

Rapport technique

Rapport technique sur l'échantillonnage et la pondération



Enquête nationale auprès des ménages, 2011



Statistique
Canada

Statistics
Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à infostats@statcan.gc.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros sans frais suivants :

- | | |
|---|----------------|
| • Service de renseignements statistiques | 1-800-263-1136 |
| • national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur | 1-877-287-4369 |

Programme des services de dépôt

- | | |
|-----------------------------|----------------|
| • Service de renseignements | 1-800-635-7943 |
| • Télécopieur | 1-800-565-7757 |

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. 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 sur le site www.statcan.gc.ca sous « À propos de nous » > « Notre organisme » > « [Offrir des services aux Canadiens](#) »

Note de reconnaissance

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

Signes conventionnels dans les tableaux

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

- . indisponible pour toute période de référence
- .. indisponible pour une période de référence précise
- ... n'ayant pas lieu de figurer
- 0 zéro absolu ou valeur arrondie à zéro
- 0^s valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
- ^p provisoire
- ^r révisé
- x confidentiel en vertu des dispositions de la *Loi sur la statistique*
- ^E à utiliser avec prudence
- F trop peu fiable pour être publié
- * valeur significativement différente de l'estimation pour la catégorie de référence ($p < 0,05$)

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2015

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'[entente de licence ouverte](#) de Statistique Canada.

Une [version HTML](#) est aussi disponible.

This publication is also available in English.

Rapport technique sur l'échantillonnage et la pondération

Table des matières

	Page
Introduction	2
1. Collecte des données	3
1.1 Collecte des données du recensement	3
1.2 Méthodologie de collecte de l'Enquête nationale auprès des ménages de 2011	7
2. Traitement des données du Recensement et de l'Enquête nationale auprès des ménages	10
2.1 Introduction	10
2.2 Réception et enregistrement.....	10
2.3 Imagerie et saisie des données à partir d'images	10
2.4 Contrôles de couverture	11
2.5 Contrôles d'intégralité et suivi des cas de rejet au contrôle	11
2.6 Codage	12
2.7 Classification et ajustements pour la non-réponse des logements inoccupés et des logements non répondants	12
2.8 Contrôle et imputation.....	13
2.9 Pondération.....	13
3. Échantillonnage	14
3.1 Échantillon de l'Enquête nationale auprès des ménages.....	14
3.2 Sous-échantillon de l'ENM.....	15
3.3 UC ciblées.....	16
3.4 Sélection du sous-échantillon pour le SCNR.....	17
4. Estimation de l'Enquête nationale auprès des ménages	19
4.1 Éléments à considérer pour choisir une méthode d'estimation	19
4.2 Univers de l'ENM	20
4.3 Poids de sondage	21
4.4 Ajustement pour la non-réponse totale.....	21
4.5 Répondants-surprises.....	22
4.6 Calage.....	22
5. Indicateurs du biais de non-réponse de l'Enquête nationale auprès des ménages.....	25
5.1 Introduction	25
5.2 Méthodologie des indicateurs de biais de la non-réponse	25
5.3 Limites des indicateurs de biais de la non-réponse de l'ENM.....	26
5.4 Quelques indicateurs de biais de la non-réponse de l'ENM.....	26
6. Régions de pondération.....	28
6.1 Formation des RP ordinaires	28
6.2 Ajustement des RP pour les régions recensées par interview et les réserves	29
7. Évaluation des procédures de pondération	33
7.1 Suppression de contraintes	33
7.2 Répartition des poids	36
7.3 Écarts entre les chiffres du recensement et les estimations de l'ENM – Canada.....	38
8. Variance.....	42
9. Conclusion	49
Annexe A Glossaire	50
Annexe B Historique de l'échantillonnage au recensement canadien	52
Annexe C Contraintes utilisées dans le processus de pondération	54
Bibliographie	56

Introduction

L'Enquête nationale auprès des ménages (ENM) de 2011 est une nouvelle enquête à participation volontaire dont le questionnaire a été distribué à environ 30 % des ménages visés par le Recensement du Canada. L'ENM sert à recueillir les renseignements obtenus antérieurement au moyen du questionnaire complet du recensement reçu par un ménage sur cinq tous les cinq ans. Le questionnaire du Recensement de 2011 contient les huit questions qui figuraient sur le questionnaire abrégé du Recensement de 2006, ainsi que deux questions supplémentaires sur les langues. Contrairement à l'ENM, la participation au recensement est obligatoire.

L'ENM a servi à recueillir des renseignements socioéconomiques dont les collectivités ont besoin pour planifier des services tels que les garderies, les écoles, les services à la famille, le logement, les routes et les transports en commun, ainsi que la formation professionnelle des candidats à l'emploi. L'objectif de l'ENM est de fournir des données pour de petites régions géographiques et de petits groupes de population.

Afin d'aider les utilisateurs à évaluer l'utilité des données de l'ENM pour leur propre usage, le présent Rapport technique sur l'échantillonnage et la pondération de l'Enquête nationale auprès des ménages de 2011 énonce en détail les définitions et le cadre conceptuel sur lesquels s'appuie l'ENM, ainsi que les procédures de collecte et de traitement des données qui ont été employées. Comme ces procédures sont reliées aux procédures de collecte et de traitement des données utilisées pour le Recensement de 2011, ces dernières sont exposées également. Ensuite, le rapport décrit la méthode d'échantillonnage et de pondération appliquée pour l'ENM de 2011, ainsi que son incidence sur les résultats. Comme certains renseignements sont recueillis par échantillonnage, puis pondérés à l'échelle de la population totale, un biais et des écarts peuvent être observés dans les estimations finales. Le présent rapport énumère certaines de ces différences observées et en explique les causes probables.

En raison de son effet possible sur les taux de réponse, le caractère volontaire de la nouvelle ENM a entraîné l'application de méthodes d'échantillonnage et de pondération plus complexes que celles utilisées pour le Recensement de 2006.

En général, les raisons qui motivent le recours à l'échantillonnage émanent d'un désir de réduire le fardeau de réponse, de réduire les coûts, d'obtenir les résultats plus rapidement ou d'une combinaison des trois. Un inconvénient de l'échantillonnage est que les résultats fondés sur un échantillon ne sont pas toujours aussi précis que ceux qui s'appuient sur l'ensemble de la population. Néanmoins, quand la perte de précision (qui peut être assez faible si l'échantillon est grand) est tolérable compte tenu des usages prévus des données, l'échantillonnage est souvent avantageux.

Pour le Recensement de la population et l'ENM de 2011, l'échantillonnage a été utilisé de diverses façons. On y a recouru pour :

- S'assurer que la qualité du travail des agents recenseurs durant la collecte des questionnaires répondait à certaines normes.
- Contrôler la qualité du codage des réponses durant le dépouillement.
- Évaluer la qualité des données du recensement et de l'ENM.

Toutefois, l'usage principal de l'échantillonnage consistait à sélectionner les logements à inclure dans l'échantillon de l'ENM. Le présent rapport décrit cet usage particulier de l'échantillonnage et évalue son incidence sur la qualité des données de l'ENM.

Le [chapitre 1](#) décrit en détail les méthodes utilisées dans le cadre du Recensement et de l'ENM de 2011. Le [chapitre 2](#) explique les procédures de traitement des données. Le [chapitre 3](#) décrit comment l'échantillonnage est mis en œuvre dans l'ENM. Le [chapitre 4](#) donne un aperçu des procédures utilisées pour appliquer les poids aux données provenant de l'échantillon de l'ENM afin d'obtenir des estimations pour la population. Le [chapitre 5](#) fournit des renseignements sur les indicateurs du biais de non-réponse à l'ENM, tandis que le [chapitre 6](#) décrit les régions de pondération. Le [chapitre 7](#) présente une évaluation des procédures de pondération et le [chapitre 8](#) donne un aperçu de la méthodologie d'estimation de la variance dans le cadre de l'ENM de 2011.

1. Collecte des données

1.1 Collecte des données du recensement

L'étape de la collecte des données du Recensement de 2011 visait à faire en sorte que chacun des 14,5 millions de logements du Canada ait reçu un questionnaire du recensement. Le recensement visait l'ensemble de la population du Canada, qui comprend les citoyens canadiens (de naissance ou par naturalisation), les immigrants reçus et les résidents non permanents, ainsi que les membres de leur famille vivant sous le même toit. Les résidents non permanents sont les personnes qui vivent au Canada et qui sont titulaires d'un permis de travail ou d'études, ou qui revendiquent le statut de réfugié, ainsi que les membres de leur famille qui vivent avec eux.

Le recensement dénombrait également les citoyens canadiens et les immigrants reçus qui étaient temporairement à l'extérieur du pays le jour du recensement, y compris les fonctionnaires fédéraux et provinciaux/territoriaux en poste à l'extérieur du Canada, le personnel des ambassades canadiennes à l'étranger, les membres des Forces canadiennes en poste à l'étranger et les Canadiens qui se trouvaient à bord de navires marchands.

Au Recensement du Canada, divers questionnaires et diverses formules sont utilisés pour recueillir les données. La formule 1 est le Registre des visites (RV) qui sert à dresser la liste des logements privés et collectifs occupés et inoccupés ainsi que des exploitants et exploitations agricoles qui se trouvent dans l'unité de collecte. Le RV sert de liste d'adresses aux fins du contrôle et des opérations sur le terrain dans le cadre de la collecte des données.

La formule 2A est le questionnaire abrégé de base servant à dénombrer tous les logements privés. Chaque ménage qui reçoit un questionnaire du recensement 2A est invité à y dénombrer tous les membres du ménage qui sont inclus dans la population du recensement et à répondre aux questions pour eux.

La formule 2C sert à dénombrer les Canadiens en poste à l'étranger, notamment les fonctionnaires (fédéraux et provinciaux/territoriaux) et les membres de leur famille, ainsi que les membres des Forces canadiennes et les membres de leur famille.

La formule 3A est un questionnaire de recensement individuel qui sert à dénombrer les personnes dans les logements collectifs¹ (chaque résident d'un logement collectif doit remplir sa propre formule 3). Elle peut également être utilisée pour dénombrer des résidents habituels d'un ménage privé qui veulent remplir leur propre questionnaire du recensement plutôt qu'être inscrits sur un questionnaire 2A.

La méthodologie par vagues est une nouvelle approche de collecte qui a été utilisée pour la première fois lors du Recensement de 2011. On communique avec les ménages à des moments clés afin de leur rappeler de participer au recensement et de les inciter à remplir le questionnaire. À chaque vague de communication, les renseignements nécessaires à la réponse sont fournis aux ménages. Cette méthode, qui s'inscrit dans un contexte où chaque ménage canadien est tenu par la loi de répondre aux questions du recensement, est conçue de façon à encourager la réponse hâtive par Internet tout en atténuant le risque de diminution de la réponse totale et le besoin de suivis coûteux en personne.

Cette nouvelle méthodologie diffère selon le mode de collecte utilisé pour distribuer les documents du recensement pour une région donnée. Ces modes de collecte sont décrits à la section suivante. En 2011, les ménages canadiens pouvaient répondre par Internet, questionnaire papier (retour par la poste) ou en contactant l'Assistance téléphonique du recensement.

1.1.1 Modes de collecte : livraison des questionnaires du recensement

Les trois modes de livraison des questionnaires utilisés au Recensement canadien sont l'envoi par la poste, le listage/livraison et le recensement par interview. Afin d'organiser la collecte du recensement le plus efficacement possible, le territoire du Canada est divisé en petites unités géographiques appelées unités de collecte (UC). Au Recensement de 2011, on comptait environ 46 000 UC au Canada.

1. Un logement collectif est un logement utilisé à des fins commerciales, institutionnelles ou communautaires (par exemple les hôpitaux, les hôtels et les prisons, etc.).

Rapport technique sur l'échantillonnage et la pondération

1.1.1.1 Envoi par la poste

Dans les unités de collecte d'envoi par la poste, on a recours au système postal pour envoyer le matériel du recensement. Ce type d'envoi permet une distribution efficace et coordonnée, sans avoir à recruter et à former un effectif important d'agents recenseurs. Les UC d'envoi par la poste sont typiquement dans les régions urbaines. Bien que ce type d'UC couvre maintenant environ 80 % des logements au Canada, il représente seulement une très petite partie de la superficie du territoire.

1.1.1.2 Listage/livraison

Les UC de listage/livraison sont typiquement des régions rurales. Dans ces régions, des agents recenseurs effectuent le listage des logements et ils font la livraison du matériel du recensement. Environ 18 % des logements du Canada se trouvent dans ce type d'UC qui couvre une partie importante de la superficie du territoire.

1.1.1.3 Recensement par interview

Les UC de recensement par interview sont habituellement situées dans des endroits éloignés ou difficiles d'accès, ainsi que dans les communautés autochtones. Afin de limiter le nombre de déplacements souvent coûteux et logistiquement complexes des agents recenseurs pour faire des suivis à ces endroits, ces derniers ne se limitent pas à faire le listage et la livraison du matériel du recensement, mais remplissent aussi immédiatement le questionnaire avec chacun des ménages. Ce type d'UC couvre un peu plus de la moitié de la superficie du territoire du Canada, mais seulement environ 2 % des logements.

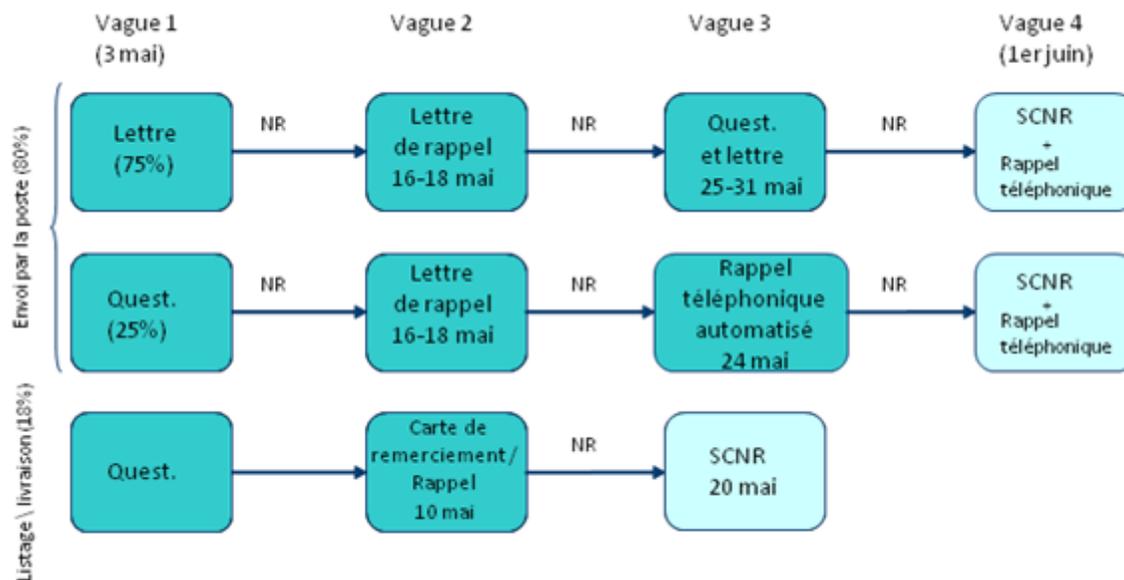
1.1.2 Méthodologie par vagues du recensement

La méthodologie par vagues a été conçue de façon à encourager la réponse par Internet tout en offrant une alternative aux ménages qui ne souhaitent pas remplir leur questionnaire par Internet. Cette approche comporte de nombreux avantages pour les taux de réponse, l'enregistrement des questionnaires, l'enchaînement des questions et la saisie des données.

La méthodologie par vagues s'applique de façon différente selon le groupe d'unités de collecte. Trois principaux groupes d'UC ont été définis, et une variante de la méthodologie par vagues a été développée pour chacun d'eux. En raison de la nature même des UC de recensement par interview, une méthodologie par vagues n'a toutefois pas été développée pour ces unités. Les sections suivantes ainsi que la [Figure 1.1.2.1](#) donnent un aperçu de la méthodologie par vagues qui a été utilisée pour le Recensement de 2011.

Rapport technique sur l'échantillonnage et la pondération

Figure 1.1.2.1 Aperçu de la méthodologie par vagues du Recensement de 2011



Note : NR = non-répondants, SCNR = suivi des cas de non-réponse.

Source : Statistique Canada, Recensement de 2011.

1.1.2.1 Unités de collecte d'envoi par la poste – Lettre à la vague 1

Premièrement, un ensemble d'UC d'envoi par la poste a été identifié pour cibler les ménages les plus susceptibles de répondre par Internet tout en évitant ceux les moins enclins à répondre au recensement. Cet ensemble d'UC correspond approximativement à 75 % des logements dans les régions d'envoi par la poste. Dans ce groupe d'UC, la première vague a consisté en l'envoi d'une lettre seulement et dans laquelle on demandait aux ménages de remplir le questionnaire en ligne au moyen du code d'accès sécurisé (CAS) qui leur était fourni ou d'utiliser un système automatisé accessible à partir d'un numéro de téléphone sans frais pour demander qu'un questionnaire papier leur soit envoyé par la poste. Les lettres de la vague 1 ont été livrées par le système postal une semaine avant le jour du recensement, soit le 3 mai 2011.

La deuxième vague a consisté en une lettre de rappel envoyée à tous les ménages non répondants de la vague 1. Cette lettre rappelait aux ménages qu'ils étaient tenus par la loi de répondre au recensement. Elle comprenait aussi le CAS et le numéro de téléphone sans frais, tout comme la lettre de la première vague. Elle a été livrée aux ménages du 16 au 18 mai, soit à partir du sixième jour suivant le jour du recensement.

À la troisième vague, on a envoyé un questionnaire papier aux ménages non répondants. Ce questionnaire a été livré aux ménages du 25 au 31 mai, soit à partir du quinzième jour après le jour du recensement. Les ménages pouvaient toujours répondre en ligne au moyen d'un CAS fourni sur la page couverture du questionnaire. De plus, une lettre accompagnait le questionnaire et mentionnait que si le questionnaire n'était pas dûment rempli au 31 mai 2011, un agent recenseur allait communiquer avec le ménage par appel téléphonique ou en personne pour remplir le questionnaire. La lettre mentionnait aussi que si le ménage refusait de répondre aux questions du recensement, le dossier pouvait être déféré au Service des poursuites pénales du Canada qui prendrait alors les mesures qui s'imposent en vertu de la *Loi sur la statistique*.

La quatrième vague, qui a débuté le 1er juin 2011, a été le suivi des cas de non-réponse (SCNR) sur le terrain, et incluait aussi un rappel téléphonique automatisé. Le SCNR est décrit à la [section 1.1.3](#) de ce document.

1.1.2.2 Unités de collecte d'envoi par la poste – Questionnaire de la vague 1

Le deuxième groupe d'UC où s'applique une variante de la méthodologie par vagues est le groupe constitué des autres UC d'envoi par la poste. Cet ensemble d'UC correspond approximativement à 25 % des logements dans les régions d'envoi par la poste. On considère les ménages dans ces UC comme étant moins susceptibles de bien répondre après la réception d'une seule lettre. Ainsi, la première vague dans ces UC est l'envoi par la poste d'un questionnaire papier. Ce questionnaire fournit un CAS, donnant ainsi l'option au ménage de répondre par Internet. Comme pour le premier groupe, cette vague a eu lieu une semaine avant le jour du recensement, soit le 3 mai 2011. La deuxième vague pour ce groupe était identique à celle du premier groupe. À la troisième vague, les ménages non répondants de ce groupe pour lesquels un numéro de téléphone était disponible sur la base du recensement ont reçu un rappel téléphonique automatisé le 24 mai. La quatrième vague était identique à celle du premier groupe.

1.1.2.3 Unités de collecte de listage/livraison

Le troisième groupe d'unités de collecte où s'applique une autre variante de la méthodologie par vagues est le groupe constitué de toutes les UC de listage/livraison. À la première vague, des agents recenseurs ont livré un questionnaire papier à tous les logements dans ces UC autour du 3 mai 2011. Ce questionnaire était assorti également d'un CAS, donnant ainsi l'option au ménage de répondre par Internet. À la deuxième vague, le 10 mai, tous les logements dans ces UC ont reçu par la poste une carte de remerciement/rappel, peu importe qu'ils aient déjà répondu ou non, car il n'est généralement pas possible dans ces régions de faire un envoi postal à des adresses de voirie. Finalement, la dernière vague consistait à passer directement au suivi des cas de non-réponse sur le terrain à partir du 20 mai.

