



N° 62F0026MIF au catalogue — N° 006
ISSN : 1708-8887
ISBN : 0-662-70279-4

Document de recherche

Série de documents de recherche sur les dépenses des ménages

Enquête sur les dépenses des ménages de 2003 : Indicateurs de qualité des données

par Yi Li, Charles Mitchell, Sylvain Nadon et Paddison Wong

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

Téléphone : 613 951-7355



Statistique
Canada

Statistics
Canada

Canada

Comment obtenir d'autres renseignements

Toute demande de renseignements au sujet du présent produit ou au sujet de statistiques ou de services connexes doit être adressée à : 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 services de dépôt	1 800 700-1033
Télécopieur pour le Programme des services de dépôt	1 800 889-9734
Renseignements par courriel	infostats@statcan.ca
Site Web	www.statcan.ca

Renseignements pour accéder au produit

Le produit n° 62F0026MIF au catalogue est disponible gratuitement. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.ca et de choisir la rubrique Nos produits et services.

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. Les normes de service sont aussi publiées dans le site www.statcan.ca sous À propos de Statistique Canada > Offrir des services aux Canadiens.



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 2003 : Indicateurs de qualité des données

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2005

Tous droits réservés. Le contenu de la présente publication peut être reproduit, en tout ou en partie, et par quelque moyen que ce soit, sans autre permission de Statistique Canada sous réserve que la reproduction soit effectuée uniquement à des fins d'étude privée, de recherche, de critique, de compte rendu ou en vue d'en préparer un résumé destiné aux journaux, et/ou à des fins non commerciales. Statistique Canada doit être cité comme suit : Source (ou « Adapté de », s'il y a lieu) : Statistique Canada, nom du produit, numéro au catalogue, volume et numéro, période de référence et page(s). Autrement, il est interdit de reproduire quelque contenu de la présente publication, ou de l'emmagasiner dans un système de recouvrement, ou de le transmettre sous quelque forme et par quelque moyen que ce soit, reproduction électronique, mécanique, photographique, pour quelque fin que ce soit, sans l'autorisation écrite préalable des Services d'octroi de licences, Division du marketing, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Octobre 2005

N° 62F0026MIF2005006

ISSN : 1708-8887

ISBN : 0-662-70279-4

Périodicité : hors série

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) :

Yi Li
Charles Mitchell
Sylvain Nadon
Paddison Wong

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	6
Introduction	9
1. Les erreurs d'échantillonnage	10
1.1 Les mesures de l'erreur d'échantillonnage.....	10
1.2 Les coefficients de variation	11
1.3 Modèle pour dériver une approximation du CV	14
1.4 Suppression des données non fiables dans les tableaux d'estimations.....	15
2. La non-réponse	15
2.1 Le taux de réponse, les taux de non-réponse et les taux de vacance	16
2.2 La non-réponse selon le niveau d'urbanisation	19
2.3 La non-réponse selon les strates de revenu	20
2.4 L'ajustement pour la non-réponse	21
3. Erreurs de couverture.....	22
3.1 Le sous-dénombrement et le surdénombrement : les taux de glissement.....	22
3.2 L'ajustement au niveau de la population et des ménages	27
4. Les erreurs de réponse	28
5. Les erreurs de traitement	30
5.1 Proportion de ménages ou personnes nécessitant imputation, à l'échelle nationale, provinciale et territoriale.....	32
Références.....	38
Annexe A Notation algébrique.....	39

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 2,3 % selon les provinces. Le CV à l'échelle nationale est de 0,7 %. Les CV sont plus élevés dans les territoires, soit 3,6 % au Yukon, 3,9 % dans les Territoires du Nord-Ouest et 5,9 % au Nunavut.
- Les coefficients de variation (CV) de l'estimation de la moyenne pour les différentes catégories de dépenses sommaires sont pour la plupart inférieurs ou égaux à 1,7 % à l'échelle nationale, et généralement inférieurs à 5,0 % à l'échelle provinciale. Les résultats pour les différentes catégories de caractéristiques des logements et l'équipement ménager sont similaires. La taille de l'échantillon étant moindre dans les territoires, les CV ont tendance à être plus élevés que ceux des provinces.

La non-réponse

- Le taux de réponse final est de 72,3 %. Les taux de réponse provinciaux et territoriaux s'échelonnent de 62,8 % en Ontario à 91,6 % au Nunavut.
- Le taux de non-réponse est de 27,7 %. La non-réponse est due aux refus (17,7 %), aux ménages qu'on n'a pas réussi à contacter (6,9 %) et aux ménages dont les données ont été considérées inutilisables (3,0 %).
- Le taux de non-réponse final a tendance à croître avec le niveau d'urbanisation. On observe un taux de non-réponse de 22,3 % en milieu rural, et un taux de 33,8 % 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 (42,2 %) est plus élevé que ceux observés dans les strates régulières (27,3 %) et dans les strates à faible revenu (27,3 %). Le taux de refus des strates à revenu élevé (30,5 %) est plus du double de celui des strates à faible revenu (13,3 %) et près du double de celui des strates régulières (17,4 %).

Les erreurs de couverture

- On observe un sous-dénombrement des ménages de 8,5 % à l'échelle nationale. On observe aussi un sous-dénombrement des ménages pour chacune des provinces et chacun des territoires, les taux variant de 4,9 % à 14,9 %. C'est au Yukon et en Colombie-Britannique que le sous-dénombrement des ménages est le plus élevé.

- On observe un sous-dénombrement des personnes de 11,3 % à l'échelle nationale. On observe aussi un sous-dénombrement des personnes pour chacune des provinces et chacun des territoires, les taux variant de 6,6 % à 22,7 %. C'est au Yukon et aux Territoires du Nord-Ouest que le sous-dénombrement des personnes est le plus élevé.
- À l'échelle nationale, on constate qu'il y a toujours un sous-dénombrement net quelque soit le groupe d'âge. À l'exception de la catégorie des 65 à 69 ans, les taux de sous-dénombrement chez les enfants (0 à 6 ans et 7 à 17 ans) sont inférieurs à ceux observés pour les adultes. Le taux de sous-dénombrement pour l'ensemble des enfants est de 5,5 %, alors qu'il est de 13,0 % 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'Enquête sur les dépenses des ménages (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 d'une heure et quarante-six 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 et les dépenses de la section sur l'impôts personnels, sécurité et dons en argent) pour 12,8 % des ménages. Pour la majorité d'entre eux, seulement une ou deux variables parmi les 229 variables de dépenses ont été imputées. À l'échelle provinciale et territoriale, les taux d'imputation varient entre un minimum de 6,6 % pour le Québec et un maximum de 18,2% pour Terre-Neuve-et-Labrador. En excluant les paiements hypothécaires réguliers et les primes d'assurance hypothécaire, le taux d'imputation est de 8,3 %, ce qui est comparable aux taux obtenus les années précédentes. À elle seule, la variable représentant les primes d'assurance hypothécaire nécessite une imputation pour 15,6 % des ménages ayant déclaré des hypothèques sur les logements dont ils étaient propriétaires et occupants.
- Près de 19 % 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.
- Environ 3 % des personnes âgées de 15 ans et plus ont nécessité l'imputation d'au moins une des variables de la section sur l'impôts personnels, sécurité et dons en argent.

