans l'EDM, différentes procédures sont appliquées à chacune des étapes afin de réduire au minimum les erreurs de traitement et les estimations de l'enquête sont comparées avec d'autres sources avant d'être publiées. Les erreurs qui peuvent découler des ajustements effectués lors de la pondération ont été décrites dans les sections 2 et 3. La présente section couvre principalement les autres types d'erreurs de traitement.

Le codage des réponses est nécessaire pour très peu de questions. Ce travail est effectué par l'intervieweur, puis vérifié par l'intervieweur principal. Avant 2001, la saisie des données était vérifiée à l'aide d'un système de vérification automatisé qui consistait à grouper les questionnaires par lots et à choisir, dans chacun de ces lots, un échantillon de questionnaires qui seraient saisis une seconde fois. Toutes les erreurs relevées étaient corrigées. Si le nombre d'erreurs dépassait un certain seuil, le lot complet était de nouveau soumis à la saisie. Contrairement aux années précédentes aucune procédure de vérification des questionnaires par lot n'a été utilisée pour 2001, suite à l'introduction d'un nouveau système de saisie des données (BLAISE). Toutefois, des règles de vérifications ont été implantées dans ce système afin de pouvoir assurer la cohérence des informations saisies. Les résultats préliminaires d'une étude sur la saisie semblent démontrer que les taux d'erreurs de saisie avec le nouveau système sont similaires à ceux avec l'ancien système.

Une première étape de vérification automatisée des questionnaires est effectuée après que chacun d'eux ait été vérifié manuellement par l'intervieweur et l'intervieweur principal. On s'assure que certaines règles essentielles de cohérence entre les réponses rapportées sur le questionnaire sont respectées. On identifie également les situations inhabituelles qui pourraient justifier des corrections. Cette étape de vérification automatisée est effectuée dans les bureaux régionaux de Statistique Canada, ce qui permet de communiquer avec les répondants lorsque des renseignements supplémentaires sont nécessaires pour résoudre des incohérences dans les réponses qu'ils ont fournies. Les problèmes identifiés au cours de cette vérification sont résolus par les membres des équipes de résolution des vérifications qui ont reçu une formation spéciale à cet égard. Par la suite, d'autres vérifications des données sont faites au bureau central et les réponses invalides sont corrigées.

Le traitement des données de l'EDM comprend également l'imputation de données pour la non-réponse partielle. Il y a non-réponse partielle lorsque le répondant refuse de répondre ou ne connaît pas la réponse à certaines questions particulières. L'approche pour imputer les données diffère selon qu'il s'agit de données catégoriques ou continues. Les données catégoriques peuvent prendre uniquement quelques valeurs (comme les questions pour lesquelles la réponse est oui ou non et les questions sur le type de logement habité) alors que les données continues peuvent prendre n'importe quelle valeur numérique (comme les revenus et les dépenses).

Les données catégoriques qu'on trouve principalement dans les sections sur les caractéristiques et l'équipement du logement, sont imputées à l'aide d'une technique "hot deck" où un ménage donneur est choisi de façon aléatoire parmi un groupe de ménages répondants possédant des caractéristiques semblables.

Les données sur le revenu et les dépenses sont quant à elles imputées au moyen de la technique du plus proche voisin. L'imputation se fait sur un groupe de variables à la fois, ce groupe étant formé de sorte à tenir compte des relations entre ces variables. Un groupe correspond généralement à une section du questionnaire. Pour chaque groupe, les valeurs manquantes d'un receveur (un ménage qui a des données manquantes pour au moins une de ces variables) sont imputées à partir des données du plus semblable parmi tous les donneurs (les ménages qui n'ont aucune donnée manquante pour ces variables). Pour chaque receveur, on détermine le donneur le plus semblable comme étant celui qui a la valeur minimale pour une certaine fonction de distance. Cette fonction est basée sur des variables d'appariement choisies parce qu'elles sont corrélées avec les variables à imputer. Par exemple, le revenu total du ménage fait partie des variables d'appariement pour presque toutes les sections portant sur les dépenses. Dans le choix du donneur, on tient compte également du fait que le receveur doit satisfaire certaines règles de cohérence après l'imputation de ces données manquantes. Pour la plupart des sections, l'imputation est effectuée au niveau du ménage mais dans certains cas, par exemple pour le revenu et les dépenses en vêtements, l'imputation est effectuée au niveau de la personne car c'est à ce niveau que l'on collecte l'information pour ces variables.