1.1.3 Suivi des cas de non-réponse

Comme mentionné à la section précédente, la dernière vague de la méthodologie par vagues correspond au suivi des cas de non-réponse (SCNR). À cette étape, les agents recenseurs appellent et visitent les ménages qui n'ont pas répondu. Chaque ménage non-répondant pour lequel un numéro de téléphone était disponible a reçu un rappel téléphonique automatisé au début de la période du SCNR. Ce message rappelait aux non-répondants leur obligation légale de répondre au recensement.

Les agents recenseurs avaient à leur disposition des informations pour gérer leur travail dans le Système de gestion du terrain. Ce système accessible via Internet est un système informatique qui a été développé pour le Recensement de 2011 afin de faciliter la cueillette d'information sur les progrès de la collecte.

1.1.4 Vérification du statut d'occupation des logements

Avant le SCNR, des opérations sur le terrain ont aussi eu lieu afin d'effectuer la vérification de l'occupation des logements (VOL). Cette activité qui commençait le 13 mai 2011, avait pour but de trouver un nombre significatif de logements inoccupés ou annulés au jour du recensement (adresses qui ne sont pas des logements privés ou collectifs) avant le début du SCNR. En identifiant ces logements à une date proche du jour du recensement, la classification du statut d'occupation devrait être plus précise et même plus facile à effectuer. De plus, la VOL permet de réduire la charge de travail du SCNR puisque les logements inoccupés ou annulés identifiés par cette vérification n'ont plus besoin de suivi.

Tout de même, des erreurs de classification d'un logement comme étant occupé ou inoccupé se produisent durant la VOL et le SCNR. Certains logements classés comme étant inoccupés sont en fait occupés et certains logements non répondants sont des logements inoccupés. Ainsi, une autre opération, l'Enquête sur la classification des logements, se déroule après le SCNR. Elle évalue et détermine le statut d'occupation des logements pour lesquels aucun questionnaire rempli n'a été reçu (logements inoccupés, logements non répondants ou cas non résolus). À partir des résultats de l'enquête, on ajuste les chiffres du Recensement de la population durant le traitement (voir la [section 2.7](#) ou le [Rapport technique sur la couverture](#), Recensement de 2011, n° 98-303-X au catalogue)

1.2 Méthodologie de collecte de l'Enquête nationale auprès des ménages de 2011

La population visée par l'Enquête nationale auprès des ménages (ENM) est constituée des personnes dénombrées par le recensement qui habitent des logements privés.

Deux questionnaires sont utilisés pour recueillir les données : les formules N1 et N2. Les deux formules contiennent les mêmes questions, mais la formule N2 est utilisée dans les régions avec recensement par interview. Les mêmes 10 questions du recensement sont également sur les formules de l'ENM.

Afin de tirer profit de l'utilisation des ressources et de l'infrastructure du recensement telles que les systèmes de gestion de collecte et les employés déjà en place, il a été décidé de mener l'ENM durant la même période que le recensement. Toutefois, cette stratégie avait le potentiel d'augmenter le fardeau des employés et des répondants. La gestion des opérations de collecte de l'ENM devait donc se faire de façon à minimiser toute incidence sur les résultats du recensement puisque ce dernier est prioritaire. Les répondants au recensement par Internet (UC d'envoi par la poste) ont eu l'opportunité de répondre à l'ENM en ligne immédiatement après avoir rempli le questionnaire du recensement alors que les répondants au recensement par questionnaire papier ont été contactés pour répondre à l'ENM à partir du 7 juin, date de livraison du questionnaire papier de l'ENM.

La stratégie de collecte de l'ENM devait aussi tenir compte des limites en ressources, des contraintes de temps et de la nature volontaire de l'enquête qui empêche les intervieweurs d'être trop insistants et qui réduit le nombre de contacts possibles avec les non-répondants.

1.2.1 Collecte de l'ENM dans les régions avec recensement par interview

Dans ces régions, la collecte des données de l'ENM avec la formule N2 se faisait en même temps que le dénombrement pour le recensement. Les réponses aux questions du recensement qui sont aussi sur la formule N2 sont transcrites d'une formule à l'autre, pourvu que le répondant accepte de participer à l'ENM.

Dans les régions où le recensement par interview n'est pas utilisé, on fait appel à une méthodologie de collecte par vagues décrite dans la section suivante.

1.2.2 Méthodologie par vagues de l'ENM

Une méthodologie par vagues a aussi été utilisée pour la collecte de l'ENM, de concert avec celle du recensement. En plus d'être basée sur le mode de collecte de l'UC, la méthodologie par vagues de l'ENM est aussi basée sur l'état de réponse et le mode de réponse (Internet ou papier) au recensement. Ainsi, parmi les logements sélectionnés pour l'ENM, les trois groupes de logements décrits ci-dessous ont été définis.

Le premier groupe de logements, appelé Groupe d'enquête 1, est constitué de répondants au recensement par Internet. Les logements du Groupe d'enquête 2 sont les répondants au recensement par questionnaire papier. Finalement, les non-répondants au recensement constituent le Groupe d'enquête 3.

La méthodologie par vagues de l'ENM est différente pour chaque groupe d'enquête, et les détails pour chacun d'entre eux sont présentés ci-dessous. Le [tableau 1.2.2.1](#) en donne aussi un aperçu. Tout comme pour le recensement, en raison de la nature même des UC de recensement par interview, la méthodologie par vagues n'a pas été appliquée dans ces unités.

Rapport technique sur l'échantillonnage et la pondération

Tableau 1.2.2.1 Aperçu de la méthodologie par vagues de l'ENM de 2011

Groupe d'enquête 1 – Répondants au recensement par Internet (UC d'envoi par la poste)	
Vague 1	Offre en ligne immédiatement après avoir rempli le questionnaire du recensement
Vague 2	Lettre de rappel reçue autour du 7 juin
Vague 3	Aucune
Vague 4	Suivi des cas de non-réponse par appel téléphonique ou en personne (débutant seulement lorsque le niveau de réponse au recensement était acceptable, à partir du 8 juin dans certaines régions où la collecte du recensement était presque terminée)
Groupe d'enquête 2 – Répondants au recensement par questionnaire papier (UC d'envoi par la poste)	
Vague 1	Questionnaire reçu autour du 7 juin
Vague 2	Lettre de rappel reçue autour du 14 juin
Vague 3	Questionnaire reçu autour du 6 juillet – Annulée
Vague 4	Suivi des cas de non-réponse par appel téléphonique ou en personne (comme pour le Groupe d'enquête 1)
Groupe d'enquête 3 – Non-répondants au recensement (UC d'envoi par la poste)	
Vague 1	Questionnaire reçu autour du 7 juin – Annulée
Vague 2	Aucune
Vague 3	Aucune
Vague 4	Suivi des cas de non-réponse par appel téléphonique ou en personne, en même temps que le suivi de la non-réponse du recensement, à partir du 8 juin
UC de listage/livraison	
Vague 1	Offre en ligne immédiatement après avoir rempli le questionnaire du recensement (Groupe d'enquête 1 seulement)
Vague 2	Questionnaire livré autour du 7 juin à tous les logements sélectionnés (sauf les répondants à l'ENM par Internet)
Vague 3	Aucune
Vague 4	Suivi des cas de non-réponse par appel téléphonique ou en personne à partir du 8 juin (en même temps que le suivi de la non-réponse du recensement pour le Groupe d'enquête 3)

Source : Statistique Canada, Enquête nationale auprès des ménages de 2011.

1.2.2.1 Envoi par la poste – Groupe d'enquête 1

Les répondants du recensement par Internet sélectionnés dans l'échantillon de l'ENM ont été invités à remplir le questionnaire de l'ENM immédiatement après avoir complété le recensement. Après avoir soumis leur questionnaire en ligne (en utilisant le code d'accès sécurisé original), ces répondants ont été automatiquement dirigés vers un écran de transition les informant qu'ils avaient été sélectionnés pour l'ENM. Cet écran contenait aussi de l'information de base à propos de l'enquête telle qu'une explication de son importance et une mention de sa nature volontaire. Les répondants désirant continuer ont alors été redirigés au questionnaire de l'ENM et l'application passait automatiquement les premières questions qui étaient les mêmes que celles du recensement et qui venaient juste d'être répondues. Ceci constituait la première vague pour le Groupe d'enquête 1.

La deuxième vague du Groupe d'enquête 1 a été l'envoi d'une lettre de rappel aux ménages non répondants. La lettre, livrée aux ménages autour du 7 juin, plaidait l'importance de l'enquête.

Aucun autre document n'a été envoyé après la vague 2 (c.-à-d. pas de vague 3) et l'étape suivante était le suivi des cas de non-réponse (SCNR) sur le terrain. Le début des suivis s'est fait progressivement par unité de collecte à partir du 8 juin, dépendamment du niveau de réponse au recensement et de la possibilité de redéployer les agents recenseurs pour prêter main-forte à la collecte du recensement dans les régions ayant du retard.

Les procédures du terrain pour la conversion de la non-réponse de l'ENM étaient similaires à celles du recensement. Les employés, les infrastructures et les systèmes du recensement ont aussi été utilisés. Lorsque c'était possible sur le plan opérationnel, les tentatives de conversion des refus étaient effectuées en personne par des chefs d'équipes et des agents recenseurs ayant reçu une formation spéciale à cet effet. La définition d'un refus était la même que celle du recensement, soit un refus clairement exprimé de répondre à l'enquête. Par contre, étant donné la nature volontaire de l'ENM, ce sont plutôt les règles habituelles de Statistique Canada de refus pour les enquêtes volontaires qui

Rapport technique sur l'échantillonnage et la pondération

s'appliquaient, c'est-à-dire interruption des activités de suivi après deux refus clairs (incluant les refus par appel téléphonique).

De plus, des intervieweurs des centres d'interviews téléphoniques assistées par ordinateur de Statistique Canada ont été utilisés comme supplément au SCNR de l'ENM. L'utilisation de ces intervieweurs était faite surtout dans les régions où la collecte de l'ENM avait du retard et où le nombre d'employés sur le terrain était insuffisant.

Il est à noter que les logements du Groupe d'enquête 1 n'ont jamais reçu de questionnaire papier de l'ENM. Puisque ces logements ont choisi de répondre au recensement par Internet, on faisait l'hypothèse qu'ils utiliseraient le même mode de réponse pour l'ENM s'ils acceptaient d'y répondre. L'envoi d'un questionnaire papier aurait eu peu d'incidence sur le taux de réponse et cela aurait augmenté significativement les coûts de collecte.

1.2.2.2 Envoi par la poste – Groupe d'enquête 2

Le questionnaire papier du recensement ne faisait aucune mention de l'enquête additionnelle volontaire que le logement pourrait recevoir plus tard. Ainsi, le premier contact avec l'ENM des logements ayant répondu au recensement avec un questionnaire papier au cours du mois de mai était environ 4 semaines après le jour du recensement (autour du 7 juin), alors qu'ils recevaient le questionnaire papier de l'ENM par la poste. Un code d'accès sécurisé pour l'utilisation d'Internet était imprimé sur chaque questionnaire. Ceci constituait la vague 1 pour ce groupe.

Pour la deuxième vague, une lettre de rappel a été livrée aux ménages non répondants une semaine plus tard, autour du 14 juin. Comme pour la lettre de rappel envoyée au Groupe d'enquête 1, cette lettre rappelait aux non-répondants l'importance de répondre à l'enquête volontaire.

Selon les plans initiaux, la vague 3 aurait consisté en l'envoi par la poste d'un deuxième questionnaire papier aux ménages non répondants. Ce questionnaire, accompagné d'une autre lettre de rappel, devait être livré autour du 6 juillet. Toutefois, en raison de l'incertitude générée par la possibilité d'un arrêt de travail à Postes Canada et de l'opportunité de commencer plus tôt les suivis sur le terrain, cette vague a été annulée.

Le SCNR sur le terrain (vague 4) a débuté lorsque le déroulement de la collecte du recensement était suffisamment avancé. Les mêmes procédures que celles décrites précédemment pour le Groupe d'enquête 1 ont été suivies.

1.2.2.3 Envoi par la poste – Groupe d'enquête 3

La méthodologie par vagues pour les non-répondants au recensement était plus simple que pour les autres groupes d'enquête. Puisque les répondants des logements du Groupe d'enquête 3 n'avaient pas réagi positivement à aucune des incitations envoyées pour répondre au recensement, lequel est de nature obligatoire; envoyer des lettres, questionnaires ou messages vocaux additionnels pour encourager la réponse à l'enquête volontaire n'aurait pas eu d'effet significatif. Seul l'envoi du questionnaire de l'ENM pour être reçu autour du 7 juin était prévu, mais cette vague a également été annulée en raison de la possibilité d'un arrêt de travail à Postes Canada afin de donner priorité à la préparation et à la distribution des questionnaires du Groupe d'enquête 2 qui s'effectuait en même temps.

Ainsi, le seul contact avec les ménages du Groupe d'enquête 3 a été la vague 4, soit le SCNR sur le terrain. Celui-ci a commencé le 8 juin, et il se faisait lors de la même visite aux logements que pour le suivi du recensement.

1.2.2.4 Listage/livraison – Groupes d'enquête 1, 2 et 3

Les logements dans les UC de listage/livraison ont reçu leur questionnaire du recensement au début du mois de mai. Similairement à la méthodologie par vagues pour les groupes d'enquête des UC d'envoi par la poste, les logements qui ont répondu au recensement par Internet en utilisant le code d'accès sécurisé imprimé sur le questionnaire ont été invités à remplir le questionnaire de l'ENM en ligne immédiatement après avoir terminé le questionnaire du recensement. Aussi, un questionnaire de l'ENM a été livré par les agents recenseurs autour du 7 juin aux logements ayant choisi de répondre au recensement par papier ou n'ayant pas répondu au recensement, tout comme à ceux qui ont refusé l'offre de l'ENM après avoir rempli le questionnaire du recensement par Internet.

Aucun autre document n'a été livré aux portes, et le SCNR par appel téléphonique ou en personne a commencé aux mêmes dates que pour les régions d'envoi par la poste. Pour le Groupe d'enquête 3, le SCNR de l'ENM se faisait aussi lors de la même visite aux logements que celle du SCNR pour le recensement.

2. Traitement des données du Recensement et de l'Enquête nationale auprès des ménages

2.1 Introduction

Le présent chapitre porte sur le traitement de tous les questionnaires remplis, ce qui englobe toutes les activités allant de la réception des questionnaires à la création d'une base de données du recensement et d'une base de données de l'Enquête nationale auprès des ménages (ENM) exactes et complètes. Les étapes du traitement décrites ci-dessous sont l'enregistrement des questionnaires, l'imagerie des questionnaires et la saisie des données, la vérification, la correction des erreurs, le suivi des cas de rejet à la vérification, le codage, la classification des logements et les ajustements pour tenir compte de la non-réponse, l'imputation et la pondération.

Les processus automatisés mis en œuvre pour le Recensement et l'ENM de 2011 ont fait l'objet d'un contrôle afin de s'assurer que toutes les résidences de Canadiens soient dénombrées une fois et seulement une fois, et d'indiquer lesquelles de ces résidences devaient être incluses dans l'ENM. Le Système de contrôle principal (SCP) a été construit afin de contrôler et de surveiller l'enchaînement des opérations, de la collecte au traitement des données. Le SCP contenait une liste principale de tous les logements existants au Canada sur laquelle figurait pour chacun d'eux un numéro d'identification unique. Ce système a été mis à jour quotidiennement à l'aide des renseignements sur le statut de chaque logement dans l'enchaînement des opérations du recensement et de l'ENM (c.-à-d. questionnaire livré, reçu, traité, etc.). Le système a produit des rapports que les gestionnaires pouvaient consulter en ligne afin de s'assurer que les opérations du recensement et de l'ENM se déroulaient de manière efficace et efficiente.

2.2 Réception et enregistrement

Les réponses reçues par Internet ou par interview téléphonique de l'assistance téléphonique ont été transmises directement à une installation centralisée de traitement des données appelée Centre des opérations des données (COD) où leur réception a été enregistrée automatiquement.

Les répondants qui ont rempli un questionnaire papier devaient le retourner au COD par la poste. Postes Canada enregistrerait automatiquement leur réception dans plusieurs emplacements au Canada (dans le cadre du processus ordinaire de traitement du courrier) en scannant par la fenêtre transparente des enveloppes-réponses le code à barres qui figurait à la première page du questionnaire. Puis, les enveloppes étaient livrées au COD. Chaque jour, Postes Canada envoyait un fichier énumérant tous les questionnaires du recensement et de l'ENM reçus à chaque établissement régional de traitement du courrier, par date de réception.

L'enregistrement de chaque questionnaire retourné était signalé dans le SCP à Statistique Canada. La liste de tous les logements pour lesquels un questionnaire n'avait pas été reçu était produite par le SCP, puis transmise aux opérations sur le terrain en vue d'un suivi. Des mises à jour concernant l'enregistrement étaient envoyées quotidiennement aux opérations sur le terrain afin d'éviter d'effectuer un suivi auprès des ménages qui avaient déjà rempli leur questionnaire.

2.3 Imagerie et saisie des données à partir d'images

En 2011, les formulaires transformés en image étaient les trois questionnaires du recensement (2A, 2C, 3A), le questionnaire du Recensement de l'agriculture (F6), et les deux questionnaires de l'ENM (N1, N2). La qualité des images a été améliorée comparativement à 2006 grâce au remplacement des scanners en noir et blanc par des scanners en couleurs. Le processus d'imagerie comprenait les étapes suivantes :

- **Préparation des documents** : les questionnaires retournés par la poste ont été retirés des enveloppes, puis débarrassés de tous corps étrangers, comme des trombones ou des agrafes, en vue du balayage. Les questionnaires ont été mis en lot selon le type de formule. Les questionnaires sous forme de cahier ont été séparés en feuilles individuelles en coupant le dos de la reliure.
- **Balayage** : à cette étape, les questionnaires papier ont été convertis en images numériques.

Rapport technique sur l'échantillonnage et la pondération

- **Évaluation automatisée de la qualité des images** : un système automatisé a analysé les images afin de déceler les erreurs ou les anomalies. Les images rejetées à cette étape ont été transmises pour examen à un opérateur spécialisé en analyse de documents.
- **Analyse des documents** : à cette étape, les images contenant des anomalies ont été présentées à un opérateur en vue de leur examen. L'opérateur pouvait accepter l'image telle quelle, l'envoyer directement à l'étape de la saisie des données ou l'envoyer en vue d'un nouveau balayage.
- **Reconnaissance automatisée de caractères** : cette étape consistait en une tentative de reconnaissance automatique des réponses et des marques manuscrites sur les questionnaires.
- **Saisie des données par clavier** : des opérateurs ont entré manuellement les réponses que le système de reconnaissance automatisée ne pouvait pas saisir avec suffisamment de certitude.
- **Vérification à la sortie** : après qu'ils aient franchi avec succès toutes les étapes susmentionnées, les questionnaires papier ont été retirés du système. La vérification à la sortie est un processus d'assurance de la qualité destiné à s'assurer que les images et les données saisies sont d'une qualité telle que les questionnaires papier ne doivent plus subir de traitement ultérieur. Les questionnaires qui ont été identifiés comme contenant des erreurs ont été retirés à la sortie et traités à nouveau.

2.4 Contrôles de couverture

Les contrôles de couverture ont été appliqués aux questionnaires du recensement ainsi qu'à ceux de l'ENM. À cette étape, un certain nombre de contrôles automatisés ont été exécutés sur les données des répondants. Ces contrôles avaient pour but de déceler les cas où des personnes non valides pouvaient avoir été créées, à cause d'une erreur du répondant ou d'une erreur de saisie des données. Les données entrées par erreur pour une personne dans la mauvaise colonne, les données supprimées qui ont été saisies par erreur, ou les données fournies plus d'une fois pour la même personne, habituellement en raison de la réception de formules en double (p. ex., une personne a répondu au questionnaire par Internet, tandis que son conjoint ou sa conjointe a rempli le questionnaire papier et l'a renvoyé par la poste) en sont des exemples. Les contrôles ont également été conçus afin de déceler l'absence éventuelle de résidents habituels, lorsque les données n'étaient pas fournies pour chacun des membres du ménage énumérés au début du questionnaire.