ii) Des variables de revenu

- Il a fallu imputer au moins une variable de revenu pour moins de 5 % des individus âgés de 15 ans et plus. Pour 70 % d'entre eux, le revenu total était fourni par le répondant mais toutes les composantes (sources de revenu) ont été imputées.

iii) Des variables catégoriques

- Il a fallu imputer au moins une variable catégorique pour 6,4 % des ménages. Pour près de 75 % d'entre eux, seulement une variable a été imputée. À l'échelle provinciale et territoriale, les taux d'imputation varient entre un minimum de 2,8 % pour l'Île-du-Prince-Édouard et un maximum de 10,7 % pour la Saskatchewan. Les variables catégoriques qui sont imputées se trouvent aux sections suivantes du questionnaire: Caractéristiques du logements; Équipement associé au logement; Mode d'occupation du logement ; et Tabac et dépenses diverses, pour les variables se rapportant aux achats par le biais de vente directe.

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 2003 est composé de 23 869 ménages¹ 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. À la suite d'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 2003. 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 réponse et 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. Au départ, l'échantillon compte 27 671 logements desquels on doit identifier et enlever les logements inadmissibles (voir section 2.1) pour en arriver à 23 869 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 bootstrap. Jusqu'à l'année de référence 2002, l'EDM utilisait la méthode de rééchantillonnage du jackknife afin de produire les estimations de l'erreur type. Il a été décidé d'utiliser la méthode du bootstrap à compter de l'EDM de 2003 principalement en raison du fait que la Division de la statistique du revenu (DSR) allait diffuser les médianes des dépenses et requérait donc les coefficients de variation de celles-ci. La méthode du bootstrap permet l'estimation de la variance de statistiques non lisses comme les quantiles. 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.

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

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]).

On doit noter que les estimations de CV ne tiennent pas compte du fait que certaines données ont été imputées; par conséquent, les CV présentés peuvent sous-estimer les valeurs réelles. Pour la plupart des variables, le taux d'imputation est peu élevé (voir section 5) et les CV fournis représentent une bonne estimation des vrais CV. Cependant, il est important de tenir compte à la fois du CV et du taux d'imputation si on s'intéresse à la fiabilité de dépenses détaillées ayant un taux d'imputation élevé.

Le tableau 1.1 donne un aperçu des CV des estimations des moyennes par ménage à l'échelle provinciale et territoriale ainsi qu'à l'échelle nationale pour l'estimation de quelques-unes des catégories de dépenses sommaires et pour le revenu.

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

Catégories de dépenses sommaires	Can.	T.-N.-L.	Î.-P.-É.	N.-É.	N.-B.	Qc.	Ont.	Man.	Sask.	Alb.	C.-B.	Yn.	T.N.-O.	Nt.
Dépenses totales	0,7	1,6	2,3	1,6	1,2	1,7	1,4	1,4	1,5	1,3	1,5	3,6	3,9	5,9
Dépenses de consommation	0,7	1,5	2,2	1,4	1,3	1,3	1,3	1,5	1,5	1,4	1,4	3,4	3,9	5,8
Alimentation	0,6	1,3	2,2	1,3	1,3	0,9	1,2	1,5	1,6	1,4	1,4	3,8	3,5	2,9
Logement	0,9	2,3	2,7	1,6	1,8	1,4	1,7	2,1	1,5	1,9	1,6	4,1	5,1	7,2
Entretien ménager	1,0	1,6	3,5	2,6	2,0	1,6	2,1	1,9	1,9	2,3	2,1	4,7	4,1	7,1
Ameublement	1,7	3,5	5,5	4,1	3,4	3,4	3,3	3,8	4,3	3,5	4,2	10,5	9,2	12,4
Habillement	1,3	2,5	3,5	3,0	2,7	2,0	2,5	2,5	3,1	2,8	3,4	10,1	6,5	8,2
Transport	1,5	3,7	4,7	3,6	3,0	3,3	2,8	4,6	3,7	3,7	3,3	7,2	10,4	19,7
Soins de santé	1,3	3,0	3,5	3,7	3,2	2,0	3,3	4,0	3,8	2,5	2,8	7,8	9,9	19,7
Soins personnels	1,0	2,2	3,6	2,1	2,4	1,6	2,0	2,2	2,2	2,8	2,5	9,9	4,6	8,2
Loisirs	1,7	3,8	6,5	3,1	4,5	3,0	3,4	3,5	3,7	3,3	3,9	7,7	7,3	9,2
Lecture et autres imprimés	1,7	4,3	5,1	3,8	4,0	2,9	3,6	4,0	3,3	4,1	3,7	7,6	8,0	19,0
Éducation	3,0	7,5	11,0	8,3	8,6	4,3	5,4	7,5	6,9	7,3	6,7	12,4	19,6	29,7
Tabac, boissons alcoolisées	1,6	3,6	6,9	4,1	4,8	2,7	3,5	4,8	4,0	4,5	4,5	8,4	8,2	8,1
Jeux de hasard (net)	4,1	6,4	13,8	8,0	9,6	6,3	7,7	8,2	7,4	13,3	13,4	14,4	24,3	20,4
Dépenses diverses	2,9	6,0	28,1	5,1	5,4	4,2	6,0	6,0	6,4	8,2	6,5	10,9	7,6	17,2
Impôts personnels	1,7	3,7	4,5	2,8	2,2	3,4	3,2	2,6	2,9	2,5	2,6	6,3	5,2	9,7
Assurances individuelles et cotisations de retraite	1,2	3,1	2,9	9,5	1,9	2,5	2,3	3,2	2,2	2,1	2,0	6,9	5,0	6,7
Dons et contributions	6,0	8,3	14,9	9,3	7,9	9,3	10,8	11,5	6,3	7,6	18,6	15,1	15,1	16,5
Revenu	0,7	1,4	2,2	1,5	1,1	1,6	1,5	1,3	1,5	1,1	1,4	3,9	4,6	6,0

Les coefficients de variation de l'estimation de la moyenne des dépenses totales par ménage varient entre 1,2 % et 2,3 % selon les provinces et il est de 0,7 % à l'échelle nationale. Les CV sont plus élevés dans les territoires, soit 3,6 % au Yukon, 3,9 % dans les Territoires du Nord-Ouest et de 5,9 % au Nunavut.

Pour les différentes catégories de dépenses sommaires, les CV à l'échelle nationale sont inférieurs ou égaux à 1,7 % à l'exception des dépenses dans les catégories suivantes: éducation, jeux de hasard, dépenses diverses ainsi que dons en argent et contributions. Ces catégories de dépenses représentent respectivement 1,6 %, 0,4 %, 1,5 % et 2,5 % du total des dépenses (données non présentées). De plus, si on fait exception de ces catégories, les CV sont généralement inférieurs ou égaux à 5 % à l'échelle provinciale. À l'exception de ces mêmes catégories, les CV dans les territoires sont pour la plupart inférieurs à 11 %. Également, la taille de l'échantillon étant moindre dans les territoires, les CV ont tendance à être plus élevés que ceux des provinces.