Notons qu'en 2001, l'imputation des données sur le revenu et les dépenses a été effectuée avec un nouveau système utilisant une méthodologie quelque peu différente de celle du système utilisé auparavant. Le nouveau système permet une meilleure utilisation des données catégoriques lors de la sélection du donneur. Des tests ont été

effectués avant l'introduction du système et les résultats obtenus étaient semblables à ceux obtenus avec l'ancien système. De plus, les résultats de cette année se comparent à ceux des années précédentes.

Le biais des estimations causé par l'imputation de la non-réponse partielle est difficile à évaluer. Il dépend de la différence entre les répondants et les non-répondants ainsi que de la capacité de la méthode d'imputation à produire des estimations non biaisées. Toutefois, les taux d'imputation donnent une indication de l'importance des non-réponses partielles. Ils sont présentés dans la section suivante.⁸

5.1 Proportion de ménages ou personnes nécessitant imputation, à l'échelle nationale, provinciale et territoriale

Une première indication de l'ampleur de la non-réponse partielle est la proportion de ménages nécessitant imputation et le nombre de variables imputées par ménage. Le questionnaire peut se diviser en deux grands groupes de variables, celles qui sont recueillies au niveau du ménage et celles collectées auprès des individus telles que le revenu et les dépenses en vêtements. Pour ces dernières catégories, il est important de noter que l'on accepte que le répondant fournisse seulement le revenu total ou les dépenses totales s'il est incapable de fournir la ventilation par source de revenu ou type de dépenses. Le niveau d'imputation pour les composantes du revenu et des dépenses vestimentaires est donc plus élevé mais n'affecte pas le revenu total, les dépenses vestimentaires totales ainsi que le total des dépenses.

Le pourcentage de ménages pour lesquels il a fallu imputer des dépenses de ménage (excluant les dépenses vestimentaires) est présenté dans la prochaine sous-section. Dans celle qui la suit, on présente le pourcentage de personnes pour lesquelles il a fallu imputer au moins une variable reliée aux dépenses vestimentaires et le pourcentage de personnes pour lesquelles au moins une variable de revenu a été imputée. À la suite de l'imputation des données par le système, certaines corrections peuvent s'avérer nécessaires, autant sur les variables imputées par le système que sur les autres, afin d'assurer la cohérence des données. Dans les faits, cette année, environ 1 % des ménages et 2 % des personnes ont été modifiés manuellement. Les résultats sont fournis tant à l'échelle nationale que provinciale et territoriale. Cela fournit une indication des provinces et territoires qui sont les plus affectées par l'imputation.

5.1.1 Imputation de dépenses du ménage par province et territoire

Le pourcentage de ménages utilisables pour lesquels il a fallu imputer au moins une variable de dépenses (excluant les dépenses vestimentaires) figure dans le tableau 5.1. Les ménages utilisables correspondent à tous les ménages demeurant dans des logements admissibles, à l'exception de ceux qui n'ont pu être contactés, qui ont refusé de participer à l'enquête, qui ont fourni des données incomplètes ou qui ne balançaient pas (voir définitions dans la section 2.1). Le tableau est réparti en fonction du nombre de variables qui ont été imputées (parmi les 251) pour un ménage.

⁸ Pour des raisons opérationnelles, ces indicateurs de qualité ne sont pas disponibles pour les données catégoriques comme les caractéristiques du logement et l'équipement ménager.