Environ 45 % des cas de rejet au contrôle ont été résolus de manière déterministe par le système. Les autres cas ont été transmis aux commis préposés au traitement en vue d'être résolus. Un système interactif a permis aux commis d'examiner les données saisies et de les comparer à l'image du questionnaire si celle-ci était disponible (les questionnaires en ligne n'ont pas d'image). Les rejets au contrôle ont été résolus en supprimant les personnes non valides ou en double et en ajoutant les personnes manquantes (c.-à-d. en créant des enregistrements de personne en blanc) au besoin et lorsque pertinent.

2.5 Contrôles d'intégralité et suivi des cas de rejet au contrôle

Les contrôles d'intégralité et le suivi des cas de rejet au contrôle ne s'appliquent qu'aux questionnaires du recensement. Après les contrôles de couverture, une autre série de contrôles automatisés a été exécutée sur les questionnaires du recensement afin de déceler les cas qui comportaient trop de réponses manquantes ou qui portaient à croire que les données n'avaient peut-être pas été fournies pour tous les résidents habituels du ménage. Les ménages dont le questionnaire a été rejeté lors de ces contrôles ont fait l'objet d'un suivi. Un intervieweur a téléphoné au répondant afin de résoudre les problèmes de couverture éventuels et d'obtenir les renseignements manquants en se servant d'une application d'interview téléphonique assistée par ordinateur. Les données ont été renvoyées au COD afin d'être réintégrées dans le système en vue d'un traitement subséquent.

2.6 Codage

Les questionnaires du recensement et de l'ENM contenaient tous deux des questions dont les réponses pouvaient être cochées sur une liste, ainsi que des questions dont les réponses devaient être inscrites par le répondant dans les cases prévues à cette fin. Ces réponses écrites ont été soumises à un codage automatisé afin d'attribuer à chacune un code numérique en se servant des fichiers de référence, des ensembles de codes et des classifications types de Statistique Canada. Les fichiers de référence utilisés pour le processus d'appariement automatisé ont été créés en se servant des réponses réelles recueillies lors des recensements antérieurs, ainsi que de fichiers de données administratives. Des codeurs ayant reçu une formation particulière et des spécialistes du domaine ont résolu les cas où un code ne pouvait pas être attribué automatiquement. Les questions du recensement ainsi que de l'ENM qui nécessitaient un codage étaient les suivantes : Lien avec la Personne 1, Langue parlée à la maison et Langue maternelle. Les questions de l'ENM seulement nécessitant un codage étaient les suivantes : Lieu de naissance, Citoyenneté, Langues non officielles, Origine ethnique, Groupe de population, Première Nation/bande indienne, Religion, Lieu de résidence 1 an auparavant, Lieu de résidence 5 ans auparavant, Lieu de naissance des parents, Principal domaine d'études, Lieu des études, Industrie, Profession, Lieu de travail et Langue de travail.

Environ 15 millions de réponses écrites ont été codées à partir des questionnaires du Recensement de 2011, tandis qu'environ 46 millions ont été codées à partir des questionnaires de l'ENM. Globalement, environ 87 % ont été codées automatiquement, mais le taux de codage automatique variait considérablement d'une variable à l'autre.

Une fois que les réponses pour une variable particulière étaient codées, les données pour cette variable étaient envoyées à l'étape du contrôle et de l'imputation.

2.7 Classification et ajustements pour la non-réponse des logements inoccupés et des logements non répondants

L'Enquête sur la classification des logements (ECL) a servi à estimer le taux d'erreur des agents recenseurs lors de la classification des logements comme étant occupés ou inoccupés dans les régions de collecte par autodénombrement du recensement. En s'appuyant sur l'information ainsi recueillie, des corrections ont été apportées à la base de données du recensement. L'ECL a été réalisé auprès d'un échantillon aléatoire de 1 729 unités de collecte (UC) autodénombrées qui ont été visitées en juillet et août 2011 afin de réévaluer le statut d'occupation le jour du recensement pour chaque logement pour lequel aucune réponse n'avait été reçue. L'ECL a montré que 13,8 % des 1 099 156 logements classés comme étant inoccupés étaient en fait occupés et que 30,8 % des 317 976 logements sans réponse qui avaient été classés comme étant occupés ou dont le statut d'occupation était inconnu, étaient en fait inoccupés. Les estimations fondées sur l'échantillon de l'ECL ont été utilisées pour corriger le statut d'occupation des logements individuels, ce qui a donné lieu à une augmentation de 3,3 % du nombre de logements occupés et à une diminution de 5,0 % du nombre de logements inoccupés à l'échelle du Canada.

Après cet ajustement du statut d'occupation par l'ECL, les logements occupés ayant une non-réponse totale ont fait l'objet d'une imputation du nombre de résidents habituels (s'il n'était pas connu) et de toutes les réponses aux questions du recensement en empruntant les réponses non imputées d'un autre ménage appartenant à la même UC. Ce processus, appelé imputation des ménages au complet (IMC), a été appliqué à 99 % des ménages ayant une non-réponse totale. L'utilisation d'un seul donneur dans le cadre de l'IMC était informatiquement plus efficace et moins susceptible de produire des résultats non plausibles que l'utilisation de plusieurs donneurs dans le cadre du processus principal de contrôle et d'imputation. Néanmoins, pour les autres ménages ayant une non-réponse totale, c'est-à-dire 1 % d'entre eux, pour lesquels aucun ménage donneur n'avait été trouvé dans le cadre de l'IMC, l'imputation a été effectuée au moyen du processus principal de contrôle et d'imputation.

Des renseignements plus détaillés sur l'ECL et la procédure d'imputation des ménages au complet figurent dans le [Rapport technique sur la couverture](#), Recensement de 2011, n° 98-303-X au catalogue.

2.8 Contrôle et imputation

Les données recueillies dans le cadre de toute enquête ou recensement contiennent des omissions ou des incohérences. Par exemple, un répondant peut éprouver de la réticence à répondre à une question, ne pas se souvenir de la bonne réponse ou mal comprendre la question. D'autres erreurs, comme le codage incorrect d'une réponse, peuvent aussi avoir lieu.

L'épuration définitive des données, à l'étape du contrôle et de l'imputation, a été presque entièrement automatisée. Deux types d'imputation ont été appliqués. Le premier, appelé « imputation déterministe », consiste à attribuer des valeurs particulières sous certaines conditions quand la résolution du problème est claire et non ambiguë. Des règles de contrôle détaillées ont été appliquées pour déterminer ces conditions, puis des valeurs prédéterminées ont été attribuées aux variables intervenant dans les règles. Le deuxième type d'imputation, appelé « imputation par donneur à changements minimaux par le plus proche voisin », consiste à appliquer une série de règles de contrôle détaillées qui permettent de déceler toute réponse manquante ou incohérente. Quand un enregistrement contenant des réponses manquantes ou incohérentes est repéré, un autre enregistrement dont la plupart des caractéristiques sont les mêmes que celles de l'enregistrement erroné est sélectionné. Les données provenant de cet enregistrement donneur sont empruntées et utilisées pour apporter le nombre minimal de changements aux variables afin de résoudre tous les problèmes de réponses manquantes ou incohérentes. Le Système canadien de contrôle et d'imputation du recensement (SCANCIR) (voir le Guide de base de l'utilisateur, version 5.2 du SCANCIR) est le système automatisé qui a été utilisé pour presque toutes les imputations déterministes et les imputations par donneur à changements minimaux par le plus proche voisin pour le Recensement de 2011 et l'Enquête nationale auprès des ménages (ENM) de 2011.

2.9 Pondération

En 2011, le questionnaire du recensement contenait les huit questions qui figuraient sur le questionnaire abrégé du Recensement de 2006, ainsi que deux questions supplémentaires sur les langues. Ces questions ont été posées à l'ensemble de la population. L'information restante a été recueillie au moyen de l'Enquête nationale auprès des ménages, dont le questionnaire a été distribué à environ 30 % des ménages. On a fait appel à la pondération pour attribuer à l'ensemble de la population l'information recueillie auprès de l'échantillon de 30 %.

La méthode d'échantillonnage utilisée pour l'ENM de 2011 différait de celle appliquée pour le questionnaire complet du Recensement de 2006. Par conséquent, la méthode de pondération était différente également. La première étape de la pondération consistait à attribuer à chaque ménage un poids de base reflétant la probabilité qu'il soit échantillonné. Ces poids ont été ajustés pour tenir compte de la non-réponse totale. Un dernier ajustement, par le plus petit nombre que possible, a été effectué afin d'établir une meilleure concordance entre les estimations de l'échantillon et les chiffres du recensement pour un certain nombre de caractéristiques liées à l'âge, au sexe, à l'état matrimonial, à l'union libre, à la langue et à la taille du ménage. La méthode de pondération est décrite en détail au [chapitre 4](#).

3. Échantillonnage²

3.1 Échantillon de l'Enquête nationale auprès des ménages

L'Enquête nationale auprès des ménages (ENM) est une enquête-échantillon conçue pour recueillir des données démographiques et socioéconomiques sur la population canadienne. L'échantillon a été tiré de la liste des logements du Recensement de la population de 2011. Seuls les logements privés occupés et les ménages correspondants faisaient partie du champ de l'ENM. Donc, contrairement au recensement, tous les logements collectifs et les ménages vivant à l'extérieur du Canada (p. ex., diplomates, personnel militaire) étaient hors du champ de l'ENM. Au moment de la sélection de l'échantillon de l'ENM, on ne savait pas systématiquement quelles adresses étaient liées à des logements hors du champ de l'enquête, ce qui signifie que certains logements ont reçu un questionnaire de l'ENM par erreur. Après avoir établi qu'un logement était hors du champ de l'enquête, toute autre activité de collecte ou de traitement des données a été abandonnée. Le questionnaire de l'ENM a été distribué à environ 30 % des ménages privés³.

La proportion de logements à sélectionner pour l'ENM a été déterminée au niveau de l'unité de collecte (UC). La méthode de collecte du recensement (p. ex., envoi par la poste, listage/livraison, recensement par interview) utilisée pour une UC donnée a également été utilisée pour déterminer la fraction d'échantillonnage de cette unité pour l'ENM.

Tous les logements privés des UC pour lesquels les renseignements du recensement avaient été recueillis par interview ont été sélectionnés pour l'ENM, y compris les logements dans les réserves indiennes. Les secteurs de recensement par interview englobaient environ 200 000 logements privés occupés. Dans les secteurs de recensement par listage/livraison, les logements privés ont été échantillonnés au taux de un sur trois.

Dans les secteurs d'envoi par la poste, les logements privés ont été sélectionnés en appliquant les fractions d'échantillonnage qui avaient été calculées au niveau provincial/territorial. Ces fractions ont été calculées de manière à atteindre la taille d'échantillon national fixe souhaitée de 4,5 millions de logements, tout en obtenant des fractions de sondage provinciales/territoriales égales pour les secteurs combinés de recensement par listage/livraison et d'envoi par la poste. Le [tableau 3.1.1](#) présente les différentes fractions de sondage pour les provinces et territoires dans lesquels des questionnaires ont été envoyés par la poste⁴.

2. Pour l'historique de l'échantillonnage dans les recensements du Canada, consulter l'[annexe B](#).

3. Par ménage privé, on entend une personne ou un groupe de personnes (autres que des résidents étrangers) qui occupe un logement privé et qui ne possède pas de lieu de résidence habituel ailleurs au Canada.

4. Il n'y avait pas d'UC à envoi par la poste dans les Territoires du Nord-Ouest (T.N.-O) et au Nunavut. Dans les T.N.-O, dans les secteurs de recensement par listage/livraison, les logements privés ont été échantillonnés au taux de un sur trois, tandis que dans les UC pour lesquelles l'information de l'ENM a été recueillie par interview, tous les logements privés ont été sélectionnés. Au Nunavut, les renseignements de l'ENM ont été recueillis par interview pour la totalité (100 %) des logements privés.

Rapport technique sur l'échantillonnage et la pondération

Tableau 3.1.1 Fractions de sondage provinciales/territoriales pour les UC avec envoi par la poste

Province et territoire	Fraction de sondage pour l'envoi par la poste
Terre-Neuve-et-Labrador	0,1973
Île-du-Prince-Édouard	0,2091
Nouvelle-Écosse	0,2488
Nouveau-Brunswick	0,2678
Québec	0,2768
Ontario	0,2752
Manitoba	0,2640
Saskatchewan	0,2473
Alberta	0,2725
Colombie-Britannique	0,2772
Yukon	0,2621
Territoires du Nord-Ouest	...
Nunavut	...

... n'ayant pas lieu de figurer

Source : Statistique Canada, Enquête nationale auprès des ménages de 2011.

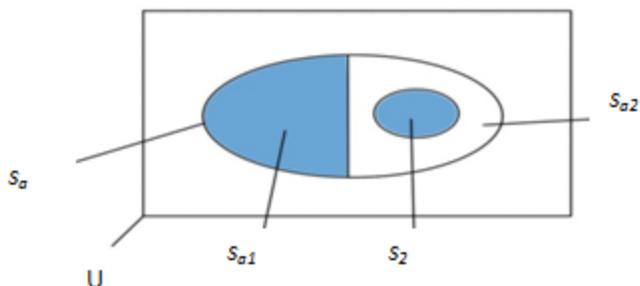
Dans les secteurs d'envoi par la poste et de listage/livraison où l'autodénombrement a été utilisé, les ménages échantillonnés ont été sélectionnés selon un plan d'échantillonnage aléatoire systématique stratifié (par province et territoire et par mode de collecte). Dans les unités de collecte par listage/livraison, un logement sur trois a été sélectionné. Dans les unités de collecte par envoi par la poste, les logements ont été sélectionnés en appliquant les fractions d'échantillonnage provinciales qui avaient été calculées de manière à 1) produire la taille totale d'échantillon souhaitée au niveau du Canada, et 2) produire la même fraction de sondage pour chaque province et territoire (modes de collecte par envoi par la poste et par listage/livraison combinés).

3.2 Sous-échantillon de l'ENM

Avant la collecte des données de l'ENM, il a été établi que les ressources disponibles pour le suivi des cas de non-réponse (SCNR) ne permettraient pas de faire le suivi de tous les non-répondants. Il a par conséquent été décidé de procéder, à une certaine date, au sous-échantillonnage des cas de non-réponse non traités. Cette décision a été prise afin de s'assurer que le nombre de cas devant être traités par le personnel sur le terrain soit gérable et afin de réduire le risque de biais de non-réponse⁵ si l'on permettait au personnel sur le terrain de se concentrer sur certains secteurs ou certains logements. Tous les cas non sélectionnés dans le sous-échantillon du SCNR ont par conséquent été exclus de toute activité de collecte ultérieure. Cette approche correspond à une forme d'échantillonnage appelée « échantillonnage à deux phases ». Les logements échantillonnés au départ pour l'ENM constituent l'échantillon de première phase, tandis que les logements sous-échantillonnés pour le SCNR constituent l'échantillon de deuxième phase. La [figure 3.2.1](#) illustre le plan d'échantillonnage. Le rectangle U représente les logements du recensement. L'échantillon de l'ENM est représenté par la grande ellipse s_a , où s_{a1} représente les logements qui avaient déjà répondu à l'ENM le 14 juillet 2011. La proportion restante de l'ovale, qui représente les logements qui n'avaient pas encore répondu à cette date, a été divisée en deux parties. Le sous-échantillon de non-répondants sélectionnés pour le SCNR est représenté par le petit ovale s_2 et les non-répondants non sélectionnés pour le SCNR se trouvent dans s_{a2} .

5. Par « biais », on entend la différence entre la moyenne des estimations, sur tous les échantillons possibles, et la valeur de population réelle.

Figure 3.2.1 Plan d'échantillonnage de l'ENM et du SCNR



Source : Statistique Canada, Enquête nationale auprès des ménages de 2011.

La méthode d'échantillonnage à deux phases a été proposée pour la première fois par Hansen et Hurwitz (1946) et s'est avérée être une méthode efficace quand le coût de collecte par unité à la deuxième phase est élevé. La méthode est également efficace lorsqu'on s'efforce de cibler un plus petit nombre de non-répondants pour le suivi des cas de non-réponse. L'objectif de l'ENM était de concentrer l'effort maximal raisonnable sur un plus petit nombre de cas sélectionnés pour le suivi au lieu de diluer l'effort en répartissant les ressources disponibles entre tous les cas de non-réponse. Parallèlement, la méthode de sous-échantillonnage pour le SCNR permet de tirer parti d'une combinaison de renseignements provenant de la base de sondage et de paradosées afin de répartir les ressources de façon efficace, ce qui donne lieu à une réduction du biais de non-réponse.

Les cas qui suivent ont été considérés comme inadmissibles au sous-échantillonnage pour le SCNR :

- Ceux qui avaient été reconnus comme des répondants au moment du sous-échantillonnage.
- Ceux pour lesquels un rendez-vous de SCNR avait déjà été pris parce qu'on estimait qu'il y avait une forte chance qu'ils deviennent des cas de réponses complètes dans un avenir proche.
- Ceux faisant partie des secteurs de recensement par interview.
- Ceux qui avaient été ajoutés à l'échantillon de l'ENM après la sélection initiale de l'échantillon, puisqu'aucun questionnaire n'avait été envoyé.
- Ceux correspondant à une ligne inutilisée ou à un logement figurant dans le Registre des visites qui auraient pu être définis avant le sous-échantillonnage comme un logement occupé par des résidents temporaires ou étrangers (RT/RE) seulement.

3.3 UC ciblées

L'un des avantages de la création d'un sous-échantillon pour le SCNR tient au fait que les non-répondants vivant dans certaines régions ou possédant certaines caractéristiques qui sont moins susceptibles de répondre à l'ENM peuvent être ciblés afin d'améliorer leur taux de réponse. Par conséquent, comme ces non-répondants ont habituellement une plus faible propension à répondre, il a été décidé de suréchantillonner certaines UC prédéterminées afin d'obtenir une bonne représentation de certaines populations dans le sous-échantillon de l'ENM. Le [tableau 3.3.1](#) présente la liste des cinq populations (domaines cibles) qui ont été suréchantillonnées parce qu'en se fondant sur d'autres enquêtes ou sur de précédents recensements, elles étaient considérées comme étant à risque de produire un taux de réponse plus faible et elles présentaient un grand intérêt pour des analyses statistiques éventuelles.

Rapport technique sur l'échantillonnage et la pondération

Tableau 3.3.1 Populations à risque désignées pour être suréchantillonnées dans le SCNR

Caractéristique	Description	Niveau
Autochtones (hors réserves)	Première Nation (Indiens de l'Amérique du Nord) – Identité unique, Métis – Identité unique, Inuk (Inuit) – Identité unique ou toute combinaison de ces trois caractéristiques	Personne
Nouveaux immigrants	Résidents non permanents et Canadiens ayant obtenu le statut d'immigrant entre 1996 et 2006	Personne
Minorités visibles	Tous les groupes de non-Blancs sauf les Chinois ⁶ et les Autochtones	Personne
Faible niveau de scolarité	Membres de la population active sans certificat d'études au-delà du diplôme d'études secondaires	Personne
Faible revenu	Ménage dont le revenu est compris entre 0 \$ ⁷ et 20 000 \$	Ménage

Source : Statistique Canada, Enquête nationale auprès des ménages de 2011.

La stratégie de suréchantillonnage a permis de déterminer des secteurs cibles qui contenaient une proportion relativement élevée de personnes ou de ménages appartenant à un ou à plusieurs domaines cibles. Les données non pondérées du Recensement de 2006 ont été utilisées pour créer les domaines cibles. Les logements recensés en 2006 ont été appariés à une UC de 2011, et seules les UC contenant des cas de collecte par envoi par la poste ou par listage/livraison ont été prises en considération pour le suréchantillonnage.