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	1,7	3,3	2,0	1,8	1,7	1,9	1,9	1,8	1,9	1,9	6,6	10,9	15,1
Locataire	1,7	5,6	8,6	5,0	5,2	2,2	4,0	4,7	5,0	4,8	3,3	10,8	16,4	8,5
Machine à laver	0,6	0,9	2,1	1,5	1,3	0,8	1,5	1,6	1,2	1,1	1,3	4,7	5,3	4,9
Sècheuse	0,6	1,1	2,3	1,5	1,3	0,9	1,6	1,6	1,1	1,2	1,5	4,4	7,0	5,5
Lave-vaisselle	1,1	3,8	4,7	3,5	3,4	1,9	2,5	2,7	2,6	2,0	1,8	7,2	8,0	12,4
Congélateur	1,0	1,4	3,0	2,3	1,8	2,0	2,1	1,7	1,5	2,3	2,2	5,6	6,1	7,9
Four à micro-ondes	0,3	0,6	1,1	0,7	0,6	0,6	0,6	0,7	0,6	0,6	1,0	2,1	4,3	4,7
Téléphone cellulaire	1,1	2,9	5,2	2,8	3,7	2,4	2,0	2,8	2,8	2,0	2,1	12,2	8,0	24,6
Lecteur de disques compacts	0,6	1,5	2,5	1,5	1,7	1,2	1,3	1,7	1,6	1,3	1,4	4,4	3,5	6,0
Télédistribution	1,0	2,4	5,0	2,7	2,9	1,9	2,0	2,7	3,0	2,6	1,6	10,4	6,2	6,8
Antenne parabolique	2,4	5,0	7,8	5,6	5,0	4,2	5,5	4,7	4,4	4,8	5,7	11,7	12,0	16,4
Lecteur DVD	1,1	2,9	4,7	2,4	2,9	2,1	2,2	2,3	2,6	2,1	2,6	5,4	5,0	8,2
Ordinateur personnel	0,8	2,5	3,8	2,0	2,4	1,6	1,5	2,0	1,8	1,6	1,6	4,8	6,8	7,9
Connexion tél. régulière à un ordinateur (Modem)	2,4	5,4	9,5	6,2	4,6	4,1	4,4	7,0	7,1	7,4	6,8	15,3	17,4	11,2
Connexion tél. haute vitesse à un ordinateur	2,7	9,4	12,0	5,6	8,5	5,1	5,7	6,7	5,2	5,3	5,5	9,2	12,8	40,0
Utilisation d'Internet (maison)	0,9	3,2	4,9	2,4	2,8	2,0	1,9	2,7	2,4	2,1	2,1	6,7	7,0	10,8
Véhicules possédés (un)	1,3	2,8	6,2	3,7	3,4	2,4	2,9	3,1	3,3	3,1	2,9	9,6	5,6	14,3
Véhicules possédés (2 ou plus)	1,4	4,4	4,9	2,9	3,5	2,9	3,1	2,8	3,2	2,4	2,7	7,1	10,2	22,3

Les coefficients de variation pour les caractéristiques des logements et l'équipement ménager sont généralement inférieurs à 5 % à l'échelle provinciale sauf pour les catégories : *locataire*, *antenne parabolique*, *connexion téléphonique régulière à un ordinateur (Modem)* et *connexion téléphonique haute vitesse à un ordinateur*. Les CV sont plus élevés dans les territoires où l'on retrouve parfois une proportion moindre des équipements. En outre, la taille de l'échantillon étant moindre dans les territoires, les CV ont tendance à être plus élevés que ceux des provinces.

Les CV pour les caractéristiques des logements et l'équipement ménager à l'échelle nationale sont inférieurs ou égaux à 1,7 % à l'exception des catégories suivantes : *antenne parabolique*, *connexion téléphonique régulière* et *connexion téléphonique haute vitesse à un ordinateur*. On retrouve une proportion moindre

de ménages possédant de tels équipements pour ces trois catégories. Cette proportion représente respectivement 22,6 %, 21,2 % et 17,3 % à l'échelle nationale (données non présentées).

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 2003 (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 2003. 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

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

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.

2.1 Le taux de réponse, 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.

Il est à noter que tous les taux de la présente section sont non pondérés. Pour l'Enquête sur les dépenses des ménages de 2003, le taux de réponse final est de 72,3 %. Le tableau 2.1a présente le taux de réponse final ainsi que la taille de l'échantillon (ménages admissibles) ventilée selon les refus, les non-contacts, les données inutilisables et les données utilisables. Ce taux est fourni tant à l'échelle nationale que provinciale et territoriale.

Cette année, un échantillon supplémentaire d'environ 2 000 ménages au Québec a été financé par le Ministère des finances du Québec afin de mieux répondre aux besoins analytiques de celui-ci. Cet échantillon a été sélectionné d'une liste de logements selon un plan stratifié à deux degrés d'échantillonnage. La base liste, qui est construite à partir du Registre des adresses de Statistique Canada, est constituée des logements d'un certain nombre d'aires de diffusion du Recensement de 2001. Pour une description détaillée du plan de sondage de l'échantillon supplémentaire, on peut consulter la référence [6].

Tableau 2.1a
Taille de l'échantillon et taux de réponse (%) par province et territoire ainsi qu'à l'échelle nationale

Province ou territoire	Ménages admissibles	Non contactés	Ayant refusé de participer	Inutilisables	Utilisables	Taux de réponse final (à l'estimation) ¹ %
Canada	23 869	1 647	4 234	723	17 265	72,3
Terre-Neuve-et-Labrador	1 772	134	198	69	1 371	77,4
Île-du-Prince-Édouard	781	52	121	27	581	74,4
Nouvelle-Écosse	2 048	150	347	132	1 419	69,3
Nouveau-Brunswick	1 845	98	303	118	1 326	71,9
Québec	4 817	360	927	68	3 462	71,9
Ontario	3 149	282	787	103	1 977	62,8
Manitoba	1 887	104	312	49	1 422	75,4
Saskatchewan	1 837	72	280	77	1 408	76,6
Alberta	2 087	128	336	18	1 605	76,9
Colombie-Britannique	2 580	205	500	45	1 830	70,9
Yukon	418	28	74	9	307	73,4
Territoires du Nord-Ouest	410	21	42	8	339	82,7
Nunavut	238	13	7	-	218	91,6

1. Utilisables/ménages admissibles x 100

Le tableau 2.1b 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.

Notons que les taux de vacance présentés aux tableaux de la section 2 incluent les logements vacants (inoccupés, saisonniers ou en construction) ainsi que les logements qui n'existent plus (démoli, maison mobile déplacée ou logement converti en entreprise).