Tableau 5.1
Ménages qui nécessitent l'imputation de dépenses par province et territoire

Province	Ménages (%) qui nécessitent l'imputation de VARIABLES DE DÉPENSES (sauf pour les dépenses en vêtements)			
	Nombre de variables imputées (parmi 251)			TOTAL
	1	2	3 ou plus	
Canada	5,4	1,0	0,8	7,2
Terre-Neuve-et-Labrador	9,9	0,8	0,2	11,0
Île-du-Prince-Édouard	2,9	0,5	0,8	4,1
Nouvelle-Écosse	7,0	1,2	1,2	9,4
Nouveau-Brunswick	3,8	0,6	0,5	4,9
Québec	2,5	0,6	0,4	3,5
Ontario	7,2	2,0	2,0	11,3
Manitoba	4,3	0,7	0,6	5,6
Saskatchewan	6,9	1,0	0,8	8,7
Alberta	3,6	0,7	0,6	4,9
Colombie-Britannique	4,4	1,2	0,8	6,4
Yukon	6,0	0,4	0,4	6,8
Territoires du Nord-Ouest	8,9	2,8	0,3	12,0
Nunavut	9,1	1,0	0,0	10,0

Le tableau 5.1 montre qu'il a fallu imputer des dépenses (mise à part la section sur les vêtements) pour 7,2 % des ménages à l'échelle nationale, mais pour près de 75 % d'entre eux, seulement une variable a été imputée. Il y a très peu de ménages à l'échelle nationale qui ont eu plus d'une variable imputée (1,8 %). Les provinces pour lesquelles on observe les plus basses proportions de ménages ayant requis l'imputation d'au moins une variable de dépense sont le Québec (3,5 %), l'Île-du-Prince-Édouard (4,1 %) et le Nouveau-Brunswick (4,9 %). On retrouve les taux les plus élevés à Terre-Neuve-et-Labrador (11 %), en Ontario (11,3 %) et aux Territoires du Nord-Ouest (12 %). Le faible pourcentage de ménages pour lesquels des variables doivent être imputées, combiné à un nombre généralement peu élevé de variables à imputer lorsque l'imputation est nécessaire, suggère que l'impact des valeurs imputées sur les estimations ne devrait pas être trop élevé.

5.1.2 Imputation de dépenses vestimentaires et de revenu par province et territoire

Étant donné qu'une partie des répondants fournissent uniquement le total pour les variables de dépenses en vêtements et les variables de revenu, celles-ci sont imputées, au niveau des individus, en deux étapes. Les individus pour lesquels il faut imputer seulement quelques composantes sont imputés en premier lieu, suivis par ceux pour lesquels seulement les totaux sont disponibles et qui nécessitent l'imputation de toutes les composantes. (Voir la référence [1] pour une description plus détaillée de ce processus).

Le pourcentage d'individus utilisables (personnes membres des ménages utilisables) pour lesquels il a fallu imputer au moins une variable de revenu est indiqué par province et territoire dans le tableau 5.2. On y retrouve le pourcentage de personnes pour lesquelles exactement une variable a été imputée, ceux dont deux variables ou plus (mais pas toutes) ont été imputées et le pourcentage des personnes pour lesquelles seul le revenu total était disponible, nécessitant ainsi l'imputation de toutes les

composantes. Le pourcentage total de personnes pour lesquelles une forme d'imputation de revenu a été effectuée est également donné. La dernière colonne du tableau 5.2 contient ce même pourcentage total de personnes pour lesquelles une forme d'imputation a été effectuée, mais pour les dépenses en vêtements.

Tableau 5.2
Personnes qui nécessitent l'imputation du revenu et personnes qui nécessitent l'imputation de dépenses en vêtements par province et territoire

Province	Pourcentage de personnes qui nécessitent l'imputation de VARIABLES DE REVENU				Pourcentage de personnes qui nécessitent l'imputation d'au moins une des 15 VARIABLES DE DÉPENSES VESTIMENTAIRES
	1 variable imputée	2 variables imputées ou plus (pas toutes)	Toutes les variables imputées (revenu total connu)	TOTAL (toute forme d'imputation)	
Canada	0,5	0,2	2,7	3,5	12,6
Terre-Neuve-et-Labrador	0,2	0,1	1,7	2,0	14,8
Î.-P.-É.	0,0	0,1	5,1	5,2	13,3
N.-É.	0,5	0,3	3,1	3,9	15,2
N.-B.	0,2	0,2	5,9	6,4	13,1
Québec	0,3	0,1	1,4	1,8	9,8
Ontario	1,3	0,5	2,4	4,2	13,5
Manitoba	1,2	0,2	2,3	3,7	13,2
Saskatchewan	0,4	0,2	2,0	2,5	11,9
Alberta	0,2	0,3	3,2	3,6	15,1
C.-B.	0,6	0,3	2,9	3,8	9,8
Yukon	0,3	0,0	1,2	1,5	12,8
T.N.-O.	0,0	0,0	1,8	1,8	8,5
Nunavut	0,1	0,0	0,1	0,2	7,6