3.4 Sélection du sous-échantillon pour le SCNR

Afin de déterminer le sous-échantillon du SCNR, le pays a été subdivisé en 23 901 strates. Toute UC ciblée pour le suréchantillonnage a servi de strate; sinon, une strate correspondait habituellement au groupement de deux UC adjacentes appelées zones de dénombrement. Toutes les strates ont été représentées dans le sous-échantillon du SCNR. Si une strate contenait un ou deux logements admissibles, ces logements ont été ajoutés automatiquement au sous-échantillon. Seuls 196 logements appartenaient à une telle strate. Pour les strates comptant trois logements admissibles ou plus, le nombre de logements à sélectionner a été calculé de manière qu'il soit proportionnel à la taille de la strate. La taille du sous-échantillon pour les strates qui étaient désignées pour le suréchantillonnage a été augmentée d'un facteur 1,6.

Une fois que les tailles des sous-échantillons de strate ont été calculées, le sous-échantillon du SCNR a été sélectionné le 14 juillet 2011 par échantillonnage systématique avec un intervalle d'échantillonnage fractionnaire. En tout, 642 442 logements ont été sélectionnés dans le sous-échantillon du SCNR, ce qui représentait environ 33,5 % des logements de l'ENM admissibles restants. Les UC ciblées ont fourni 169 657 logements, c'est-à-dire 26,4 % du sous-échantillon total du SCNR. La répartition du sous-échantillon par province et territoire est illustrée au [tableau 3.4.1](#).

6. Les Canadiens de descendance chinoise ont été exclus du groupe des minorités visibles parce que leur taux de réponse est habituellement relativement élevé.

7. La valeur déclarée du revenu du ménage est parfois négative. Pour les besoins de l'échantillonnage, ces ménages ne sont pas considérés comme ayant un faible revenu.

Rapport technique sur l'échantillonnage et la pondération

Tableau 3.4.1 Répartition du sous-échantillon du SCNR par province et territoire

Province et territoire	Total
Terre-Neuve-et-Labrador	11 884
Île-du-Prince-Édouard	3 160
Nouvelle-Écosse	19 832
Nouveau-Brunswick	16 098
Québec	144 767
Ontario	244 309
Manitoba	23 770
Saskatchewan	22 666
Alberta	66 889
Colombie-Britannique	88 311
Yukon	613
Territoires du Nord-Ouest	143
Total	642 442

Source : Statistique Canada, Enquête nationale auprès des ménages de 2011.

4. Estimation de l'Enquête nationale auprès des ménages

Tout processus d'échantillonnage nécessite une méthode d'estimation connexe pour convertir les données d'échantillon à l'échelle de la population et pour s'assurer que les estimations d'après les données d'enquête sont représentatives de la population. Le choix d'une méthode d'estimation est généralement régi par des contraintes de nature opérationnelle ainsi que théorique. Du point de vue opérationnel, la méthode doit être applicable dans le système de traitement des données dont elle fait partie, et du point de vue théorique, la méthode doit réduire au minimum l'erreur statistique dans les estimations produites. La [section 4.1](#) décrit les considérations de nature opérationnelle et théorique pertinentes pour choisir la méthode d'estimation. Les [sections 4.2 à 4.6](#) portent sur les aspects détaillés de l'estimation dans le cas de l'Enquête nationale auprès des ménages (ENM) de 2011, y compris la définition de l'univers de l'ENM, les poids de sondage, les ajustements de la pondération pour tenir compte de la non-réponse totale, ainsi que les processus de calage des poids finaux.

4.1 Éléments à considérer pour choisir une méthode d'estimation

4.1.1 Considérations de nature opérationnelle

Mathématiquement, une méthode d'estimation peut être décrite au moyen d'une formule algébrique ou d'un estimateur qui montre comment l'estimation pour la population est calculée sous forme d'une fonction des valeurs d'échantillon observées et d'autres données provenant du plan de sondage ou de sources externes. La plupart du temps, cet estimateur est une simple fonction des poids et de la variable d'intérêt pour les unités répondantes. L'utilisation d'un jeu unique de poids pour produire toutes les estimations garantit un certain niveau de cohérence entre les diverses estimations de l'enquête.

Par conséquent, l'approche adoptée pour l'ENM (et la plupart des enquêtes-échantillon) consiste à subdiviser la méthode d'estimation en deux étapes : a) le calcul des poids (appelé procédure de pondération) et b) l'utilisation des poids pour produire les estimations, comme l'estimation d'une caractéristique particulière par totalisation des poids appliqués aux personnes ou aux ménages possédant la caractéristique. La plupart de la complexité mathématique est alors confinée à l'étape a) qui est exécutée une seule fois. L'étape b), quant à elle, est réduite à un simple processus, comme la sommation des poids chaque fois qu'une totalisation est nécessaire. Il convient de souligner que la cohérence entre les différentes estimations fondées sur les données d'échantillon est assurée étant donné que le poids appliqué à chaque unité de l'échantillon est le même pour toute totalisation faisant intervenir l'unité en question.

4.1.2 Considérations d'ordre théorique

Pour un plan de sondage donné et une méthode d'estimation donnée, il est possible, en s'appuyant sur la théorie de l'échantillonnage, de formuler un énoncé quant aux chances qu'un certain intervalle contienne la valeur de population inconnue qu'il faut estimer. Le principal critère relatif au choix d'une méthode d'estimation est la minimisation de l'étendue de ces intervalles pour un niveau donné de confiance, de façon que les énoncés au sujet des valeurs inconnues de population soient aussi précis que possible. La mesure de précision habituelle pour comparer les méthodes d'estimation est appelée erreur type. Pourvu que certaines conditions soient respectées, les intervalles de plus ou moins deux erreurs types par rapport à la valeur estimée contiendront la valeur de la population pour approximativement 95 % de tous les échantillons possibles.

Outre la minimisation de l'erreur type, un second objectif relatif au choix d'une méthode d'estimation pour l'échantillon de l'ENM, est de garantir autant que possible que les estimations d'après l'échantillon des caractéristiques évaluées au recensement concordent avec les valeurs de recensement connues. Heureusement, ces deux objectifs sont habituellement complémentaires en ce sens que l'erreur d'échantillonnage a tendance à être réduite lorsque l'on veille à ce que les estimations de certaines caractéristiques de base de l'échantillon concordent avec les chiffres du recensement correspondants. Cependant, bien que cela soit vrai en général, forcer les estimations de l'échantillon de l'ENM de caractéristiques évaluées au recensement à concorder avec les chiffres du recensement correspondants pour de très petits sous-groupes peut avoir un effet indésirable sur l'erreur type des estimations pour les caractéristiques de l'échantillon proprement dites.

Si l'on ne dispose d'aucune information au sujet de la population échantillonnée à part celle recueillie pour les unités échantillonnées et en l'absence de non-réponse totale, la méthode d'estimation est limitée à la pondération des unités

échantillonnées par l'inverse de leurs probabilités de sélection. Par exemple, si une unité avait une chance sur trois d'être sélectionnée, elle recevrait un poids de 3. En présence de non-réponse totale, le poids doit faire l'objet d'un ajustement supplémentaire en utilisant la probabilité de réponse de l'unité. En outre, en pratique, on dispose presque toujours de certains renseignements supplémentaires au sujet de la population (p. ex., sa taille totale et, parfois, la ventilation en fonction d'une certaine variable – peut-être par province et territoire). Ce genre d'information peut être utilisé pour améliorer la formule d'estimation afin de produire des estimations ayant plus de chance d'être proches de la valeur de population inconnue. Dans le cas de l'échantillon de l'ENM, une grande quantité de renseignements très détaillés au sujet de la population échantillonnée était disponible sous forme de données de recensement à chaque niveau géographique. Nous pouvons profiter de cette riche information sur la population pour améliorer les estimations faites sur l'échantillon de l'ENM. Ce point sera discuté plus loin dans ce rapport.

4.1.3 Autres considérations concernant la pondération des données de l'Enquête nationale auprès des ménages

Comme dans le cas des recensements précédents, un poids a été attribué à chaque ménage ayant participé à l'ENM afin que les caractéristiques des répondants à l'ENM puissent être pondérées et combinées pour produire des estimations pour la population. Cependant, au moins deux problèmes importants ont rendu le processus plus complexe en 2011. Premièrement, la collecte des données du recensement et des données de l'ENM représente des événements distincts. Le fait d'utiliser deux jeux de questionnaires administrés à deux moments différents a donné lieu à des complications telles que l'appariement des ménages et des incohérences entre les données provenant des deux sources. Ce problème a été mentionné à la [section 2.9](#). Deuxièmement, l'ENM a dû composer avec un taux de non-réponse des ménages relativement élevé. Le problème de la non-réponse totale et la conception du processus de suivi des cas de non-réponse (SCNR) en vue d'améliorer le taux de réponse sont décrits au [chapitre 3](#).

Dans une enquête, la méthode de pondération des unités échantillonnées par l'inverse de leur probabilité de sélection (ou de leurs probabilités de réponse en présence d'un ajustement pour la non-réponse totale) devrait aboutir à de faibles différences entre les estimations de l'échantillon et les chiffres de recensement pour les grands sous-groupes de la population. Cependant, les différences pourraient être importantes pour les sous-groupes plus petits. Dans l'ENM, ces différences ont généralement été accentuées par les cas de non-réponse totale.

Il est difficile de faire concorder les estimations de l'échantillon de l'ENM pour les caractéristiques du recensement avec les chiffres du recensement à chaque niveau géographique. Les différences entre les estimations de l'échantillon et les chiffres du recensement deviennent visibles lorsqu'on produit un tableau croisé d'une variable-échantillon et de la variable de recensement correspondante. Le tableau des totaux estimés d'après l'échantillon pour des caractéristiques particulières ne concordera pas nécessairement avec le tableau équivalent des totaux des chiffres du recensement pour ces caractéristiques.

L'ajustement aussi faible que possible des poids correspondant à l'inverse des probabilités de réponse afin d'obtenir une concordance parfaite entre les estimations et les chiffres du recensement pour certains sous-groupes porte le nom de « calage ». Cette procédure est discutée plus en détail à la [section 4.6](#).

4.2 Univers de l'ENM

L'univers des ménages du recensement a été divisé en trois parties : les ménages privés, les ménages collectifs et les ménages à l'extérieur du Canada. L'univers des ménages de l'ENM correspond aux ménages privés qui étaient admissibles au Recensement de 2011. Sauf indication contraire, le terme « dans le champ de l'enquête » sera utilisé dans le présent document pour indiquer qu'un ménage fait partie de l'univers de l'ENM (c.-à-d. les ménages privés), tandis que le terme « hors du champ de l'enquête » fait référence aux ménages non compris dans l'univers (c.-à-d. les ménages collectifs et à l'extérieur du Canada).

Les ménages privés de cinq subdivisions de recensement (SDR) correspondant à cinq réserves indiennes représentent une exception à l'univers de l'ENM. Ces ménages ont été exclus de l'univers en raison d'un très faible taux de réponse à l'ENM. S'ils ne l'avaient pas été, les régions environnantes dans les mêmes régions de pondération (RP) auraient été très affectées par la non-réponse et les ajustements des poids par calage qui auraient été nécessaires pour compenser les faibles taux de réponse. L'exclusion de ces cinq SDR est un exemple des différences que l'on peut observer lorsque l'on compare une publication fondée sur les données du recensement à une publication

Rapport technique sur l'échantillonnage et la pondération

fondée sur les données de l'ENM. Le nombre de ménages privés et la population de ces cinq SDR sont présentés au [tableau 4.2.1](#).

Tableau 4.2.1 Subdivisions de recensement non incluses dans l'univers de l'ENM

Province	Division de recensement	Subdivision de recensement	Nombre de ménages privés	Population
Québec	La Haute-Côte-Nord	Essipit	111	268
Ontario	Cochrane	Factory Island 1	384	1 414
Ontario	Kenora	Sandy Lake 88	459	1 861
Manitoba	Division n° 19	The Narrows 49	337	826
Saskatchewan	Division n° 14	Opaskwayak Cree Nation 27A (Carrot River)	72	286
Total			1 363	4 655

Sources : Statistique Canada, Recensement de 2011 et Enquête nationale auprès des ménages de 2011.

4.3 Poids de sondage

À chaque logement de l'échantillon de première phase (c.-à-d. sélectionné pour l'ENM), on a attribué un poids de base de première phase égal à l'inverse de la probabilité de sélection du logement dans l'échantillon de première phase. À chaque logement sous-échantillonné dans l'échantillon de deuxième phase (c.-à-d. sélectionné pour le suivi des cas de non-réponse (SCNR)), on a aussi attribué un poids de base de deuxième phase égal à l'inverse de la probabilité de sélection de ce logement dans le sous-échantillon.

Les poids de sondage ont, en général, été calculés comme suit :

1. Aux logements qui n'étaient pas admissibles au sous-échantillonnage pour le SCNR, on a attribué un poids de sondage égal au poids de l'échantillon de première phase. La grande majorité de ces logements correspondait à des ménages identifiés comme ayant répondu à l'ENM.
2. Aux logements qui étaient admissibles au sous-échantillonnage pour le SCNR et qui n'ont pas été sélectionnés pour le SCNR, on a attribué un poids de sondage dont la valeur était fixée à 0.
3. Aux logements qui ont été sous-échantillonnés pour le SCNR, on a attribué un poids de sondage égal au produit de leurs poids de base de première et de deuxième phase.

4.4 Ajustement pour la non-réponse totale

Il existe différentes catégories de ménages non répondants à l'ENM. Les ménages qui avaient été désignés comme ayant répondu au moment du sous-échantillonnage n'étaient pas admissibles à la sélection du sous-échantillon. Cependant, après un examen minutieux des questionnaires qu'ils avaient remplis, on a constaté que certains d'entre eux avaient retourné le questionnaire, mais n'avaient en fait pas fourni les réponses et étaient donc effectivement des non-répondants. En outre, de nombreux logements sélectionnés pour le SCNR n'ont pas fourni les renseignements requis malgré les tentatives en vue d'obtenir une réponse. Le taux de réponse global non pondéré était de 68,6 % et le taux de réponse pondéré était de 77,2 %.

Diverses stratégies de traitement de la non-réponse totale faisant appel aux données auxiliaires disponibles pour les répondants et les non-répondants ont été étudiées. Les méthodes d'imputation étaient intéressantes dans le contexte de l'ENM étant donné que les données de recensement étaient disponibles pour la grande majorité des non-répondants. Autrement dit, la non-réponse totale à l'ENM peut être considérée comme un problème de non-réponse partielle. Malheureusement, les méthodes d'imputation n'ont pas toujours été fructueuses, car le grand nombre de variables d'appariement rendait souvent impossible la découverte d'un donneur parfait. Un donneur parfait

Rapport technique sur l'échantillonnage et la pondération

serait un ménage répondant ayant la même valeur que le ménage non-répondant pour chaque variable de recensement utilisée pour l'appariement. Cette situation a donné lieu à une augmentation importante de l'occurrence dans les données de combinaisons des variables du recensement et des variables imputées non issues du recensement qui sont rares dans la population. Après une analyse approfondie, ces stratégies d'imputation ont été écartées et il a été décidé d'attribuer un poids nul aux ménages non répondants qui avaient été désignés au départ comme ayant répondu, ainsi qu'aux ménages du sous-échantillon qui n'ont pas répondu. Les poids qu'ils auraient eus s'ils avaient répondu ont été transférés à leurs plus proches voisins. La méthode, qui est une approche de repondération, peut être subdivisée en trois grandes étapes.

La première étape consistait à déterminer les variables auxiliaires prédisant le mieux la propension à répondre des ménages. De nombreuses variables auxiliaires provenant du recensement et de l'ENM, ainsi que de couplages avec les fichiers de données fiscales de 2010, le Registre des Indiens, et les fichiers d'immigration pour la période de 1980 à 2011, ainsi que les variables se rapportant aux réserves indiennes (s'il y avait lieu) ont été prises en considération pour construire les modèles de régression logistique au niveau du Canada par une méthode de sélection ascendante des variables. Ensuite, un poids relatif a été appliqué à chacune des variables auxiliaires en fonction de leur pouvoir prédictif dans le modèle de réponse qui avait été sélectionné. Même si certaines variables auxiliaires n'ont pas été incluses dans les modèles finaux, un poids minimal positif a été attribué à chacune.

À la deuxième étape, le Système canadien de contrôle et d'imputation du recensement (SCANCIR) (voir le Guide de base de l'utilisateur de la version 5.2 du SCANCIR) a été utilisé pour repérer les 20 ménages répondants se trouvant à proximité qui correspondaient le mieux à chaque ménage non-répondant. Les ménages choisis comprenaient le même nombre de membres et résidaient habituellement dans le même quartier ou dans un quartier proche du ménage non-répondant. Le processus d'appariement a été effectué en se servant des variables auxiliaires sélectionnées à la première étape et de leur poids relatif. Un score d'appariement compris entre 0 et 1 a été attribué à chaque variable d'appariement, un score de 0 étant observé si les valeurs étaient les mêmes et une valeur supérieure à 0 (habituellement 1) étant attribuée si les valeurs étaient différentes. Ces scores ont été multipliés par le poids d'imputation de la variable dans le SCANCIR et totalisés sur l'ensemble des variables pour calculer un score de distance entre le non-répondant et le répondant. Les 20 répondants dont le score était le plus faible ont été identifiés.

La troisième étape consistait à transférer le poids du ménage non-répondant à chacun des 20 ménages répondants identifiés à la deuxième étape. La part du poids attribuée était proportionnelle à l'inverse de la distance entre le ménage répondant et le ménage non-répondant. Par conséquent, les ménages répondants correspondant mieux aux ménages non-répondants ont reçu une part plus importante du poids qui était transféré.

4.5 Répondants-surprises

Les répondants-surprises sont les ménages occupant un logement sélectionné dans l'échantillon de première phase qui n'avaient pas répondu avant la sélection du sous-échantillon, mais qui ont répondu par après sans faire partie du sous-échantillon du SCNR. Les répondants-surprises ont été regroupés avec les ménages répondants dans l'échantillon de deuxième phase. Leur poids de sondage a été fixé à 0 comme il est décrit à la [section 4.3](#). Au lieu de laisser à ces répondants-surprises leur poids initial de valeur 0, celui-ci a été forcé afin d'être égal à 1 et le poids de certains répondants inclus dans le sous-échantillon a été réduit en conséquence. Cette mesure a été prise parce qu'un poids de 1 réduisait au minimum le biais tout en s'assurant que les répondants-surprises soient autoreprésentés.

Comme dans le cas de l'ajustement pour la non-réponse totale, le SCANCIR a été utilisé pour déterminer l'ensemble de voisins donneurs correspondant aux 20 ménages répondants les plus proches dans l'échantillon de deuxième phase pour chaque répondant-surprise. Un poids de 1/20 a ensuite été soustrait du poids de ces répondants voisins et transféré au répondant-surprise.

4.6 Calage

Après avoir été ajustés pour tenir compte de la non-réponse totale et des répondants-surprises, les poids de sondage de l'ENM ont été ajustés une dernière fois durant le processus de calage. Ce processus consistait à ajuster légèrement les poids afin qu'ils satisfassent à un certain nombre de contraintes. Tout en assurant la cohérence entre

Rapport technique sur l'échantillonnage et la pondération

les totaux estimés d'après l'ENM et les chiffres du recensement pour de nombreuses contraintes, le calage a pour effet de réduire la variabilité d'échantillonnage des estimations dérivées de l'enquête. Durant cette dernière étape d'ajustement, il était important de limiter l'intervalle de valeurs des poids calés pour éviter qu'une part excessive de poids ne soit appliquée à un ménage ou à une personne. Par conséquent, les poids ont été contraints de prendre une valeur comprise entre 1 et 100.

L'[annexe C](#) énumère toutes les contraintes qui ont été prises en considération durant le processus de calage. Elles comprenaient les mêmes contraintes qu'en 2006, ainsi que des contraintes supplémentaires faisant intervenir des variables de la famille du recensement et de la famille économique ainsi que des variables de langue. Les caractéristiques pour lesquelles des données provenant du recensement et de l'ENM étaient disponibles et pour lesquelles on s'est efforcé d'établir une concordance comprennent l'âge, le sexe, l'état matrimonial, l'union libre, la taille du ménage, le type de logement, la langue maternelle, la langue parlée à la maison, etc.