Tableau 2.1b
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 équilibré	
Canada	11,0	24,6	6,9	17,7	3,0	1,3	1,7	27,7
Terre-Neuve-et-Labrador	15,4	18,7	7,6	11,2	3,9	1,8	2,1	22,6
Île-du-Prince-Édouard	18,1	22,2	6,7	15,5	3,5	1,7	1,8	25,6
Nouvelle-Écosse	13,8	24,3	7,3	16,9	6,4	1,2	5,2	30,7
Nouveau-Brunswick	13,7	21,7	5,3	16,4	6,4	1,9	4,5	28,1
Québec	7,3	26,7	7,5	19,2	1,4	1,2	0,2	28,1
Ontario	8,2	33,9	9,0	25,0	3,3	1,9	1,4	37,2
Manitoba	9,6	22,0	5,5	16,5	2,6	0,7	1,9	24,6
Saskatchewan	15,7	19,2	3,9	15,2	4,2	1,3	2,9	23,4
Alberta	9,0	22,2	6,1	16,1	0,9	0,7	0,1	23,1
Colombie-Britannique	8,9	27,3	7,9	19,4	1,7	1,4	0,4	29,1
Yukon	15,3	24,4	6,7	17,7	2,2	0,2	1,9	26,6
Territoires du Nord-Ouest	12,2	15,4	5,1	10,2	2,0	0,2	1,7	17,3
Nunavut	14,8	8,4	5,5	2,9	0,0	0,0	0,0	8,4

Le taux de non-réponse final au Canada est de 27,7 %. Il est dû aux refus (17,7 %), aux ménages qu'il a été impossible de rejoindre (6,9 %), et finalement aux ménages dont les données étaient inutilisables (3,0 %). 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 le Nunavut pour lequel le taux de refus est inférieur au taux de non-contacts ainsi que les provinces du Nouveau-Brunswick et de la Saskatchewan pour lesquelles la proportion de données inutilisables est supérieure à la proportion de non-contacts.

Les taux de non-réponse finaux varient selon les provinces. C'est au Nunavut que l'on observe le taux de non-réponse le plus bas, soit de 8,4 %. C'est aussi dans ce territoire que la proportion de refus est la plus faible (2,9 %). De plus, au Nunavut, toutes les données recueillies étaient utilisables, ce qui veut dire que la non-réponse finale est attribuable seulement à la non-réponse obtenue lors de la collecte. Les taux de non-réponse du Nunavut et des Territoires du Nord-Ouest sont inférieurs à 20 %, tandis que les taux sont supérieurs à 30 % en Nouvelle-Écosse et en Ontario. Le taux de non-réponse final plus élevé en Nouvelle-Écosse est en partie attribuable à un taux plus élevé de questionnaires non équilibrés (5,2 %). Le taux de non-réponse de l'Ontario est particulièrement élevé atteignant 37,2 %. C'est aussi dans cette province que l'on trouve le plus haut taux de non-contacts (9,0 %) et de refus (25,0 %).

Le taux de vacance est présenté au tableau 2.1b, 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 2003 est de 11,0 %.

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 équilibré	
Urbain								
1 000 000 ou plus	4,8	31,8	9,4	22,4	2,0	1,5	0,5	33,8
500 000 à 999 999	5,7	25,3	6,3	19,0	1,6	0,8	0,7	26,9
250 000 à 499 999	5,4	30,1	8,2	21,9	6,2	1,2	5,0	36,3
100 000 à 249 999	7,6	24,0	6,5	17,5	4,5	1,3	3,2	28,5
30 000 à 99 999	6,7	26,4	7,1	19,4	3,4	1,7	1,7	29,8
Moins de 30 000	10,6	20,0	5,8	14,2	2,7	1,2	1,5	22,7
Rural	23,1	18,9	5,4	13,5	3,4	1,3	2,0	22,3
Total	11,0	24,6	6,9	17,7	3,0	1,3	1,7	27,7

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, celui des 100 000 à 249 999 et, dans une moindre mesure, le groupe des 1 000 000 ou plus, qui ne suivent pas cette tendance. Pour le groupe des 500 000 à 999 999, ceci est dû dans une certaine mesure au fait que c'est pour celui-ci que l'on trouve le plus faible taux de données inutilisables (1,6 %). Pour le groupe des 100 000 à 249 999, ceci est dû à un taux de refus (17,5 %) plus faible que pour les autres régions urbaines de plus de 30 000 habitants. Le taux de non-réponse final de 36,3 % de la catégorie d'urbanisation «250 000 à 499 999» est dû à la fois à la non-réponse obtenue lors de la collecte (30,1 %), laquelle est due en partie à celle obtenue en Ontario (38,1 %, donnée non présentée), ainsi qu'au taux de données inutilisables (6,2 %), lequel est dû en partie à celui obtenu en Nouvelle-Écosse (9,8 %, donnée non présentée).

Le taux de non-réponse à la collecte a lui aussi tendance à croître avec le niveau d'urbanisation. Il y a un écart de près de 12 % entre les catégories d'urbanisation

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.

«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 et plus, c'est pour celle-ci que la proportion de non-réponse totale due aux refus des ménages de participer à l'enquête est la plus élevée (70,6 %). C'est dans la catégorie d'urbanisation «1 000 000 ou plus» que l'on retrouve le plus haut taux de non-contacts (9,4 %) et de refus (22,4 %).

En examinant les taux de vacance par niveau d'urbanisation, il ressort que le taux de vacance est deux fois plus élevé en région rurale (23,1 %) que dans les régions urbaines à faible population (10,6 %). 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. Ceci explique entre autre pourquoi le taux de vacance est plus élevé dans les provinces de l'Atlantique, tel qu'il est illustré dans le tableau 2.1b, et surtout à l'Île-du-Prince-Édouard puisqu'on y observe une forte proportion de logements ruraux.

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. Il est à noter que les résultats de ce tableau excluent l'échantillon supplémentaire de la province de Québec. Puisque l'échantillon supplémentaire ne provient pas de la base de sondage de l'EPA, une telle classification de strates (revenu élevé, régulières et revenu faible) n'est pas disponible.

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 ¹	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 équilibré	
Revenu élevé	4,1	39,8	9,4	30,5	2,4	1,5	0,9	42,2
Régulières	11,7	24,1	6,6	17,4	3,2	1,3	1,9	27,3
Revenu faible	3,7	25,2	11,9	13,3	2,1	1,4	0,7	27,3
Total	11,4	24,5	6,7	17,7	3,2	1,3	1,9	27,6

1. Exclut l'échantillon supplémentaire au Québec.

Dans les strates à revenu élevé, le taux de non-réponse final (42,2 %) est de beaucoup supérieur à celui des strates régulières et des strates à faible revenu. Le taux de refus des strates à revenu élevé est de plus de 30 %, ce qui est environ le double de ce qui est observé pour les autres types de strates.