À partir de ces résultats, on constate qu'il y a moins de 4 % des personnes des ménages utilisables pour lesquelles il a fallu imputer au moins une variable de revenu. Pour environ 80 % d'entre elles, le revenu total était disponible mais toutes les composantes ont dû être imputées. Pour une très grande partie des autres personnes qui ont nécessité imputation, seulement une composante de revenu (une variable) a été imputée. À l'échelle provinciale et territoriale, le pourcentage de personnes nécessitant l'imputation d'au moins une variable de revenu est également peu élevé, variant entre un minimum de 0,2 % pour le Nunavut et un maximum de 6,4 % pour le Nouveau-Brunswick.

Dans la dernière colonne du tableau, on constate qu'environ 13 % des personnes ont nécessité une imputation pour au moins une des variables de dépenses vestimentaires. Les taux provinciaux et territoriaux s'échelonnent de 7,6 % pour le Nunavut à environ 15 % (Terre-Neuve-et-Labrador, Nouvelle-Écosse et Alberta). Presque toutes ces personnes ont fourni le total de leurs dépenses vestimentaires, mais il a fallu imputer les composantes. Le niveau élevé d'imputation effectuée sur les composantes des dépenses vestimentaires implique que les estimations de ces composantes pourraient être grandement affectées par l'imputation, alors que l'effet sur les estimations du total sera négligeable.

Références

- [1] Tremblay J. et Arsenault, S. 2001. *La méthodologie de l'Enquête sur les dépenses des ménages*. Ottawa. Division des méthodes d'enquêtes auprès des ménages, Statistique Canada.
- [2] Wolter, K.M. 1985. *Introduction to Variance Estimation*. New York. Springer-Verlag.
- [3] Statistique Canada, Division de la statistique de revenu. 2001. *Guide de l'utilisateur — Enquête sur les dépenses des ménages*. N° 62F0026MIF2002002 au catalogue. Ottawa.
- [4] Statistique Canada, Division de la statistique de revenu. 2001. *Les habitudes de dépenses au Canada*. N° 62-202-XPF au catalogue. Ottawa.
- [5] Statistique Canada, Division des méthodes d'enquêtes auprès des ménages. 2001. *Enquête sur la population active, Rapport sur les opérations*. Enquête 200112. Ottawa.

Annexe A

Notation algébrique

1. Ajustement pour la non-réponse

Le sous-poids (c-à-d. le poids de sondage ajusté pour la non-réponse) d'un ménage k , noté w_k^{NR} , est

$$w_k^{NR} = \pi_k^{-1} * \frac{1}{\text{taux}_g} \quad \text{avec} \quad \text{taux}_g = \frac{\sum_{k \in S_{g,r}} \pi_k^{-1}}{\sum_{k \in S_{g,r}} \pi_k^{-1} + \sum_{k \in S_{g,nr}} \pi_k^{-1}}$$

où

$S_{g,r}$ est l'ensemble des ménages répondants dans le groupe de non-réponse g ,

$S_{g,nr}$ est l'ensemble des ménages non-répondants (refus, non-contacts, données inutilisables) dans le groupe de non-réponse g , et

B_k^{-1} est le poids de sondage attribué au ménage k .

2. Calcul du taux de glissement

Le taux de glissement pour un groupe de contrôle c , noté taux_c , est,

$$\text{taux}_c = 100 * \frac{\left(\sum_{k \in S_{c,r}} w_k^{NR} \right) - t_c}{t_c}$$

où

$S_{c,r}$ est l'ensemble des ménages répondants dans le groupe de contrôle c ,

w_k^{NR} est le sous-poids du ménage k , et

t_c est le total de la donnée auxiliaire pour le groupe de contrôle c .