Le choix des contraintes et le calage ont été exécutés indépendamment dans les régions géographiques appelées régions de pondération (RP). Le Canada comptait 5 884 RP, dont la plupart étaient constituées d'une zone contiguë de terrain à l'intérieur d'une division de recensement (DR). En moyenne, une RP contenait environ 2 300 ménages au recensement, et la plupart des RP contenaient de 300 à 699 ménages qui étaient soit des répondants initiaux à l'ENM ou des ménages qui avaient été sélectionnés pour le SCNR. Voir le [chapitre 6](#) pour des renseignements supplémentaires sur la conception des RP.

La première étape du processus de calage consistait à classer les contraintes en trois groupes :

1. Contraintes obligatoires – Ces contraintes doivent être utilisées dans le calage parce qu'il faut absolument qu'il y ait concordance entre les chiffres du recensement et les estimations de l'ENM aux niveaux géographiques qui sont des agrégats habituels des RP (p. ex., Canada, provincial et territorial). Le nombre de personnes dans la RP et le nombre de ménages dans la RP étaient les deux contraintes obligatoires.
2. Contraintes à faible réponse – Les contraintes qui seraient appliquées à moins de 30 personnes ne doivent pas être utilisées dans le calage parce qu'elles peuvent rendre les estimations de l'enquête instables.
3. Toutes les autres contraintes – Ces contraintes ont été examinées plus en détail afin de déterminer s'il fallait les utiliser dans le calage.

La deuxième étape consistait à déterminer quelles contraintes du troisième groupe devaient être utilisées dans le processus de calage en plus des contraintes obligatoires. Les contraintes de ce troisième groupe ont été ajoutées une par une en commençant par celles qui étaient moins linéairement dépendantes des contraintes à faible réponse et des contraintes déjà incluses. Les contraintes qui étaient trop linéairement dépendantes n'ont pas été incluses.

Le calage a ensuite été exécuté sur l'ensemble final de contraintes de l'étape 2 en ajustant les poids des ménages aussi peu que possible de manière que la somme des estimations pondérées soit égale aux chiffres du recensement pour ces contraintes. En pratique, le processus de calage a été exécuté en se servant du logiciel StatMx développé par Statistique Canada, qui était programmé entièrement en langage SAS. Voir Verret (2013) pour des renseignements plus détaillés.

Il existe quelques raisons pour lesquelles les estimations de l'échantillon peuvent différer des chiffres du recensement, particulièrement pour les petites régions, même après l'étape du calage. Ces principales raisons sont énumérées ci-dessous.

1. Contraintes supprimées durant le processus de calage : Comme il est décrit plus haut, des contraintes pouvaient être supprimées parce que le nombre de réponses était faible, parce qu'elles étaient linéairement dépendantes d'autres contraintes choisies ou parce qu'elles étaient quasi linéairement dépendantes des contraintes pour lesquelles le nombre de réponses était faible. Le calage n'a pas été effectué sur les contraintes supprimées, ce qui a généralement donné lieu à des différences entre les chiffres du recensement et les estimations de l'ENM pour ces variables.

Rapport technique sur l'échantillonnage et la pondération

2. Sous-région de pondération : La RP était la région géographique la plus petite pour laquelle le système de pondération a essayé d'obtenir une concordance entre les chiffres du recensement et les estimations de l'ENM pour un nombre de variables auxiliaires aussi grand que possible. Toute entité plus petite qu'une RP, comme la plupart des aires de diffusion (AD), est appelée sous-RP. Des écarts pouvaient exister entre les chiffres du recensement et les estimations de l'ENM pour ces sous-RP.

Le point qui précède est important dans les régions où le taux de non-réponse est élevé. Cela tient au fait que les RP devaient être plus grandes que d'habitude pour contenir 300 à 699 ménages répondants. Par conséquent, certaines petites collectivités ou municipalités où les taux de réponse étaient faibles étaient des sous-RP, ce qui signifie que la concordance de leurs estimations avec les chiffres du recensement ne peut pas être garantie.

5. Indicateurs du biais de non-réponse de l'Enquête nationale auprès des ménages

5.1 Introduction

En raison du caractère volontaire de l'ENM, de nombreux efforts ont été mis en place par Statistique Canada pour maximiser la participation des ménages, notamment au niveau de la collecte, à l'aide de méthodes de collecte mixte ainsi que la mise en place de la méthodologie de collecte par vague (voir le [chapitre 1](#) sur la collecte). Malgré ces efforts, et un taux de réponse équivalent à celui des autres enquêtes volontaires menées par Statistique Canada (68,6 %), un peu moins d'un ménage sur trois n'a pas participé. À l'étape de traitement, la non-réponse totale a été traitée en utilisant une méthodologie de repondération. Le but de la repondération est principalement de réduire le biais de non-réponse dans les estimations.

La non-réponse a deux effets possibles sur les données : d'une part, elle contribue à faire augmenter la variance totale des estimations, car la taille observée de l'échantillon est réduite par rapport à la taille initialement prévue. D'autre part, elle peut introduire un biais dans les estimations lorsque les non-répondants diffèrent des répondants par rapport aux caractéristiques mesurées⁸. On parle alors de biais dû à la non-réponse. Par conséquent, il est recommandé pour évaluer la qualité des données, de déterminer s'il existe un biais potentiel dû à la non-réponse et, si possible, dériver un indicateur de son ampleur.

L'indicateur de biais de non-réponse est un outil qui tente d'identifier la présence d'un biais résiduel potentiel, c.-à-d. après la repondération. Le biais de non-réponse survient lorsque les caractéristiques d'intérêt des répondants diffèrent de celles des non-répondants. Il ne diminue pas nécessairement avec une augmentation de la taille de l'échantillon contrairement à la variance due à l'échantillonnage.

Un indicateur de biais de non-réponse positif indique que l'estimateur surestime en moyenne le paramètre d'intérêt et un indicateur de biais de non-réponse négatif indique une sous-estimation. L'amplitude de l'indicateur du biais de non-réponse indique l'étendue du biais potentiel dans l'estimation. Ainsi, une grande valeur de l'indicateur du biais de non-réponse peut indiquer des différences marquées entre les caractéristiques des répondants et celles des non-répondants.

5.2 Méthodologie des indicateurs de biais de la non-réponse

Les indicateurs de biais dû à la non-réponse totale ont été produits à partir d'un fichier des données couplées des recensements de 2011 et de 2006. En se servant du nom de famille, de l'adresse et de la date de naissance, 73 % des répondants au Recensement de 2011 ont été appariées à leurs réponses du Recensement de 2006. Cela comprenait les réponses au questionnaire complet de 2006 pour une grande partie de l'échantillon de l'ENM de 2011, y compris les répondants et les non-répondants à l'ENM.

Les poids obtenus pour l'échantillon apparié ont été calés sur les estimations du nombre de personnes dans la population commune au Recensement de 2006 et au Recensement de 2011. Ce calage a été exécuté séparément pour les répondants appariés et les non-répondants appariés à l'ENM de 2011. Une deuxième série de poids a été calculée pour les répondants appariés. Ces poids ont été produits en utilisant les facteurs d'ajustement pour la non-réponse totale de 2011 et ont été ajustés aux estimations du nombre total de personnes dans la population commune au Recensement de 2006 et au Recensement de 2011. En utilisant les deux séries de poids, des totaux étalonnés ont été obtenus pour l'ensemble de l'échantillon apparié (c.-à-d. les répondants et les non-répondants à l'ENM de 2011) en utilisant les données du questionnaire complet du Recensement de 2006, et les totaux ajustés ont été obtenus à partir de l'échantillon de répondants appariés, de nouveau en utilisant les données du questionnaire complet de 2006.

8. Statistique Canada (2009). Lignes directrices concernant la qualité, 5^e édition, produit n°12-539-XWF au catalogue.

Rapport technique sur l'échantillonnage et la pondération

La différence entre le total ajusté et le total étalonné a servi d'indicateur du biais dû à la non-réponse totale dans l'ENM de 2011. Algébriquement, pour la catégorie k de la variable x_j de 2006, cet indicateur de biais peut être exprimé comme $\hat{B}(\hat{X}_{jk}^a) = \hat{X}_{jk}^a - \hat{X}_{jk}$ avec le total étalonné $\hat{X}_{jk} = \sum_{i \in s_m} w_i x_{ijk}$ et le total ajusté $\hat{X}_{jk}^a = \sum_{i \in s_{m,r}} a_i w_i x_{ijk}$, où, pour l'unité i , a_i est un ajustement pour la non-réponse totale, w_i est un poids calé, x_{ijk} est la valeur de la variable j pour la catégorie k , s_m est l'échantillon apparié et $s_{m,r}$ est l'ensemble des répondants de 2011 dans l'échantillon apparié.

5.3 Limites des indicateurs de biais de la non-réponse de l'ENM

Ces indicateurs présentent plusieurs limites. L'indicateur reflète le biais dû à la non-réponse totale et, donc, ne tient pas compte de la non-réponse partielle. Il s'agit d'un indicateur qui a trait à la population commune aux recensements de 2006 et de 2011. Autrement dit, il est fait abstraction du biais possible dû aux naissances, aux nouveaux immigrants (moins de 5 ans) et aux résidents non permanents (moins de 5 ans). L'indicateur est fondé sur un échantillon apparié, si bien que les unités non appariées qui faisaient partie de la population commune ne sont pas représentées. L'ajustement des poids qui tient compte de la partie non appariée est vraisemblablement imparfait. L'indicateur est fondé sur les variables de 2006 qui sont des mesures indirectes pour les variables de 2011. L'indicateur manque de précision pour les régions géographiques pour lesquelles l'échantillon apparié est de petite taille, c'est-à-dire souvent les petites régions géographiques. Enfin, la fiabilité de l'indicateur est également affectée par la non-réponse au questionnaire abrégé et au questionnaire complet du Recensement de 2006, ainsi qu'au Recensement de 2011. En résumé, les indicateurs de biais sont exactement ce que leur nom indique, des indicateurs et non des mesures ou des estimations du biais.

5.4 Quelques indicateurs de biais de la non-réponse de l'ENM

À titre illustratif, les indicateurs de biais de la non-réponse présentés dans le [tableau 5.4.1](#) ont été calculés au niveau du Canada pour quelques caractéristiques présentant potentiellement un risque de biais plus important :

- Statut d'immigration (immigrant reçu, résident non-permanent)
- Identité autochtone (Première Nation (Indien de l'Amérique du Nord) - Identité unique, Métis - Identité unique, Inuk (Inuit) – Identité unique)
- Minorité visible (Sud-Asiatique, Chinois, Noir, Philippin, Latino-Américain)
- Plus haut certificat, diplôme ou grade (aucun certificat, diplôme ou grade, diplôme d'études secondaires ou l'équivalent, certificat d'une école de métiers ou diplôme d'études collégiales⁹, baccalauréat).

9. Cette catégorie comprend les personnes dont le plus haut certificat, diplôme ou grade est un certificat ou diplôme d'apprenti inscrit ou d'une école de métiers, ou un certificat ou diplôme d'un collège, d'un cégep ou d'un autre établissement non universitaire.

Rapport technique sur l'échantillonnage et la pondération

Tableau 5.4.1 Indicateurs de biais au niveau du Canada pour les caractéristiques « Statut d'immigration », « Identité autochtone », « Minorités visibles » et « Plus haut certificat, diplôme ou grade »

Caractéristique	Indicateur de biais en (%)
Statut d'immigration	
Immigrant reçu	-1,2
Résident non permanent	-0,8
Identité autochtone	
Première Nation (Indien de l'Amérique du Nord) – Identité unique	1,7
Métis – Identité unique	0,2
Inuk (Inuit) – Identité unique	6,1
Minorité visible	
Sud-Asiatique	1,3
Chinois	0,9
Noir	-3,4
Philippin	7,3
Latino-Américain	-1,1
Plus haut certificat, diplôme ou grade	
Aucun certificat, diplôme ou grade	-0,7
Diplôme d'études secondaires ou l'équivalent	-0,9
Certificat d'une école de métiers ou Diplôme d'études collégiales	0,0
Baccalauréat	1,9

Source : Statistique Canada, Enquête nationale auprès des ménages de 2011.

6. Régions de pondération

Les régions de pondération (RP) sont les subdivisions géographiques qui ont été utilisées durant les procédures de pondération de l'Enquête nationale auprès des ménages (ENM). Pour l'ENM de 2011, le Canada a été subdivisé en 5 884 RP. Le processus de pondération visait à obtenir la concordance entre certaines estimations de l'échantillon et les chiffres du recensement correspondants pour chaque RP. Certaines modifications ont été apportées à la conception des RP en 2011, parce que le caractère non obligatoire de la participation à l'ENM a produit différents taux de réponse pour différentes parties du pays. En 2011, les RP ont été conçues en deux étapes : la formation des RP ordinaires et l'ajustement pour les secteurs de recensement par interview et les réserves.

6.1 Formation des RP ordinaires

La première étape de la création d'un ensemble de RP consistait à regrouper des aires de diffusion (AD) tout en satisfaisant aux conditions suivantes :

- a) Une RP doit respecter les limites des divisions de recensement (DR) (c.-à-d. qu'une RP peut seulement être trouvée dans une seule DR).
- b) Une RP doit, dans la mesure du possible, contenir de 300 à 699 ménages qui ont répondu à l'ENM ou qui ont été sélectionnés dans le sous-échantillon (répondants ainsi que non-répondants du sous-échantillon). Pour simplifier, dans le présent chapitre, ces ménages seront considérés comme des ménages répondants.
- c) Une RP doit, dans la mesure du possible, respecter (par ordre de priorité) les limites des subdivisions de recensement (SDR) et celles des secteurs de recensement (SR).
- d) Une RP doit, dans la mesure du possible, être formée d'AD contiguës (c.-à-d. ne pas être fractionnée en deux parties ou plus ni contenir des « trous ») et doit être aussi compacte que possible.

Les conditions susmentionnées sont très semblables à celles appliquées en 2006, la principale exception étant la condition b). En 2006, la taille de la RP était fondée sur le nombre de ménages au recensement qui étaient soumis à l'échantillonnage, nombre qui dans la plupart des cas était compris entre 1 000 et 3 000 ménages. Si les RP avaient été créées de la même façon en 2011, le nombre de ménages répondants aurait varié fortement d'une RP à l'autre. Le calage n'aurait pas été efficace dans certaines RP à faible taux de réponse.

Pour remédier à ce problème en 2011, les RP ont été créées en se fondant avant tout sur le nombre de ménages répondants, même si cela signifiait que les RP dans les régions à faible taux de réponse devraient contenir un nombre plus élevé de logements dans le champ de l'enquête que par le passé. Au Recensement de 2006, la fraction de sondage était d'environ un sur cinq, ce qui signifie que chaque RP comprenant de 1 000 à 3 000 ménages dans le champ de l'enquête aurait compté environ de 200 à 600 ménages répondants dans l'échantillon du questionnaire complet. Cet intervalle a été augmenté légèrement en 2011 afin d'inclure de 300 à 699 ménages pour tenir compte de la non-réponse supplémentaire et obtenir une plus grande concordance entre les chiffres du recensement et les estimations de l'ENM. Une augmentation plus importante de la taille des RP pourrait entraîner un nombre significativement plus élevé des conflits avec les limites des SDR mentionnés à la condition c).

La quatrième condition était généralement facile à appliquer dans les régions urbaines. Cependant, l'obligation d'avoir des AD contiguës a dû être retirée dans les régions où deux enclaves disjointes ou plus de terrain appartenaient à la même municipalité. La condition a également été retirée quand une DR contenait plusieurs municipalités de petite taille ou à faible taux de réponse qui devaient être regroupées dans une seule RP afin de satisfaire à la condition b). Cette situation s'est produite le plus fréquemment dans les régions comprenant des réserves et dans le Nord.

L'algorithme qui a été utilisé en 2011 à la première étape du processus de formation des RP était le même qu'en 2006. Dans les deux cas, de nombreux ajustements manuels ont été appliqués aux RP anormales en divisant, en fusionnant ou en réalignant les RP afin de mieux respecter les conditions décrites plus haut. Après les ajustements manuels de la première étape en 2011, 99,3 % de toutes les RP contenaient le nombre souhaité de ménages, c'est-à-dire le même

Rapport technique sur l'échantillonnage et la pondération

taux que celui observé en 2006. Le [tableau 6.1.1](#) présente certaines statistiques indiquant dans quelle mesure les RP sont bien ajustées aux SDR et aux SR. Chacune des catégories est décrite après le tableau.

Tableau 6.1.1 Nombre de subdivisions de recensement et de secteurs de recensement qui respectent les limites des régions de pondération, première étape, Recensement de 2011

Scénario	Description	SDR		SR	
		Nombre	Pourcentage (%)	Nombre	Pourcentage (%)
1	La SDR ou le SR était de taille suffisamment petite pour entrer entièrement dans une RP, et la même RP ne comprenait que des SDR ou des SR entiers. Aucune SDR ni aucun SR dans la RP ne faisait aussi partie d'une autre RP.	3 729	70,6	2 408	44,1
2	La SDR ou le SR était de taille suffisamment petite pour entrer entièrement dans une RP, mais une autre SDR ou un autre SR dans la même RP faisait aussi partie d'une autre RP.	669	12,7	203	3,7
3	La SDR ou le SR était de taille suffisamment grande pour contenir des RP entières. Aucune des RP ne faisait aussi partie d'une autre SDR ou d'un autre SR.	674	12,8	2 725	49,9
4	La SDR ou le SR faisait partie d'au moins deux RP.	181	3,4	129	2,4
Total		5 253	100,0	5 465	100,0

Source : Statistique Canada, Recensement de 2011.

Dans le scénario 1, la condition c) était satisfaite. Ce scénario se produisait fréquemment pour les SDR, parce qu'il existait de nombreuses municipalités très petites, telles que des réserves indiennes et des villages qui comptaient une population plus faible que celle nécessaire pour créer une RP. Dans le scénario 2, la condition c) n'est pas satisfaite. Dans le scénario 3, la condition c) est satisfaite. Dans le scénario 4, la condition c) n'est pas satisfaite.

6.2 Ajustement des RP pour les régions recensées par interview et les réserves

La deuxième étape de création des RP était une nouveauté en 2011. Afin de créer des RP plus homogènes, les ménages auprès desquels les données ont été recueillies par un intervieweur ont été traités ensemble. Ces ménages ont été retirés de leurs RP originales et assignés dans un petit nombre de RP spéciales appelées RP recensées par interview. Afin de créer des RP contenant de 300 à 699 ménages répondants, la contrainte voulant qu'ils se trouvent dans la même DR et parfois dans la même province ou le même territoire a été retirée. De même, les ménages vivant dans les réserves indiennes ont été traités ensemble avec la même tribu autochtone dans la mesure du possible. Les ménages qui résidaient dans les réserves indiennes, à l'exclusion des terres qui avaient été louées pour des logements non autochtones, ont été assignés dans un petit nombre de RP spéciales appelées RP pour réserves. La plupart des RP pour réserves contenaient aussi de 300 à 699 ménages répondants et pouvaient dépasser les limites de la DR et parfois les limites de la province ou du territoire. En fait, 15 des 5 884 RP chevauchaient les limites de la province ou du territoire et faisaient partie d'exactly deux provinces ou territoires, et une RP figurait simultanément en Colombie-Britannique, en Alberta et dans les Territoires du Nord-Ouest. Le [tableau 6.2.1](#) montre le nombre de RP qui chevauchaient les limites de la DR, et en particulier le nombre de DR touchées.

Tableau 6.2.1 Nombre de RP associées à plusieurs divisions de recensement

Nombre de DR faisant partie d'une RP	Nombre de RP
1	5 793
2	58
3	20
4	8
5	2
6	1
7	0
8	1
9	0
10	0
11	1
Total	5 884

Source : Statistique Canada, Enquête nationale auprès des ménages de 2011.

Le [tableau 6.2.1](#) montre que 5 793 des 5 884 RP se trouvaient dans une seule DR, tandis que 58 RP se retrouvaient dans deux DR. Le nombre le plus élevé de DR touchées par une seule RP était égal à 11.