Les ménages des strates à faible revenu ont un taux de non-réponse final identique et un taux de non-réponse à la collecte similaire à ceux des strates régulières. Toutefois les causes de la non-réponse sont différentes pour les deux types de strates. Pour les strates régulières, la non-réponse à la collecte est majoritairement due aux refus tandis que pour les strates à faible revenu, la non-réponse à la collecte se divise à peu près également entre les refus et les non-contacts.

Comme pour les enquêtes de 1997 à 2002, 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 faible revenu, les taux de vacance sont semblables. Le taux de vacance dans les strates à faible revenu est de 3,7 %, en hausse par rapport au taux de 1,8 % observé en 2002.

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 qu'à l'exception de l'échantillon supplémentaire sélectionné au Québec cette année, 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 (à l'exception des Territoires). L'enquête couvre donc près de 98 % de la population dans les 10 provinces. Au Yukon, la couverture de la population est de 88%, celle des Territoires du Nord-Ouest correspond à 92% et la couverture du Nunavut est de 89%.⁵

Nous n'avons pas recueilli de renseignements auprès des personnes qui vivent temporairement loin de leur famille (par exemple, les étudiants d'université) parce que nous obtenons l'information auprès de leur famille, si elle est choisie dans l'échantillon.

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.

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

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

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	-1,0	-6,5	-3,7	
	7-17 ans	-6,0	-6,8	-6,4	
	18-24 ans	-16,9	-14,7	-15,8	
	25-34 ans	-24,2	-15,7	-20,0	
	35-54 ans	-15,0	-10,1	-12,5	
	55-59 ans	-12,2	-8,0	-10,1	
	60-64 ans	-10,1	-6,8	-8,4	
	65-69 ans	-4,5	-5,2	-4,9	
	70 ans et +	-8,9	-7,5	-8,1	
	Total		-12,8	-9,8	-11,3

1. Exclut les territoires.

Pour l'EDM de 2003, le taux de sous-dénombrement de la population était 11,3 %. En analysant le tableau 3.1 on constate qu'à l'échelle nationale, il y a toujours un sous-dénombrement net quelque soit le groupe d'âge. À l'exception de la catégorie des 65 à 69 ans, les taux de sous-dénombrement chez les enfants (0 à 6 ans et 7 à 17 ans) sont inférieurs à ceux observés pour les adultes. Le taux de sous-dénombrement pour l'ensemble des enfants est de 5,5 %, alors qu'il est de 13,0 % pour les adultes (données non présentées). Le léger sous-dénombrement observé pour les garçons de 0 à 6 ans est dû en grande partie au surdénombrement obtenu au Québec (1,7 %) et en Ontario (3,6 %) pour ce groupe d'âge-sexe (voir le tableau 3.2). De la même façon, le taux de sous-dénombrement plus faible pour le groupe d'âge des 65 à 69 ans est dû au surdénombrement des femmes de ce groupe d'âge au Québec (11,9 %) ainsi qu'au surdénombrement des hommes de ce groupe d'âge en Ontario (11,2 %) (voir le tableau 3.2).

Les taux les plus élevés à l'échelle nationale sont observés chez les hommes de 25 à 34 ans, chez ceux de 18 à 24 ans, ainsi que chez les femmes de 25 à 34 ans. À l'exception de la catégorie des 65 à 69 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 10,7 % (référence [5]), ce qui est légèrement inférieur au taux de 12,6 % 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	Age	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	-10,7	9,9	-4,5	2,5	1,7	3,6	-3,5	-17,3	-2,2	-14,3	-0,9	-11,4	-12,7
	7-17	-6,0	-4,6	-6,5	-7,2	4,0	-10,7	-0,9	-12,3	-4,2	-9,6			
	18-24	-35,5	-21,0	-27,4	-21,5	-2,9	-21,2	-30,4	-33,3	-11,5	-20,4			
	25-34	-29,6	-18,7	-27,2	-18,6	-18,1	-30,0	-20,2	-19,0	-16,4	-26,5			
	35-54	-11,4	-12,5	-3,6	-15,7	-9,5	-16,3	-14,6	-22,5	-23,2	-16,4			
	55-59	-20,2	-24,6	-29,4	-11,1	-13,2	-3,9	-26,1	-14,9	-20,0	-18,4			
	60-64	-20,7	-9,4	-9,2	-12,9	-22,0	2,0	-30,6	-5,9	-11,6	-14,2			
	65-69	-0,3	-3,6	-3,8	1,9	-16,3	11,2	-32,5	-0,2	-11,5	-19,6			
	70 +	-2,7	3,8	-19,5	-10,6	-2,4	-13,6	7,8	5,9	-10,1	-12,8			
	Total	-15,5	-10,2	-12,7	-12,6	-7,9	-13,6	-14,0	-16,6	-14,4	-16,9			
Femme	0-6	-10,3	14,3	0,5	9,9	6,3	-13,3	-7,7	-5,6	-10,6	-8,6	-20,4	-25,3	-16,8
	7-17	-9,8	-3,4	-4,0	-10,7	-5,4	-6,0	7,4	-19,8	-3,5	-15,1			
	18-24	-17,1	-20,5	-10,9	-11,5	-6,8	-17,6	-9,4	-19,3	-8,0	-26,6			
	25-34	-18,3	-12,5	-16,3	-17,9	-7,2	-16,9	-22,1	-24,2	-9,7	-27,0			
	35-54	-8,5	-5,0	-0,3	-6,6	-3,9	-12,0	-3,0	-15,4	-17,1	-14,4			
	55-59	-6,9	-12,4	-22,0	-10,3	-17,4	6,5	-22,0	-3,4	-17,9	-18,2			
	60-64	-8,1	0,7	-12,1	-5,4	-14,2	-1,0	-2,3	-1,7	-5,1	-10,6			
	65-69	-5,4	-21,1	-1,5	-8,7	11,9	-11,2	-18,8	-3,9	-25,5	-3,8			
	70 +	-13,0	-7,1	-1,0	-0,5	-8,3	-9,0	12,0	9,1	-11,4	-11,9			
	Total	-11,1	-7,0	-6,0	-7,6	-5,3	-10,5	-4,9	-12,2	-11,9	-16,4			
Total	-13,3	-8,6	-9,3	-10,1	-6,6	-12,1	-9,4	-14,4	-13,1	-16,6	-22,7	-18,7	-13,4	

On observe un sous-dénombrement net pour chacune des provinces et chacun des territoires, les taux variant de 6,6 % à 22,7 %. L'ampleur du sous-dénombrement est plus importante pour les territoires. C'est au Québec que l'on retrouve le plus faible taux de sous-dénombrement (6,6 %). Toutefois, le taux plus faible au Québec est dû dans une certaine mesure à l'échantillon supplémentaire de cette province en 2003. En effet, lorsque l'on omet l'échantillon supplémentaire au Québec, le taux de sous-dénombrement net passe de 6,6 % à 8,8 % (donnée non présentée) pour cette province. 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 (-8,6 %) cache le pire cas de surdénombrement à l'échelle provinciale des groupes d'âge-sexe (14,3 % chez les filles de 0 à 6 ans). En revanche, bien que le taux de glissement général observé en Colombie-Britannique soit de -16,6 %, cette province n'affiche que le douzième pire cas de sous-dénombrement (27,0 % chez les femmes de 25 à 34 ans).

À l'échelle provinciale, le taux de sous-dénombrement le plus élevé est observé chez les hommes de 18 à 24 ans à Terre-Neuve-et-Labrador (35,5 %). Le sous-dénombrement chez les hommes de 25 à 34 ans est plus important. En effet, pour toutes les provinces on observe un taux de sous-dénombrement supérieur à 16,0 %, ce qui explique le taux global de 24,2 % à l'échelle nationale (voir le Tableau 3.1). 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 le Yukon⁷ 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 ¹	-8,5	-2,2	-7,5	-13,2
Terre-Neuve-et-Labrador	-8,1	11,0	-5,1	-17,2
Île-du-Prince-Édouard	-8,3	-9,8	-3,3	-11,3
Nouvelle-Écosse	-6,7	0,6	-8,2	-9,8
Nouveau-Brunswick	-4,9	6,2	3,1	-17,6
Québec	-5,0	-3,8	-3,0	-8,0
Ontario	-9,6	-2,9	-8,6	-13,9
Manitoba	-5,7	5,5	-9,5	-10,5
Saskatchewan	-9,4	2,8	-8,0	-19,4
Alberta	-10,6	-5,7	-14,2	-10,5
Colombie-Britannique	-11,8	-0,3	-9,6	-21,5
Yukon	-14,9	-2,0	-12,7	-26,7

1. Exclut les Territoires du Nord-Ouest et le Nunavut.

On observe une sous-estimation du nombre de ménages de 8,5 % à l'échelle nationale. Ce taux de glissement est légèrement inférieur au taux de glissement pour les personnes qui est de -11,3 %.

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 -9,8 % à l'Île-du-Prince-Édouard jusqu'à 11,0 % pour Terre-Neuve-et-Labrador. Le même phénomène est observable pour les ménages de deux personnes mais avec une variation moindre. Pour les ménages de deux personnes, les taux de glissement varient de -14,2 % pour l'Alberta à 3,1 % pour le Nouveau-Brunswick. Pour les ménages de trois personnes et plus, les taux de glissement varient de -26,7 % pour le Yukon à -8,0 % au Québec.

⁷ Pour l'année de référence 2003, les poids de sondage n'ont pas été ajustés pour refléter les estimations post-censitaires du nombre de ménages aux Territoires du Nord-Ouest et au Nunavut (voir section 3.2).

Notons qu'à l'exception du Nouveau-Brunswick, on observe toujours un sous-dénombrement pour les catégories des ménages de deux personnes et trois personnes et plus. Soulignons que pour les ménages de trois personnes et plus, le sous-dénombrement est généralement supérieur à 10 %. Encore une fois, le taux plus faible au Québec (8,0 %) pour cette dernière catégorie est dû dans une certaine mesure à l'échantillon supplémentaire de cette province en 2003. Lorsque l'on omet l'échantillon supplémentaire au Québec, le taux de sous-dénombrement net de cette catégorie passe de 8,0 % à 10,3 % (donnée non présentée) pour cette province. À l'exception de l'Alberta, le sous-dénombrement le plus élevé est observé pour les ménages de trois personnes et plus.

3.2 L'ajustement au niveau de la population et des ménages

Afin de corriger le problème de représentativité 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 et territoire. Pour plus de détails sur la méthodologie de l'ajustement, voir la référence [1]. Cet ajustement réduit le biais mais ne l'élimine pas complètement si les 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 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, les contrôles selon le nombre de ménages calculés pour l'EDM présentaient un bris dans la série des contrôles qui avaient été produits

jusqu'à ce jour pour les Territoires du Nord-Ouest et le Nunavut. Alors que les contrôles jusqu'en janvier 2001 étaient similaires aux données du recensement, ceux de janvier 2002 s'en écartaient. Afin de diminuer l'impact de cette brisure dans la série sur les données de l'EDM, certains changements avaient été apportés à la méthodologie utilisée pour la production des données auxiliaires sur le nombre de ménages dans ces deux territoires pour la pondération de l'EDM de 2001. Par ailleurs, dans les provinces et au Yukon aucun problème n'apparaît dans la série de contrôles basés sur le recensement de 1996. Quant à l'EDM de 2003, la brisure dans la série de contrôles selon le nombre de ménages existait toujours pour ces deux territoires. Notons que la nouvelle série de contrôle basée sur le recensement de 2001 est disponible et cette dernière série ne présente plus de bris pour ces deux territoires. Toutefois avant d'utiliser la nouvelle série basée sur le recensement de 2001, des évaluations doivent être faites pour déterminer l'impact des nouveaux contrôles sur les estimations de l'enquête. Pour toutes ces raisons et puisque les changements apportés en 2001 ne devaient être qu'une mesure temporaire, il a été décidé de ne pas utiliser les contrôles sur le nombre de ménages au Nunavut et aux Territoires du Nord-Ouest pour la pondération de l'EDM de 2003. Bien que les poids de sondage n'ont pas été ajustés pour tenir compte du nombre de ménages (basé sur le recensement de 1996) aux Territoires du Nord-Ouest et au Nunavut, les estimations du nombre total de ménages obtenues à l'aide des poids finaux sont similaires aux contrôles du nombre de ménages fondés sur le recensement de 2001 pour ces deux territoires. On obtient ainsi un surdénombrement net de 1,3 % aux Territoires du Nord-Ouest et un sous-dénombrement net de 4,0 % au Nunavut.

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. Cependant, en 2002 lorsque les questions additionnelles ont été enlevées, l'estimation pour les dépenses en soins personnels a diminué de nouveau.

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 depuis 2001, à la suite de 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. 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 sur le revenu et les dépenses sont 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 que depuis 2001, l'imputation des données sur le revenu et les dépenses est effectuée à l'aide du Système canadien de contrôle et d'imputation du recensement (SCANCIR) de Statistique Canada. Le nouveau système utilise une méthodologie quelque peu différente de celle du système utilisé auparavant. Ce système permet une meilleure utilisation des données catégoriques lors de la sélection du donneur. De plus, ce système permet à la fois l'imputation des données continues et des données catégoriques. Des tests ont été effectués avant l'introduction du système et les résultats obtenus étaient semblables à ceux obtenus avec l'ancien système. À compter de 2003, les données catégoriques, qu'on trouve principalement dans les sections sur les caractéristiques et l'équipement du logement, sont imputées à l'aide du SCANCIR. Les données catégoriques étaient imputées auparavant 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.

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.

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 et les dépenses de la section sur l'impôts personnels, sécurité et dons en argent) 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, le pourcentage de personnes pour lesquelles au moins une variable de revenu a été imputée et le pourcentage de personnes pour lesquelles il a fallu imputer au moins une variable de la section sur l'impôts personnels, sécurité et dons en argent. Finalement dans la dernière sous-section on présente les résultats pour le pourcentage de ménages pour lesquels il a fallu imputer au moins une des variables catégoriques. 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 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 et les dépenses de la section sur l'impôts personnels, sécurité et dons en argent) 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 229) pour un ménage.