Tous les ménages répondants qui n'ont pas été assignés dans une RP recensée par interview ou une RP pour réserves sont demeurés dans les RP ordinaires. Dans la plupart des cas, ces ménages sont demeurés dans leur RP originale. Toutefois, en raison de l'extraction de ménages dans les RP recensées par interview et dans les RP pour réserves, le nombre de ménages dans certaines RP originales a été réduit. Si le nombre de ménages répondants dans la RP originale était réduit à un faible nombre (habituellement moins de 200), les ménages ont été réaffectés à la RP ordinaire la plus proche et la RP originale a été supprimée. Cependant, si le nombre de ménages dans la RP originale restait important (habituellement au moins 200), la RP n'a pas été modifiée, même si le nombre de ménages était inférieur à la limite de l'intervalle de 300 à 699 ménages souhaité. Le [tableau 6.2.2](#) montre la ventilation du nombre de RP selon le type de RP après l'achèvement de la deuxième étape.

Rapport technique sur l'échantillonnage et la pondération

Tableau 6.2.2 Fréquence du type de région de pondération

Type de RP	Région	Nombre de RP
Ordinaire	Terre-Neuve-et-Labrador	82
	Île-du-Prince-Édouard	24
	Nouvelle-Écosse	158
	Nouveau-Brunswick	125
	Québec	1 501
	Ontario	2 063
	Manitoba	199
	Saskatchewan	162
	Alberta	578
	Colombie-Britannique	756
	Yukon	5
	Territoires du Nord-Ouest	10
	Nunavut	0
	Canada	5 663
Recensé par interview	Canada	39
Réserves	Canada	182
Total	Canada	5 884

Note : Les statistiques pour les RP recensées par interview et les RP pour réserves n'ont été fournies que pour le Canada, parce que ces RP pouvaient chevaucher les limites provinciales et territoriales.

Source : Statistique Canada, Enquête nationale auprès des ménages de 2011.

Le [tableau 6.2.3](#) montre le nombre de RP selon le type de RP et les intervalles de nombre de ménages répondants. Comme il est précisé à la [section 6.1](#), la plupart des RP doivent contenir de 300 à 699 ménages répondants. Le tableau montre que 5 736 (97,5 %) des 5 884 RP finales se trouvaient dans l'intervalle de 300 à 699 en 2011.

Tableau 6.2.3 Répartition des RP selon le nombre de ménages répondants et le type de RP

Ménages répondants	Enquête nationale auprès des ménages de 2011				
	RP ordinaires	RP recensées par interview	RP pour réserves	Total	Pourcentage (%)
0 à 99	2	0	6	8	0,1
100 à 199	5	0	8	13	0,2
200 à 299	84	1	5	90	1,5
300 à 399	1 981	16	84	2 081	35,4
400 à 499	1 797	10	38	1 845	31,4
500 à 599	1 212	8	13	1 233	21,0
600 à 699	564	3	10	577	9,8
700 à 799	11	1	8	20	0,3
800 à 899	1	0	5	6	0,1
900 à 999	1	0	4	5	0,1
1 000+	5	0	1	6	0,1
Total	5 663	39	182	5 884	100,0

Source : Statistique Canada, Enquête nationale auprès des ménages de 2011.

Rapport technique sur l'échantillonnage et la pondération

L'ensemble de logements privés occupés et les ménages correspondants du Recensement de 2011 ont été considérés comme faisant partie du champ de l'ENM de 2011. Cependant, étant donné les taux de réponse variables à l'ENM, la taille de chaque RP dépendait du nombre de ménages répondants à l'ENM et non du nombre de ménages dans le champ de l'enquête. Après avoir déterminé les RP dans l'ENM, le nombre de ménages du recensement dans le champ de l'enquête a également été déterminé. Le [tableau 6.2.4](#) montre le nombre de RP selon le type de RP, ainsi que les intervalles de nombre de ménages du recensement dans le champ de l'enquête.

Tableau 6.2.4 Répartition des RP selon le nombre de ménages du recensement dans le champ de l'enquête et le type de RP

Ménages dans le champ de l'enquête	Enquête nationale auprès des ménages de 2011				
	RP ordinaires	RP recensées par interview	RP pour réserves	Total	Pourcentage (%)
0 à 499	1	12	102	115	2,0
500 à 999	5	27	72	104	1,8
1 000 à 1 499	217	0	5	222	3,8
1 500 à 1 999	1 595	0	2	1 597	27,1
2 000 à 2 499	1 790	0	0	1 790	30,4
2 500 à 2 999	1 285	0	1	1 286	21,9
3 000 à 3 499	615	0	0	615	10,5
3 500 à 3 999	143	0	0	143	2,4
4 000 à 4 499	10	0	0	10	0,2
4 500 à 4 999	0	0	0	0	0,0
5 000+	2	0	0	2	0,0
Total	5 663	39	182	5 884	100,0

Sources : Statistique Canada, Recensement de 2011 et Enquête nationale auprès des ménages de 2011.

La majorité des RP contenaient de 1 500 à 2 999 ménages de recensement dans le champ de l'enquête, ce qui est comparable avec le Recensement de 2006 dans lequel les RP étaient conçues de manière à contenir de 1 000 à 3 000 ménages.

7. Évaluation des procédures de pondération

Comme il est décrit au [chapitre 4](#), la méthode d'estimation appliquée pour l'Enquête nationale auprès des ménages (ENM) comprenait l'attribution de poids. Pour commencer, à chaque ménage, on a appliqué un poids d'échantillonnage qui était déterminé par le plan de sondage de l'ENM. Certains ajustements de ces poids ont ensuite été nécessaires pour tenir compte de la non-réponse totale. Ces poids ajustés, appelés poids initiaux ou poids avant calage, ont ensuite été ajustés durant le processus de calage afin de produire les poids finaux des ménages. Ces poids finaux ont permis d'obtenir une concordance généralement meilleure entre les chiffres du recensement et les estimations de l'ENM pour les variables communes aux deux enquêtes. Durant le calage, les caractéristiques énoncées à [annexe C](#) ont été utilisées comme contraintes. La façon dont certaines contraintes ont été supprimées afin d'obtenir une meilleure concordance entre les chiffres du recensement et les estimations de l'ENM pour d'autres caractéristiques est discutée au [chapitre 4](#).

Le présent chapitre présente et évalue certains aspects des procédures de pondération de l'ENM. On y examine la fréquence à laquelle certaines contraintes ont été supprimées durant le calage, ainsi que leur effet sur les estimations de l'ENM. Le chapitre examine aussi la distribution des poids et, pour diverses caractéristiques, les écarts entre les chiffres du recensement et les estimations d'échantillon à divers niveaux d'agrégation géographique.

7.1 Suppression de contraintes

L'objectif du calage est d'ajuster les poids des ménages afin que les estimations de l'ENM obtenues soient aussi proches que possible des chiffres du recensement pour de nombreuses caractéristiques communes. Le calage doit aussi produire une concordance exacte entre les chiffres du recensement et les estimations de l'ENM pour toute contrainte obligatoire. Voir l'[annexe C](#) pour la liste des contraintes et la [section 4.6](#) pour d'autres renseignements sur le calage. L'ajout des variables de langue pour le Recensement de 2011 signifie que le nombre de contraintes était beaucoup plus grand que pour le Recensement de 2006.

Le calage a été effectué pour 5 884 régions de pondération (RP) qui ont été traitées indépendamment les unes des autres. Dans chaque RP, les 60 contraintes ont toutes été entrées dans le calage et n'ont été supprimées qu'au besoin. Le nombre total de personnes et le nombre total de ménages étaient les variables constituant l'ensemble minimum de contraintes, ce qui signifie qu'elles ne pouvaient être supprimées pour aucune des RP. Les autres contraintes ont été abandonnées ou supprimées au besoin pour les raisons qui suivent :

1. Population nulle – Si la population pour la variable choisie comme contrainte est nulle dans la RP, l'estimation pour cette contrainte doit également être nulle. Cette contrainte ne contribue donc pas au processus de calage. Nous ne classons pas ces contraintes comme étant supprimées, mais plutôt comme n'étant pas admissibles pour le calage.
2. Petit échantillon – Si le nombre de répondants à l'ENM pour une contrainte dans une RP en particulier est supérieur à 0, mais inférieur à 30, l'utilisation de la contrainte en question réduira l'exactitude des estimations agrégées. Les contraintes de ce type sont par conséquent supprimées.
3. Linéairement dépendant et quasi linéairement dépendant – S'il est possible de calculer la valeur d'une contrainte en combinant les valeurs d'autres contraintes, l'une de ces contraintes doit être supprimée, à cause de la dépendance linéaire. Par exemple, la somme des valeurs pour les contraintes d'état matrimonial Marié, Célibataire, Séparé, Divorcé et Veuf est égale à la valeur de la contrainte TOTPERS. Donc, au moins une de ces contraintes n'est pas nécessaire et peut être supprimée. Les variables d'âge et de langue contribuent à d'autres groupes de contraintes qui donnent lieu à une dépendance linéaire parce que leur somme est également égale à la valeur de la contrainte TOTPERS. La somme des valeurs des contraintes de taille de ménage est aussi égale à la valeur de la contrainte Nombre de ménages, si bien qu'au moins une de ces contraintes doit être supprimée.

Si une contrainte est supprimée parce que la taille de l'échantillon est petite (raison 2), sa valeur pourrait être déterminée en soustrayant la combinaison d'autres valeurs de contraintes qui n'ont pas encore été

Rapport technique sur l'échantillonnage et la pondération

supprimées. Dans ce cas, une des contraintes restantes doit aussi être supprimée, à cause de la dépendance linéaire. Si une autre contrainte n'est pas supprimée, cela revient à garder la contrainte supprimée à cause d'un petit échantillon, soit raison 2). Par exemple, supposons que la contrainte d'état matrimonial Veuf soit supprimée dans une RP parce que le nombre de répondants est inférieur à 30. Toutefois, cette valeur de la contrainte Veuf peut être obtenue en soustrayant les contraintes restantes Marié, Célibataire, Séparé et Divorcé de la contrainte TOTPERS. Par conséquent, une contrainte supplémentaire doit être supprimée. La sommation des contraintes restantes, Marié, Célibataire, Séparé et Divorcé, aurait été approximativement égale à la valeur de la contrainte TOTPERS.

La dépendance linéaire est équivalente à l'existence d'une redondance avec les contraintes choisies, ou avec les contraintes choisies et les contraintes à faible population. La dépendance quasi linéaire est équivalente à l'existence de contraintes presque redondantes.

4. Valeur aberrante – Si l'application d'une contrainte fait que le poids d'un ménage se situe en dehors des limites acceptables de l'intervalle de 1 à 100, la contrainte en question est supprimée afin qu'elle ne donne pas lieu à des poids aberrants.

Sauf dans le premier cas où la population est nulle, chaque fois qu'une contrainte est supprimée, le processus de calage n'essaie plus de faire concorder les chiffres du recensement et les estimations de l'ENM pour cette contrainte dans la RP en question. L'écart entre le chiffre du recensement et l'estimation de l'ENM sera habituellement plus grand pour une contrainte supprimée fréquemment que pour une contrainte supprimée nettement moins souvent. Ce fait est évident si l'on compare le [tableau 7.1.1](#) et le [tableau 7.3.1](#).

Le [tableau 7.1.1](#) présente la liste de toutes les contraintes au niveau de la RP, ainsi que le nombre de fois que les contraintes ont été supprimées pour chacune des raisons. Les situations décrites ci-dessus sont énumérées respectivement dans les colonnes Population nulle, Petit échantillon, Linéairement dépendant, Quasi linéairement dépendant et Valeur aberrante.

Tableau 7.1.1 Fréquence de suppression des contraintes au niveau de la RP en 2011 dans l'ajustement final des poids

Variables/ contraintes	Population nulle	Petit échantillon	Linéairement dépendant	Quasi linéairement dépendant	Valeur aberrante
TOTPERS	0	0	0	0	0
TPERGE15	0	3	3 878	1 970	0
MALE	0	4	1	3 114	0
MALEGE15	0	5	0	373	0
AGE4	1	592	390	427	0
AGE9	1	804	301	325	0
AGE14	1	707	356	299	0
AGE19	3	459	278	179	0
AGE24	2	299	88	168	0
AGE29	2	387	88	246	0
AGE34	2	279	196	322	0
AGE39	2	210	326	341	0
AGE44	1	145	358	290	0
AGE49	1	70	400	237	0
AGE54	0	66	119	62	0
AGE59	1	97	10	13	0
AGE64	3	253	5	17	0
AGE74	0	186	187	212	0
AGE75PL	1	1 070	209	190	0
SINGLE	0	4	38	5 804	0
MARRIED	0	5	1	5 611	0

Rapport technique sur l'échantillonnage et la pondération

Tableau 7.1.1 Fréquence de suppression des contraintes au niveau de la RP en 2011 dans l'ajustement final des poids (suite)

Variables/ contraintes	Population nulle	Petit échantillon	Linéairement dépendant	Quasi linéairement dépendant	Valeur aberrante
SEPARATED	1	4 192	0	0	0
DIVORCED	1	213	0	1	0
WIDOWED	2	1 392	0	11	0
COMLAW	0	341	0	164	0
COUPLE	0	5	2	5 844	0
CHILD	0	14	2	5 835	0
LONEPAR	1	1 242	0	25	0
INEFAM	0	4	3 030	2 817	0
NOINEFAM	1	44	0	2 778	1
NOTINFAM	0	9	0	306	2
EN_EN_BI	32	2 044	0	12	0
EN_EN_EN	117	824	0	2 425	0
FR_EN_BI	119	4 952	0	0	0
FR_FR_BI	195	3 082	0	148	0
FR_FR_FR	1 722	1 091	0	847	0
LAN_ABOR	2 571	1 317	0	60	0
LAN_AFR	1 562	2 532	0	0	0
LAN_ARAB	808	3 255	0	2	0
LAN_CHIN	610	2 994	0	12	0
LAN_GER	217	4 248	0	29	0
LAN_INDO	1 230	2 653	0	18	0
LAN_IRAN	1 830	2 533	0	0	0
LAN_KOR	1 843	2 312	0	1	0
LAN_PILI	1 769	2 696	0	0	0
LAN_ROM	161	3 298	0	3	0
LAN_SLAV	347	3 793	0	0	0
CHILDFAM	0	30	2	5 813	0
NOCLDFAM	0	15	0	616	0
TOTHLD	0	0	0	0	0
HHSIZE1	1	214	2	5 202	0
HHSIZE2	0	14	0	1 218	0
HHSIZE3	0	223	0	531	0
HHSIZE4	3	714	0	4 060	0
HHSIZE5	2	4 007	0	569	0
HHSIZEG6	9	5 219	14	467	0
SNGLDET	81	404	2	55	0
APTLT5	274	2 111	0	19	0

Source : Statistique Canada, Enquête nationale auprès des ménages de 2011.

Comme certaines langues non officielles ont tendance à être observées dans certaines parties du pays et non dans d'autres, de nombreuses RP auront un chiffre de population faible ou nul pour une langue particulière. Par conséquent, pour les contraintes correspondant aux langues non officielles, les chiffres du recensement et les estimations sont souvent nuls, ou bien les contraintes sont supprimées parce que le nombre de répondants ayant la caractéristique en question est faible. De même, pour certaines contraintes dans lesquelles interviennent le français ou l'anglais, la population est faible ou nulle dans certaines régions du pays. Par conséquent, des contraintes telles que FR_EN_BI, FR_FR_BI et EN_EN_EN sont souvent supprimées.

Rapport technique sur l'échantillonnage et la pondération

En général, il n'existait pas de déséquilibre important en ce qui concerne les contraintes d'âge supprimées, à l'exception peut-être de la contrainte correspondant à 75 ans et plus. Le contrôle (calage) a été effectué régulièrement sur tous les groupes d'âge, ce qui signifie que les écarts entre le chiffre du recensement et l'estimation de l'ENM n'étaient pas trop extrêmes.

Quelques autres contraintes ont été supprimées fréquemment. Le nombre de personnes de 15 ans et plus a été presque systématiquement supprimé parce qu'il était linéairement dépendant ou quasi linéairement dépendant. Cela tient au fait que le nombre de personnes âgées de 15 ans et plus peut être déterminé au moyen des contraintes d'âge, AGE19, AGE24, AGE29 jusqu'à AGE75PL. Enfin, les contraintes représentant des personnes dans des familles économiques ou ayant des enfants ont été supprimées fréquemment parce que celles représentant le nombre de personnes ne faisant pas partie d'une famille économique ou d'une famille avec enfants en faisaient des contraintes quasi linéairement dépendantes.

Les écarts actuels entre les chiffres du recensement et les estimations seront examinés à la [section 7.3](#).

Le [tableau 7.1.2](#) montre le nombre de fois que chaque raison de supprimer ou d'éliminer une contrainte au niveau de la RP est survenue. Le nombre total de contraintes supprimées est égal à la somme des valeurs des catégories Petit échantillon, Linéairement dépendant, Quasi linéairement dépendant et Valeur aberrante. Comme il est mentionné plus haut, la catégorie Population nulle n'est pas incluse dans le total, parce qu'elle ne représente pas réellement des contraintes supprimées. Le nombre moyen de contraintes supprimées par RP est simplement égal au total pour la catégorie divisé par 5 884, c'est-à-dire le nombre de RP.

Tableau 7.1.2 Statistiques sommaires concernant la suppression des contraintes au niveau de la RP en 2011 dans l'ajustement final des poids

Contraintes	Population nulle	Petit échantillon	Linéairement dépendant	Quasi linéairement dépendant	Valeur aberrante	Total
Nombre total de contraintes supprimées (niveau de la RP)	15 530	69 676	10 283	60 088	3	140 050
Nombre moyen de contraintes supprimées par RP	2,6	11,8	1,7	10,2	0,0	23,8

Source : Statistique Canada, Enquête nationale auprès des ménages de 2011.

7.2 Répartition des poids

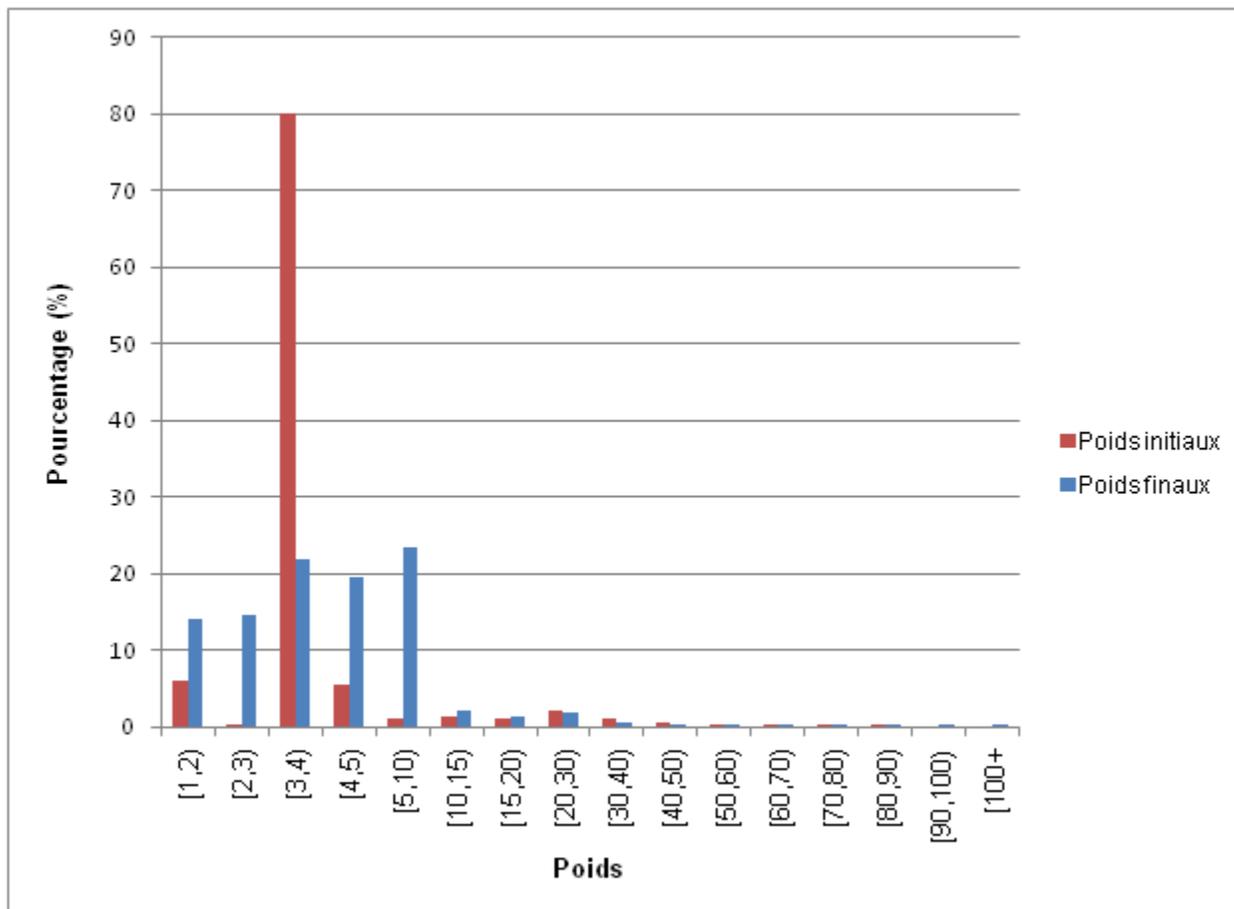
La [figure 7.2.1](#) montre la répartition des poids avant le calage (poids initiaux) et des poids après le calage (poids finaux). Les poids sont groupés par intervalles de taille 1 pour les poids faibles, par intervalles de taille 5 pour les poids compris entre 5 et 20, et par intervalles de taille 10 pour les poids supérieurs à 20 qui sont moins fréquents. La figure montre le pourcentage de fois que les poids se trouvent dans chaque intervalle. L'examen de la figure permet de constater que les poids initiaux, ou poids avant calage, étaient compris entre 3 et 4 pour environ 80 % des ménages. Cela est dû au plan de sondage de l'ENM, selon lequel environ 1 ménage sur 3 a reçu un questionnaire de l'ENM dans la plupart des régions. Des poids initiaux compris entre 1 et 2 sont observés environ 6 % du temps, de même que ceux compris dans l'intervalle de 4 à 5. Les 8 % restants de poids initiaux étaient répartis entre les catégories correspondant à l'intervalle de valeurs de 2 à 3 et l'intervalle de valeurs supérieures à 5. Le poids initial moyen était de 5,0 et le poids initial médian était de 3,6.