Il importe de préciser que les paiements hypothécaires réguliers ainsi que les primes d'assurance hypothécaire sont inclus au titre des dépenses de logement et donc dans les dépenses totales. Depuis 2002, ces deux variables sont ajoutées au calcul des taux d'imputation présentés au Tableau 5.1. Ceci a pour effet d'augmenter le taux global d'imputation.

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

Province ou territoire	Ménages (%) qui nécessitent l'imputation de variables de dépenses ¹ (sauf pour les dépenses en vêtements et les dépenses de la section sur l'impôts personnels, sécurité et dons en argent)			
	Nombre de variables imputées (parmi 229)			Total
	1	2	3 ou plus	
Canada	9,9	1,8	1,1	12,8
Terre-Neuve-et-Labrador	15,5	2,0	0,7	18,2
Île-du-Prince-Édouard	16,0	1,5	0,0	17,6
Nouvelle-Écosse	13,2	2,5	1,8	17,5
Nouveau-Brunswick	8,4	1,2	0,7	10,3
Québec	5,8	0,5	0,2	6,6
Ontario	11,3	3,2	3,0	17,5
Manitoba	8,9	1,6	0,4	10,9
Saskatchewan	8,7	1,5	0,7	10,9
Alberta	9,3	1,3	0,7	11,3
Colombie-Britannique	9,9	3,2	2,5	15,7
Yukon	13,0	1,3	0,7	15,0
Territoires du Nord-Ouest	12,1	3,2	0,3	15,6
Nunavut	7,3	1,4	0,5	9,2

1. Inclut les paiements hypothécaires réguliers et les primes d'assurance hypothécaire.

Le tableau 5.1 montre qu'il a fallu imputer des dépenses (mise à part la section sur les vêtements et la section sur l'impôts personnels, sécurité et dons en argent) pour 12,8 % des ménages à l'échelle nationale, mais pour près de 77 % d'entre eux, seulement une variable a été imputée. Le taux global d'imputation en excluant les paiements hypothécaires réguliers et les primes d'assurance hypothécaire est de 8,3 % (donnée non présentée), ce qui est comparable aux taux obtenus les années précédentes. À elle seule, la variable représentant les primes d'assurance hypothécaire nécessite une imputation pour 5,3 % des ménages (soit 15,6 % des ménages lorsque l'on se restreint aux ménages ayant déclaré des hypothèques sur les logements dont ils étaient propriétaires et occupants).

Il y a très peu de ménages à l'échelle nationale qui ont eu plus d'une variable imputée (2,9 %). Les provinces ou territoires pour lesquels on observe les plus basses proportions de ménages ayant requis l'imputation d'au moins une variable

de dépense sont le Québec (6,6 %) et le Nunavut (9,2 %). Seuls cette province et ce territoire affichent un taux global d'imputation inférieur à 10 %. On retrouve les taux les plus élevés à Terre-Neuve-et-Labrador (18,2 %), à l'Île-du-Prince-Édouard (17,6 %), en Nouvelle-Écosse (17,5 %) et en Ontario (17,5 %). C'est en Ontario et en Colombie-Britannique où l'on trouve le pourcentage le plus élevé de ménages pour lesquels il a fallu imputer plus d'une variable de dépenses. Dans ces deux provinces, on a imputé au moins deux variables de dépenses pour plus du tiers des ménages qui nécessitaient une imputation. Le faible pourcentage (en particulier si on exclut les paiements hypothécaires réguliers et les primes d'assurance hypothécaire) 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 des dépenses des personnes et du 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. L'avant 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. La dernière colonne contient le pourcentage total de personnes pour lesquelles une forme d'imputation a été effectuée pour les variables de la section du questionnaire portant sur l'impôts personnels, sécurité et dons en argent.

Il importe de préciser que seuls les membres du ménage âgés de 15 ans et plus au 31 décembre de l'année de référence doivent répondre aux questions portant sur le revenu personnel et celles de la section sur les impôts personnels, sécurité et dons en argent. Par conséquent, cette année le pourcentage de personnes qui nécessitent l'imputation de variables de revenu ou pour celles de la section sur l'impôts personnels, sécurité et dons en argent a été calculé en se restreignant aux personnes de 15 et plus et non sur l'ensemble des personnes comme pour les années précédentes. Ceci a pour effet d'augmenter légèrement le taux d'imputation pour ces variables. Tout comme pour les années précédentes, le pourcentage de personnes qui nécessitent l'imputation de dépenses vestimentaires est

calculé sur l'ensemble des personnes, puisque tous les membres d'un ménage doivent répondre aux questions portant sur ce type de dépenses.

Tableau 5.2
Personnes qui nécessitent l'imputation du revenu, personnes qui nécessitent l'imputation de dépenses en vêtements et personnes qui nécessitent l'imputation de variables de la section sur l'impôts personnels, sécurité et dons en argent par province et territoire

Province ou territoire	Pourcentage de personnes qui nécessitent l'imputation de variables de revenu				Pourcentage de personnes qui nécessitent l'imputation d'au moins une des 11 variables de dépenses vestimentaires	Pourcentage de personnes qui nécessitent l'imputation d'au moins une des 15 variables de la section sur l'impôts personnels, sécurité et dons en argent
	1 variable imputée	2 variables imputées ou plus (pas toutes)	Toutes les variables imputées (revenu total connu)	Total (toute forme d'imputation)		
Canada	1,0	0,3	3,3	4,7	18,6	3,2
T.-N.-L.	0,5	0,1	2,9	3,7	7,4	1,9
Î.-P.-É.	0,1	0,2	8,1	8,6	12,5	1,0
N.-É.	0,6	0,1	3,9	4,7	15,2	5,1
N.-B.	0,5	0,2	6,3	7,1	14,8	2,4
Qc	0,6	0,1	1,9	2,6	19,9	1,2
On.	2,3	0,7	3,2	6,2	16,3	6,8
Man.	1,6	0,4	2,1	4,2	19,1	2,4
Sask.	0,4	0,6	3,0	4,2	24,5	3,0
Alb.	0,3	0,3	3,1	3,8	20,9	2,3
C.-B.	2,4	0,4	4,1	7,0	20,0	5,6
Yn	0,5	0,4	4,4	5,3	40,9	4,2
T.N.-O.	0,7	0,0	1,3	2,0	34,1	1,3
Nt	0,0	0,0	0,8	0,8	26,3	0,8

À partir de ces résultats, on constate qu'il y a près de 5 % des personnes des ménages utilisables pour lesquelles il a fallu imputer au moins une variable de revenu. Pour 70 % 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. Rappelons que cette année les taux d'imputation pour les variables de revenu sont calculés en se restreignant aux personnes de 15 ans et plus. Le taux global d'imputation calculé sur l'ensemble des personnes est de 3,8 % (donnée non présentée), ce qui se compare au taux de 4,1 % obtenu en 2002. À l'échelle provinciale et territoriale, le pourcentage de personnes nécessitant l'imputation d'au moins une variable de revenu varie entre un

minimum de 0,8 % pour le Nunavut et un maximum de 8,6 % pour l'Île-du-Prince-Édouard. On observe un taux d'imputation plus élevé à l'Île-du-Prince-Édouard (8,6 %), au Nouveau-Brunswick (7,1 %), en Colombie-Britannique (7,0 %) et en Ontario (6,2 %). Pour l'Île-du-Prince-Édouard et le Nouveau-Brunswick, la majorité des cas à imputer correspond à la situation où le total du revenu est fourni par le répondant mais les toutes composantes doivent être imputées. Pour l'Ontario et la Colombie-Britannique, l'imputation se divise principalement entre le cas où il n'y a qu'une seule composante à imputer et le cas où total du revenu est fourni par le répondant mais toutes les composantes doivent être imputées.