La [figure 7.2.1](#) permet également de constater l'effet qu'ont eu les ajustements par calage sur les poids. Une différence très évidente est que le pourcentage de ménages dont le poids initial était compris entre 3 et 4 a été réduit considérablement pour passer de 80 % à environ 22 %. En outre, les poids finaux étaient répartis plus uniformément entre les catégories comprises entre 1 et 10. Par contre, le pourcentage de ménages dont le poids initial était

Rapport technique sur l'échantillonnage et la pondération

supérieur à 10 n'a pas varié considérablement. La moyenne des poids finaux était de nouveau égale à 5,0, mais la médiane a augmenté pour atteindre 4,0.

Figure 7.2.1 Répartition des poids initiaux et des poids finaux



Source : Statistique Canada, Enquête nationale auprès des ménages de 2011.

Les changements entre les poids initiaux et les poids finaux peuvent être observés dans le [tableau 7.2.1](#). Ce tableau montre où les changements entre les catégories ont eu lieu. Par exemple, on peut voir que l'intervalle le plus stable était celui de 1 à 2, dans lequel 93,4 % des ménages dont le poids initial était compris entre 1 et 2 sont restés dans le même intervalle après le calage. La deuxième catégorie la plus stable était l'intervalle de 5 à 10 dans lequel sont restés 53,7 % des ménages dont les poids initiaux étaient compris entre 5 et 10. Cette stabilité est vraisemblablement attribuable au fait que cette catégorie est plus large que toutes celles qui lui sont inférieures.

Environ 55 % des ménages dont le poids initial était compris entre 2 et 3 ont vu diminuer leur poids pour être compris entre 1 et 2. Cependant, la situation était différente pour les ménages dont le poids était initialement compris entre 3 et 4; 26,6 % d'entre eux sont passés dans un groupe de poids plus faibles, 47,9 % d'entre eux sont passés dans un groupe de poids plus élevés, et 25,5 % seulement ont gardé un poids compris entre 3 et 4. Certains ménages dont le poids initial était très élevé ont vu diminuer leur poids. Dans les catégories de poids supérieurs à 20, le nombre de ménages a diminué après le calage.

Rapport technique sur l'échantillonnage et la pondération

Tableau 7.2.1 Répartition des poids initiaux et des poids finaux

Poids initiaux	Poids finaux									Total
	[1,2)	[2,3)	[3,4)	[4,5)	[5,10)	[10,15)	[15,20)	[20,30)	[30+)	
[1,2)	153 152	8 701	1 609	372	177	5	0	0	0	164 016
[2,3)	1 217	747	164	44	34	7	0	1	0	2 214
[3,4)	205 266	361 855	542 330	477 016	529 771	11 819	540	30	1	2 128 628
[4,5)	10 001	17 565	31 677	34 953	54 061	1 848	77	9	0	150 191
[5,10)	1 460	1 486	2 647	3 929	16 777	4 113	681	142	6	31 241
[10,15)	954	474	692	979	10 416	13 052	6 028	1 951	121	34 667
[15,20)	606	254	354	520	5 168	9 960	8 885	5 842	583	32 171
[20,30)	968	314	425	492	4 975	10 598	15 317	22 083	5 819	60 991
[30+)	762	161	203	280	2 188	4 346	7 110	18 454	19 838	53 342
Total	374 386	391 557	580 101	518 585	623 567	55 748	38 638	48 512	26 367	2 657 461

Note :

Le symbole « [» signifie que le chiffre est inclus dans l'intervalle, et le symbole «) » signifie que le chiffre n'est pas inclus dans l'intervalle.

Source : Statistique Canada, Enquête nationale auprès des ménages de 2011.

7.3 Écarts entre les chiffres du recensement et les estimations de l'ENM – Canada

Le [chapitre 4](#) décrit les méthodes utilisées pour calculer les poids finaux des ménages et la [section 7.2](#) montre certaines relations entre les poids initiaux et finaux. Le calage a réduit ou éliminé les écarts entre les chiffres du recensement et les estimations correspondantes de l'ENM pour les 58 contraintes au niveau du RP (voir l'[annexe C](#)). Cependant, certains écarts persistent, puisque des contraintes sont parfois supprimées (voir la [section 7.1](#)). L'écart entre le chiffre du recensement et l'estimation de l'ENM est défini comme suit :

$$\text{écart chiffre du recensement/estimation de l'ENM} = \frac{\text{estimation de l'ENM} - \text{chiffre du recensement}}{\text{chiffre du recensement}}$$

Le numérateur de l'expression (estimation de l'ENM – chiffre du recensement) est appelé « différence entre le chiffre du recensement et l'estimation de l'ENM ». En divisant cette valeur par le chiffre du recensement, on obtient la différence entre le chiffre du recensement et l'estimation de l'ENM par rapport à la grandeur du chiffre du recensement. Autrement dit, le ratio indique de quel pourcentage la caractéristique a été surestimée (valeur positive) ou sous-estimée (valeur négative).

Le [tableau 7.3.1](#) montre les différences entre le chiffre du recensement et l'estimation de l'ENM au niveau du Canada pour les 58 contraintes au niveau de la RP pour les poids initiaux ainsi que les poids finaux. La caractéristique « FEMALE » n'est pas l'une des 58 contraintes au niveau de la RP énumérées à l'[annexe C](#) parce qu'elle peut être calculée par différence entre le nombre total de personnes et le nombre d'hommes. Cependant, elle a été incluse dans ce tableau en raison de l'intérêt qu'elle présente. On peut voir que la somme de la différence entre les poids finaux pour les hommes et les femmes est égale à 0. Il convient aussi de mentionner que le chiffre du recensement pour la caractéristique TOTPERS est inférieur au chiffre du Recensement de 2011 publié (33 476 688). Deux raisons expliquent cette différence. Premièrement, le processus de pondération ne s'appliquait qu'aux ménages privés. Les logements collectifs ne faisaient pas partie du champ de l'ENM et ont été exclus de l'enquête. Cependant, ils ont été utilisés pour produire les chiffres du recensement. Deuxièmement, les cas mentionnés à la [section 4.2](#) (les cinq subdivisions de recensement (SDR) correspondant à cinq réserves indiennes qui ont été exclues de l'univers

Rapport technique sur l'échantillonnage et la pondération

parce que le taux de réponse à l'ENM y était très faible) n'étaient pas inclus dans le processus de pondération, mais faisaient partie du calcul du chiffre publié.

Il ne suffit pas d'observer simplement la différence entre l'estimation de l'ENM et le chiffre du recensement. Il est préférable de prendre en considération la différence par rapport à la valeur du chiffre du recensement. Par conséquent, le [tableau 7.3.1](#) montre les écarts fondés sur les poids initiaux et les poids finaux. Dans le [tableau 7.3.1](#), l'écart est compris entre -1,00 % et 1,00 % dans la plupart des cas.

La différence entre le chiffre du recensement et l'estimation de l'ENM avait tendance à être plus grande pour les poids initiaux que pour les poids finaux. Cela montre l'importance du processus de calage. Comme il est mentionné à la [section 7.1](#), une différence entre le chiffre du recensement et l'estimation de l'ENM pourrait avoir lieu dans une RP pour une caractéristique si la contrainte est supprimée durant le calage. Autrement dit, le processus n'effectuait un contrôle (calage) sur aucune contrainte supprimée pour une RP donnée. Si la contrainte est supprimée dans de nombreuses RP, les différences pourraient s'annuler partiellement, ou elles pourraient s'additionner pour créer une différence plus importante au niveau du Canada. Le nombre total de personnes (TOTPERS) et le nombre total de ménages (TOTHLDS) étaient les deux seules contraintes obligatoires pour lesquelles la concordance entre le chiffre du recensement et l'estimation de l'ENM devait être garantie pour toutes les RP, de sorte que la différence ainsi que l'écart pour les poids finaux étaient nuls. Toutefois, toutes les autres contraintes ont dû être supprimées pour certaines RP.

Les contraintes qui ont été supprimées fréquemment et les cas pouvant présenter des différences ou des écarts importants ont été mentionnés à la [section 7.1](#). L'effet de la suppression de ces contraintes peut être constaté dans le [tableau 7.3.1](#). Les contraintes au niveau de la personne Nombre total de personnes de 15 ans et plus, Célibataire, Marié, Nombre d'enfants, Couple, Appartenance à une famille économique et Appartenance à une famille avec un enfant ont presque toujours été supprimées. Parmi ces contraintes, les différences observées pour Marié et Couple étaient assez importantes (55 107 et 54 021, respectivement). Cependant, comme les chiffres du recensement étaient très élevés, les écarts pour ces contraintes n'étaient pas trop élevés (0,43 % et 0,34 %, respectivement). Par ailleurs, l'écart pour le nombre de personnes séparées était de -3,35 %. Cette contrainte relativement petite a été supprimée 4 192 fois et a créé une sous-estimation de -27 154.

Les plus grandes différences et les plus grands écarts ont généralement été constatés pour certaines variables de langue. Comme, pour de nombreuses langues, le nombre de personnes n'était pas élevé dans plusieurs parties du pays, ces variables ont souvent été supprimées durant le calage, ce qui a donné lieu à certaines différences entre les chiffres du recensement et les estimations de l'ENM. En outre, comme le chiffre du recensement est relativement faible pour bon nombre de ces variables, les différences sont amplifiées, ce qui donne des écarts plus importants. La différence la plus importante a été observée pour la contrainte EN_EN_EN, c'est-à-dire les personnes qui ont déclaré que l'anglais était leur langue maternelle, la langue parlée à la maison et la langue officielle. La surestimation de 166 801 était, de loin, la différence la plus grande observée pour une variable. Cependant, comme il s'agissait du scénario de langue le plus fréquent, l'écart était de 0,99 %, ce qui signifie que la caractéristique a été surestimée d'environ 1 %. L'écart le plus grand a été observé pour la caractéristique FR_EN_BI, c'est-à-dire les personnes qui ont déclaré que le français était leur langue maternelle, que l'anglais était la langue parlée à la maison et qu'elles étaient capables de parler l'anglais ainsi que le français. Cette contrainte a été supprimée fréquemment parce qu'il existait au moins un répondant, mais moins de 30 répondants dans la RP. Par conséquent, cette caractéristique, pour laquelle le chiffre du recensement était de 351 000 environ, a fait l'objet d'une surestimation de près de 41 000 et d'un écart représentant 11,68 %. Les autres contraintes de langue pour lesquelles les écarts étaient importants étaient les langues africaines (8,49 %) et les langues des Philippines (5,53 %).

Pour la plupart des autres caractéristiques au niveau de la personne, les écarts étaient plus faibles, particulièrement pour l'âge et le sexe, pour lesquels l'écart entre l'estimation et le chiffre du recensement ne dépassait pas 0,2 %. Seul le groupe des 75 ans et plus pour lequel l'écart est de 0,28 % dépasse 0,2 %. Il s'agissait aussi du groupe d'âge pour lequel la différence entre le chiffre du recensement et l'estimation de l'ENM était la plus grande (-5 365).

Pour les variables au niveau du ménage, les résultats étaient partagés. Le nombre de ménages de 1 personne, 2 personnes et 6 personnes ou plus étaient tous sous-estimés. C'est pour les ménages comptant au moins 6 personnes que la sous-estimation était la plus importante (-14 493), ce qui a donné lieu à un grand écart, soit -

Rapport technique sur l'échantillonnage et la pondération

3,69 %. Par contre, la différence la plus importante entre le chiffre du recensement et l'estimation a été observée pour les ménages de 4 personnes (24 124), ce qui a donné un écart de 1,27 %.

Les estimations les plus proches pour une contrainte non obligatoire étaient celles calculées pour le groupe d'âge de 5 à 9 ans et pour le groupe de langues coréennes pour lesquels les différences étaient de 69 et de 72, respectivement. Cependant, pour le petit groupe de langues coréennes, l'écart était de 0,05 %. Les écarts les plus faibles ont été obtenus pour les logements individuels non attenants (0,00 %), le groupe d'âge de 5 à 9 ans (0,00 %), le groupe d'âge de 50 à 54 ans (0,01 %) et la famille économique (-0,02 %). Il convient de noter que ces écarts ont été arrondis au centième (0,01) près.

Tableau 7.3.1 Différences et écarts (%) entre les chiffres du recensement et les estimations de l'ENM de 2011 pour le Canada

Caractéristique	Recensement	Poids initiaux, ENM		Poids finaux, ENM		
	chiffre	estimation	différence	estimation	différence	Écart (%)
TOTPERS	32 852 323	32 570 247	-282 076	32 852 323	0	0,00
TPERGE15	27 266 716	26 792 223	-474 493	27 259 524	-7 192	-0,03
MALE	16 151 561	15 943 350	-208 211	16 163 113	11 552	0,07
FEMALE*	16 700 762	16 626 898	-73 864	16 689 210	-11 552	-0,07
MALEGE15	13 291 793	12 983 638	-308 155	13 295 352	3 559	0,03
AGE4	1 869 871	1 938 216	68 345	1 873 691	3 820	0,20
AGE9	1 803 353	1 875 881	72 528	1 803 422	69	0,00
AGE14	1 912 383	1 963 928	51 545	1 915 687	3 304	0,17
AGE19	2 161 432	2 109 902	-51 530	2 160 275	-1 157	-0,05
AGE24	2 167 042	2 017 460	-149 582	2 163 793	-3 249	-0,15
AGE29	2 150 627	2 054 640	-95 987	2 149 302	-1 325	-0,06
AGE34	2 145 318	2 115 595	-29 723	2 144 647	-671	-0,03
AGE39	2 156 499	2 149 240	-7 259	2 155 348	-1 151	-0,05
AGE44	2 305 179	2 272 305	-32 874	2 306 085	906	0,04
AGE49	2 651 229	2 585 140	-66 089	2 655 581	4 352	0,16
AGE54	2 633 806	2 567 868	-65 938	2 633 979	173	0,01
AGE59	2 316 094	2 267 609	-48 485	2 314 650	-1 444	-0,06
AGE64	2 027 821	2 024 784	-3 037	2 024 329	-3 492	-0,17
AGE74	2 619 303	2 671 238	51 935	2 624 535	5 232	0,20
AGE75PL	1 932 366	1 956 441	24 075	1 927 001	-5 365	-0,28
SINGLE	15 460 197	15 203 908	-256 289	15 449 727	-10 470	-0,07
MARRIED	12 823 743	12 994 368	170 625	12 878 850	55 107	0,43
SEPARATED	810 778	761 604	-49 174	783 624	-27 154	-3,35
DIVORCED	2 323 491	2 193 851	-129 640	2 319 013	-4 478	-0,19
WIDOWED	1 434 114	1 416 516	-17 598	1 421 110	-13 004	-0,91
COMLAW	3 135 282	3 002 520	-132 762	3 126 473	-8 809	-0,28
COUPLE	15 722 178	15 786 489	64 311	15 776 199	54 021	0,34
CHILD	9 969 150	9 972 358	3 208	9 950 695	-18 455	-0,19
LONEPAR	1 527 427	1 487 512	-39 915	1 499 376	-28 051	-1,84
INEFAM	27 954 234	27 893 392	-60 842	27 948 792	-5 442	-0,02
NOINEFAM	4 898 089	4 676 855	-221 234	4 903 531	5 442	0,11
NOTINFAM	5 633 568	5 323 889	-309 679	5 626 052	-7 516	-0,13
EN_EN_BI	1 613 378	1 584 730	-28 648	1 610 801	-2 577	-0,16
EN_EN_EN	16 806 378	16 990 679	184 301	16 973 179	166 801	0,99
FR_EN_BI	351 034	388 803	37 769	392 021	40 987	11,68
FR_FR_BI	2 646 831	2 731 130	84 299	2 666 821	19 990	0,76

Rapport technique sur l'échantillonnage et la pondération

Tableau 7.3.1 Différences et écarts (%) entre les chiffres du recensement et les estimations de l'ENM de 2011 pour le Canada (suite)

Caractéristique	Recensement	Poids initiaux, ENM		Poids finaux, ENM		Écart (%)
	chiffre	estimation	différence	estimation	différence	
FR_FR_FR	3 773 941	3 731 106	-42 835	3 806 924	32 983	0,87
LAN_ABOR	196 928	186 974	-9 954	191 850	-5 078	-2,58
LAN_AFR	76 951	77 959	1 008	83 485	6 534	8,49
LAN_ARAB	327 059	325 422	-1 637	323 124	-3 935	-1,20
LAN_CHIN	1 067 628	1 048 989	-18 639	1 066 952	-676	-0,06
LAN_GER	542 498	511 944	-30 554	536 789	-5 709	-1,05
LAN_INDO	877 645	896 995	19 350	875 699	-1 946	-0,22
LAN_IRAN	191 909	189 444	-2 465	192 127	218	0,11
LAN_KOR	137 141	131 245	-5 896	137 213	72	0,05
LAN_PILI	111 744	106 386	-5 358	117 923	6 179	5,53
LAN_ROM	1 113 783	1 112 691	-1 092	1 096 133	-17 650	-1,58
LAN_SLAV	683 708	665 284	-18 424	668 188	-15 520	-2,27
CHILDFAM	5 636 164	5 612 297	-23 867	5 633 721	-2 443	-0,04
NOCLDFAM	3 752 352	3 768 460	16 108	3 753 755	1 403	0,04
TOTHLD	13 319 251	13 283 671	-35 580	13 319 251	0	0,00
HHSIZE1	3 673 018	3 696 360	23 342	3 662 801	-10 217	-0,28
HHSIZE2	4 544 542	4 557 124	12 582	4 536 175	-8 367	-0,18
HHSIZE3	2 081 661	2 057 036	-24 625	2 086 205	4 544	0,22
HHSIZE4	1 903 101	1 885 895	-17 206	1 927 225	24 124	1,27
HHSIZE5	724 278	713 786	-10 492	728 688	4 410	0,61
HHSIZEG6	392 651	373 470	-19 181	378 158	-14 493	-3,69
SNGLDET	7 327 904	7 341 850	13 946	7 327 785	-119	0,00
APTLT5	2 431 005	2 392 925	-38 080	2 423 005	-8 000	-0,33

* La caractéristique FEMALE n'est pas l'une des 58 contraintes au niveau de la RP énumérées à l'[annexe C](#).

Sources : Statistique Canada, Recensement de 2011 et Enquête nationale auprès des ménages de 2011.