Dans l'avant dernière colonne du tableau, on constate qu'environ 19 % des personnes ont nécessité une imputation pour au moins une des variables de dépenses vestimentaires. Les taux provinciaux et territoriaux s'échelonnent de 7,4 % pour Terre-Neuve-et-Labrador à 40,9 % au Yukon. Presque toutes ces personnes ont fourni le total de leurs dépenses vestimentaires, mais il a fallu imputer les composantes. Terre-Neuve-et-Labrador est la seule province où le taux d'imputation est inférieur à 10 %. On observe également que ces taux sont plus faibles dans les provinces de l'Atlantique et à l'opposé, plus élevés dans les territoires que pour les autres provinces. 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.

Dans la dernière colonne du tableau, on constate qu'environ 3 % des personnes ont nécessité une imputation pour au moins une des variables de la section sur l'impôts personnels, sécurité et dons en argent. Rappelons que cette année le taux d'imputation pour les variables de cette section est calculé en se restreignant aux personnes de 15 ans et plus. Le taux global d'imputation calculé sur l'ensemble des personnes est de 2,6 % (donnée non présentée), ce qui est identique au taux obtenu en 2002. À l'échelle provinciale et territoriale, ce pourcentage est également peu élevé, variant entre un minimum de 0,8 % au Nunavut et un maximum de 6,8 % pour l'Ontario. Seules les provinces de la Nouvelle-Écosse, de l'Ontario et de la Colombie-Britannique affichent un taux supérieur à 5 %. Dans ces trois provinces, le taux d'imputation plus élevé est dû à la variable représentant l'impôt personnel sur le revenu de 2003 qui nécessite en proportion plus d'imputation que pour les autres provinces et territoires.

5.1.3 Imputation des variables catégoriques par province et territoire

Le pourcentage de ménages utilisables pour lesquels il a fallu imputer au moins une variable catégorique figure dans le tableau 5.3. Le tableau est réparti en fonction du nombre de variables qui ont été imputées (parmi les 56) pour un ménage. Les variables catégoriques qui sont imputées se trouvent aux sections suivantes du questionnaire: Caractéristiques du logements (à l'exception de la variable genre de logement); Équipement associé au logement; Mode d'occupation du logement (à l'exception des variables se rapportant au changement dans le mode d'occupation durant l'année de référence); Tabac et dépenses diverses, pour les variables se rapportant aux achats par le biais de vente directe (questions pour lesquelles la réponse est oui ou non). Il est à noter que les autres variables catégoriques du questionnaire telles que celles se rapportant à la composition du ménage ou encore aux divers branchements du

questionnaire sont vérifiées et validées par les spécialistes du sujet de la Division de la statistique du revenu. Par conséquent, ces dernières variables ne sont pas imputées au moyen de la technique du plus proche voisin.

Tableau 5.3
Ménages qui nécessitent l'imputation de variables catégoriques par province et territoire

Province ou territoire	Ménages (%) qui nécessitent l'imputation de variables catégoriques			
	Nombre de variables imputées (parmi 56)			Total
	1	2	3 ou plus	
Canada	4,7	0,6	1,2	6,4
Terre-Neuve-et-Labrador	3,1	0,1	0,9	4,2
Île-du-Prince-Édouard	2,1	0,3	0,3	2,8
Nouvelle-Écosse	3,0	0,2	0,7	3,9
Nouveau-Brunswick	2,5	0,2	0,8	3,5
Québec	3,5	0,5	0,6	4,6
Ontario	6,5	1,0	2,4	9,9
Manitoba	4,7	0,6	1,5	6,8
Saskatchewan	7,5	1,3	1,9	10,7
Alberta	5,8	0,2	1,0	7,0
Colombie-Britannique	6,7	1,0	2,1	9,8
Yukon	6,2	0,3	1,6	8,1
Territoires du Nord-Ouest	3,5	0,3	0,3	4,1
Nunavut	2,8	0,9	0,5	4,1

Le tableau 5.3 montre qu'à l'échelle nationale, il a fallu imputer au moins une variable catégorique se rapportant aux caractéristiques du logements et équipement associé au logement, au mode d'occupation du logement et aux achats par le biais de vente directe pour 6,4 % des ménages. Toutefois, pour près de 75 % d'entre eux seulement une variable a été imputée. Les provinces ou territoires pour lesquels on observe les plus basses proportions de ménages ayant requis l'imputation d'au moins une variable catégorique sont l'Île-du-Prince-Édouard (2,8 %), le Nouveau-Brunswick (3,5 %) et la Nouvelle-Écosse (3,9 %). On retrouve les taux les plus élevés en Saskatchewan (10,7 %), en Ontario (9,9 %) et en Colombie-Britannique (9,8 %).

Références

- [1] Tremblay J. et Arsenault, S. 2001. *La méthodologie de l'Enquête sur les dépenses des ménages*. N° 62F0026MIF2001003 au catalogue. Ottawa. Division des méthodes d'enquêtes auprès des ménages, Statistique Canada.
- [2] Rao, J.N.K., Wu, C.F.J. et Yue K. 1992. « Quelques travaux récents sur les méthodes de rééchantillonnage applicables aux enquêtes complexes », *Techniques d'enquête*, 18(2), p.225-234 (Statistique Canada, N° 12-001 au catalogue).
- [3] Statistique Canada, Division de la statistique de revenu. 2003. *Guide de l'utilisateur — Enquête sur les dépenses des ménages*. N° 62F0026MIF2004003 au catalogue. Ottawa.
- [4] Statistique Canada, Division de la statistique de revenu. 2003. *Les habitudes de dépenses au Canada*. N° 62-202 au catalogue. Ottawa.
- [5] Statistique Canada, Division des méthodes d'enquêtes auprès des ménages. 2003. *Enquête sur la population active, Rapport sur les opérations*. Enquête 200312. Ottawa.
- [6] Lapierre, B., Nadeau, C., Tremblay, J. et Gaudet, J. 2004. « Amélioration de la qualité des estimations pour une population à faible revenu : utilisation d'une base duale à l'enquête sur les dépenses des ménages », Recueil du Symposium 2004 de Statistique Canada (à paraître).

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

π_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.