8. Variance

L'erreur d'une estimation est la différence entre l'estimation et la valeur réelle de ce que l'on tente d'estimer. L'erreur d'échantillonnage et l'erreur de non-réponse des ménages sont deux sources d'erreur dans l'ENM. L'erreur d'échantillonnage provient du fait que les estimations sont produites à partir d'observations provenant d'un échantillon et non de l'ensemble de la population. L'erreur de non-réponse des ménages survient quand des ménages sélectionnés dans l'échantillon ne répondent pas à l'enquête. Les méthodes d'estimation ont également une incidence sur l'erreur. En effet, certaines sont plus précises que d'autres pour estimer une caractéristique donnée de la population.

L'erreur a deux composantes : une aléatoire, la variance; et une systématique, le biais. La variance mesure la variabilité de l'estimation par rapport à l'estimation moyenne qui résulterait de répétitions hypothétiques du processus d'enquête. Cette variance peut être estimée à partir des données des répondants de l'échantillon. Le biais est la différence entre la valeur moyenne d'une estimation qui résulterait de répétitions hypothétiques du processus d'enquête, et la valeur réelle de la caractéristique estimée. Le [chapitre 5](#) traite du biais de non-réponse. Les méthodes d'échantillonnage et d'estimation utilisées dans l'ENM génèrent pour leur part un biais négligeable. Le présent chapitre traite de la variance due à l'échantillonnage et à la non-réponse des ménages.

Pour estimer la variance, il est d'abord nécessaire de dériver des estimateurs selon des formules mathématiques. Ces formules doivent tenir compte des différentes étapes d'échantillonnage et de pondération présentées dans les chapitres précédents. Ces formules dictent la façon d'utiliser les données des répondants pour produire les estimations de variance. Pour faire les dérivations, certaines hypothèses et simplifications ont dû être faites. Tout d'abord, il a été supposé que l'échantillon et le sous-échantillon tirés tous deux selon un plan d'échantillonnage stratifié systématique avaient été tirés selon un plan d'échantillonnage aléatoire simple sans remise stratifié. Il est nécessaire de faire une hypothèse de la sorte lorsqu'un échantillonnage systématique est effectué. L'hypothèse est raisonnable, car les strates sont petites et leur population peut alors être considérée comme étant homogène. La seconde hypothèse concerne le mécanisme générant la non-réponse des ménages. Il a été supposé qu'il était équivalent à un échantillonnage de type Poisson où la probabilité de réponse/sélection d'un ménage donné est égale au ratio du poids qui lui est attribué avant l'ajustement de non-réponse par transfert de poids sur le poids attribué suite à ce transfert. Par exemple, un ménage répondant qui a vu son poids être doublé suite au transfert aurait une probabilité de réponse estimée de 50 %. Il a aussi été supposé que cette probabilité était connue et non estimée. Faire ces hypothèses sur le mécanisme de non-réponse signifie que le biais de non-réponse dans l'estimation de la variance est ignoré. La dernière simplification est que les répondants-surprises ont été exclus du processus d'estimation de la variance. Ceci facilite grandement le processus et a peu d'incidence sur l'estimation étant donné que les répondants-surprises représentent une petite fraction de l'ensemble des répondants. Les détails techniques de l'estimation de la variance sont donnés dans Verret (2013).

Comme les estimateurs de variance selon les formules discutées précédemment ont tendance à sous-estimer la variance réelle, la dernière étape consistait à les corriger à l'aide de simulations. Pour ce faire, une population artificielle à partir de données-échantillon a d'abord été générée. De multiples échantillons ont ensuite été sélectionnés de façon indépendante (c'est-à-dire des répliques) et les méthodes de pondération ont été appliquées à chacun. Les estimateurs de variance calculés ont été comparés à la variabilité réelle observée d'une réplique à l'autre. Ceci fournit un ajustement à la hausse qui est appliqué à toutes les estimations de variance calculées selon les formules.

La variance estimée peut être utilisée pour créer plusieurs types de mesures de la variabilité d'une estimation. Par exemple, elle peut être utilisée pour calculer des erreurs types ou des coefficients de variation (CV). L'erreur type est égale à la racine carrée de la variance. Le CV est le ratio de l'erreur type de l'estimation et de l'estimation elle-même. Plus ces mesures de variabilité sont faibles, plus l'estimation est précise. Le CV est une mesure de variabilité particulièrement intéressante, car elle ne dépend pas de l'unité de mesure de la caractéristique étudiée et peut être exprimée en pourcentage. Il est important de bien distinguer ces mesures de variabilité d'autres mesures de la qualité qui ne sont pas des mesures de variabilité proprement dites. Des exemples de telles mesures sont les indicateurs de biais, le taux de réponse et le taux global de non-réponse de l'ENM. En effet, le taux de réponse est un indicateur du risque associé à l'erreur de non-réponse des ménages. Le taux global de non-réponse quant à lui est un indicateur du risque d'erreur due à la non-réponse des ménages et à la non-réponse d'item aux questions individuelles. Le [tableau 8.1](#) présente les CV de diverses estimations aux niveaux du Canada, des provinces et des territoires. Le

Rapport technique sur l'échantillonnage et la pondération

contenu de ce tableau peut être consulté ou encore téléchargé à l'adresse suivante :

<http://www12.statcan.gc.ca/nhs-enm/2011/ref/cv/index.cfm?Lang=F>.

Les méthodes d'échantillonnage et d'estimation utilisées dans l'ENM 2011 sont différentes de celles qui ont été utilisées avec le questionnaire complet du Recensement de 2006. L'ampleur et l'effet de la non-réponse sont très différents en 2011. Tout cela a une incidence sur la variabilité des estimations. Une note méthodologique comparant les CV des estimations de 2011 à ceux des estimations de 2006 pour certaines variables est disponible à l'adresse <http://www12.statcan.gc.ca/nhs-enm/2011/ref/cv/cvnote.cfm?Lang=F>.

Rapport technique sur l'échantillonnage et la pondération

Tableau 8.1 Coefficients de variation (%) de l'Enquête nationale auprès des ménages de 2011 – Canada, provinces et territoires

Caractéristique	Canada	Terre-Neuve-et-Labrador	Île-du-Prince-Édouard	Nouvelle-Écosse	Nouveau-Brunswick	Québec	Ontario
Caractéristique de la population							
Immigration et citoyenneté							
Population totale dans les ménages privés selon la citoyenneté	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Citoyens canadiens	0,01	0,06	0,18	0,06	0,06	0,02	0,03
Ne sont pas des citoyens canadiens	0,23	6,06	5,87	2,40	3,01	0,50	0,35
Population totale dans les ménages privés selon le statut d'immigrant	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Non-immigrants	0,02	0,07	0,24	0,09	0,08	0,03	0,04
Immigrants	0,08	3,55	3,88	1,41	1,87	0,21	0,10
Résidents non permanents	0,61	9,99	16,29	4,97	5,98	1,28	1,02
Population immigrante totale dans les ménages privés selon le lieu de naissance	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Amériques	0,29	6,74	7,96	2,94	3,11	0,60	0,41
Europe	0,17	4,79	6,15	2,08	2,98	0,46	0,22
Afrique	0,50	16,08	41,14	7,51	9,04	0,72	0,83
Asie	0,13	7,95	6,87	3,07	4,43	0,50	0,17
Océanie et autres	1,32	28,55	55,64	12,39	15,27	6,86	2,51
Minorités visibles							
Population totale dans les ménages privés selon les minorités visibles	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Population des minorités visibles	0,10	5,27	6,21	1,76	2,90	0,30	0,14
Pas une minorité visible	0,02	0,07	0,20	0,10	0,07	0,04	0,05
Peuples autochtones							
Population totale dans les ménages privés selon l'identité autochtone	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Identité autochtone	0,26	1,61	8,37	1,71	2,10	0,72	0,60
Identité non autochtone	0,01	0,12	0,14	0,07	0,07	0,01	0,02
Mobilité							
Total – Mobilité 5 ans auparavant	0,00	0,03	0,04	0,02	0,02	0,01	0,00
Personnes n'ayant pas déménagé	0,05	0,35	0,72	0,27	0,29	0,09	0,08
Personnes ayant déménagé	0,08	0,86	1,52	0,54	0,60	0,16	0,14
Scolarité							
Population totale âgée de 15 ans et plus selon le plus haut certificat, diplôme ou grade	0,01	0,05	0,10	0,04	0,03	0,01	0,01
Aucun certificat, diplôme ou grade	0,10	0,72	1,61	0,60	0,61	0,19	0,17
Diplôme d'études secondaires ou l'équivalent	0,09	0,86	1,44	0,59	0,59	0,20	0,14
Certificat, diplôme ou grade postsecondaire	0,04	0,45	0,80	0,29	0,35	0,08	0,07
Certificat, diplôme ou grade universitaire au baccalauréat ou supérieur	0,09	1,11	1,79	0,62	0,77	0,19	0,14
Travail							
Population totale âgée de 15 ans et plus selon la langue le plus souvent utilisée au travail	0,03	0,29	0,41	0,20	0,20	0,06	0,05
Anglais, réponse unique	0,04	0,30	0,44	0,20	0,27	0,30	0,05
Français, réponse unique	0,08	8,67	7,41	2,63	0,66	0,08	0,75
Langues non officielles, réponse unique	0,51	6,10	15,65	7,38	7,02	1,20	0,83

Rapport technique sur l'échantillonnage et la pondération

Tableau 8.1 Coefficients de variation (%) de l'Enquête nationale auprès des ménages de 2011 – Canada, provinces et territoires (suite)

Caractéristique	Manitoba	Saskatchewan	Alberta	Colombie-Britannique	Yukon	Territoires du Nord-Ouest	Nunavut
Caractéristique de la population							
Immigration et citoyenneté							
Population totale dans les ménages privés selon la citoyenneté	0,00	0,00	0,00	0,00	0,00	0,02	0,00
Citoyens canadiens	0,08	0,07	0,05	0,05	0,34	0,16	0,02
Ne sont pas des citoyens canadiens	1,20	1,84	0,68	0,52	6,26	5,91	3,01
Population totale dans les ménages privés selon le statut d'immigrant	0,00	0,00	0,00	0,00	0,00	0,02	0,00
Non-immigrants	0,11	0,09	0,07	0,07	0,50	0,22	0,04
Immigrants	0,56	1,19	0,29	0,18	3,95	2,94	1,74
Résidents non permanents	4,10	4,76	1,58	1,37	21,79	12,52	6,22
Population immigrante totale dans les ménages privés selon le lieu de naissance	0,00	0,00	0,00	0,00	0,00	0,02	0,00
Amériques	1,81	3,54	1,08	0,88	9,58	8,62	3,71
Europe	1,20	2,19	0,64	0,44	5,92	5,72	2,77
Afrique	3,61	5,97	1,65	1,73	28,75	13,57	5,28
Asie	0,84	1,87	0,46	0,23	9,38	5,51	3,22
Océanie et autres	10,08	11,08	3,56	1,83	22,02	24,68	6,46
Minorités visibles							
Population totale dans les ménages privés selon les minorités visibles	0,00	0,00	0,00	0,00	0,00	0,02	0,00
Population des minorités visibles	0,69	1,39	0,33	0,19	6,62	3,54	2,19
Pas une minorité visible	0,10	0,09	0,07	0,07	0,43	0,25	0,04
Peuples autochtones							
Population totale dans les ménages privés selon l'identité autochtone	0,00	0,00	0,00	0,00	0,00	0,02	0,00
Identité autochtone	0,62	0,71	0,72	0,64	2,72	0,67	0,11
Identité non autochtone	0,12	0,13	0,05	0,04	0,82	0,72	0,67
Mobilité							
Total – Mobilité 5 ans auparavant	0,02	0,02	0,01	0,01	0,00	0,06	0,04
Personnes n'ayant pas déménagé	0,26	0,29	0,17	0,15	1,56	0,89	0,47
Personnes ayant déménagé	0,42	0,47	0,21	0,20	1,89	0,94	0,62
Scolarité							
Population totale âgée de 15 ans et plus selon le plus haut certificat, diplôme ou grade	0,03	0,04	0,02	0,02	0,00	0,09	0,06
Aucun certificat, diplôme ou grade	0,46	0,55	0,33	0,30	2,91	0,86	0,32
Diplôme d'études secondaires ou l'équivalent	0,44	0,52	0,27	0,23	2,71	1,49	0,93
Certificat, diplôme ou grade postsecondaire	0,27	0,32	0,14	0,12	1,23	0,70	0,50
Certificat, diplôme ou grade universitaire au baccalauréat ou supérieur	0,54	0,70	0,29	0,24	2,75	1,73	0,99
Travail							
Population totale âgée de 15 ans et plus selon la langue le plus souvent utilisée au travail	0,14	0,16	0,08	0,08	0,51	0,34	0,25
Anglais, réponse unique	0,15	0,17	0,08	0,09	0,56	0,36	0,35
Français, réponse unique	3,04	6,39	3,39	3,25	17,74	13,75	4,66
Langues non officielles, réponse unique	2,78	4,28	2,14	0,94	21,23	3,78	0,83

Rapport technique sur l'échantillonnage et la pondération

Tableau 8.1 Coefficients de variation (%) de l'Enquête nationale auprès des ménages de 2011 – Canada, provinces et territoires (suite)

Caractéristique	Canada	Terre-Neuve-et-Labrador	Île-du-Prince-Édouard	Nouvelle-Écosse	Nouveau-Brunswick	Québec	Ontario
Population totale âgée de 15 ans et plus selon la situation d'activité	0,01	0,05	0,10	0,04	0,03	0,01	0,01
Population active	0,03	0,33	0,47	0,22	0,23	0,06	0,05
Personnes occupées	0,04	0,41	0,62	0,26	0,27	0,07	0,06
Chômeurs	0,22	1,47	2,87	1,25	1,32	0,45	0,34
Inactifs	0,06	0,48	1,03	0,37	0,39	0,12	0,10
Logement							
Nombre total de logements privés occupés selon l'état du logement	0,00	0,00	0,01	0,00	0,00	0,00	0,00
Seulement entretien régulier ou réparations mineures requises	0,02	0,18	0,33	0,15	0,16	0,04	0,03
Réparations majeures requises	0,25	2,09	4,07	1,37	1,50	0,49	0,45
Nombre total de logements privés occupés selon le nombre de pièces	0,00	0,00	0,01	0,00	0,00	0,00	0,00
Nombre moyen de pièces par logement	0,03	0,24	0,45	0,16	0,18	0,04	0,04
Nombre total de ménages privés selon le mode d'occupation	0,00	0,00	0,01	0,00	0,00	0,00	0,00
Propriétaire	0,03	0,26	0,47	0,19	0,19	0,07	0,05
Locataire	0,07	0,90	1,31	0,47	0,60	0,11	0,13
Logement de bande	0,12	1,07	3,63	0,30	0,65	0,55	0,56
Coûts d'habitation							
Nombre total de ménages propriétaires et locataires dont le revenu total du ménage est supérieur à zéro, dans les logements privés non agricoles, hors réserve selon le rapport des frais de logement au revenu	0,01	0,02	0,12	0,04	0,04	0,01	0,01
Moins que 30 % du revenu total du ménage est consacré aux coûts d'habitation	0,04	0,29	0,61	0,23	0,23	0,07	0,07
30 % ou plus du revenu total du ménage est consacré aux coûts d'habitation	0,12	1,32	2,46	0,79	0,99	0,24	0,19
De 30 % à moins de 100 % du revenu total du ménage est consacré aux coûts d'habitation	0,14	1,50	2,78	0,93	1,15	0,28	0,22
Revenu des particuliers							
Revenu total en 2010 de la population âgée de 15 ans et plus	0,01	0,05	0,10	0,04	0,03	0,01	0,01
Revenu moyen (\$)	0,09	0,50	0,98	0,44	0,41	0,15	0,14
Revenu après impôt en 2010 de la population de 15 ans et plus	0,01	0,05	0,10	0,04	0,03	0,01	0,01
Revenu moyen après impôt (\$)	0,07	0,42	0,84	0,33	0,34	0,13	0,11
Revenu de la famille							
Revenu de la famille en 2010 des familles économiques	0,01	0,09	0,20	0,07	0,07	0,02	0,02
Revenu moyen de la famille (\$)	0,10	0,55	1,11	0,51	0,46	0,18	0,15
Revenu moyen de la famille après impôt (\$)	0,08	0,46	0,95	0,38	0,38	0,15	0,13
Nombre moyen de personnes par famille	0,01	0,09	0,20	0,07	0,07	0,02	0,02

Rapport technique sur l'échantillonnage et la pondération

Tableau 8.1 Coefficients de variation (%) de l'Enquête nationale auprès des ménages de 2011 – Canada, provinces et territoires (suite)

Caractéristique	Manitoba	Saskatchewan	Alberta	Colombie-Britannique	Yukon	Territoires du Nord-Ouest	Nunavut
Population totale âgée de 15 ans et plus selon la situation d'activité	0,03	0,04	0,02	0,02	0,00	0,09	0,06
Population active	0,16	0,19	0,09	0,09	0,74	0,37	0,28
Personnes occupées	0,18	0,21	0,10	0,10	0,91	0,46	0,33
Chômeurs	1,27	1,48	0,74	0,59	4,80	1,88	1,14
Inactifs	0,33	0,42	0,24	0,16	2,50	1,13	0,50
Logement							
Nombre total de logements privés occupés selon l'état du logement	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Seulement entretien régulier ou réparations mineures requises	0,13	0,15	0,06	0,05	0,84	0,35	0,34
Réparations majeures requises	1,05	1,24	0,83	0,70	4,34	1,68	0,90
Nombre total de logements privés occupés selon le nombre de pièces	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Nombre moyen de pièces par logement	0,13	0,17	0,08	0,07	0,91	0,39	0,19
Nombre total de ménages privés selon le mode d'occupation	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Propriétaire	0,15	0,19	0,11	0,10	1,05	0,66	0,92
Locataire	0,38	0,57	0,30	0,24	2,37	0,72	0,24
Logement de bande	0,16	0,21	0,16	0,65	2,49	3,78	...
Coûts d'habitation							
Nombre total de ménages propriétaires et locataires dont le revenu total du ménage est supérieur à zéro, dans les logements privés non agricoles, hors réserve selon le rapport des frais de logement au revenu	0,06	0,11	0,04	0,02	0,16	0,06	0,01
Moins que 30 % du revenu total du ménage est consacré aux coûts d'habitation	0,20	0,25	0,13	0,13	1,00	0,51	0,14
30 % ou plus du revenu total du ménage est consacré aux coûts d'habitation	0,79	0,89	0,40	0,29	4,02	3,05	2,07
De 30 % à moins de 100 % du revenu total du ménage est consacré aux coûts d'habitation	0,91	1,01	0,47	0,35	4,59	3,44	2,33
Revenu des particuliers							
Revenu total en 2010 de la population âgée de 15 ans et plus	0,03	0,04	0,02	0,02	0,00	0,09	0,06
Revenu moyen (\$)	0,33	0,46	0,33	0,23	1,48	0,65	0,38
Revenu après impôt en 2010 de la population de 15 ans et plus	0,03	0,04	0,02	0,02	0,00	0,09	0,06
Revenu moyen après impôt (\$)	0,27	0,39	0,28	0,19	1,30	0,58	0,35
Revenu de la famille							
Revenu de la famille en 2010 des familles économiques	0,06	0,08	0,04	0,04	0,30	0,21	0,10
Revenu moyen de la famille (\$)	0,38	0,52	0,35	0,27	1,80	0,70	0,43
Revenu moyen de la famille après impôt (\$)	0,31	0,45	0,30	0,22	1,58	0,62	0,40
Nombre moyen de personnes par famille	0,06	0,08	0,04	0,04	0,35	0,21	0,10

Rapport technique sur l'échantillonnage et la pondération

Tableau 8.1 Coefficients de variation (%) de l'Enquête nationale auprès des ménages de 2011 – Canada, provinces et territoires (suite)

Caractéristique	Canada	Terre-Neuve-et-Labrador	Île-du-Prince-Édouard	Nouvelle-Écosse	Nouveau-Brunswick	Québec	Ontario
Revenu du ménage							
Revenu du ménage en 2010 des ménages privés	0,00	0,00	0,01	0,00	0,00	0,00	0,00
Revenu total moyen des ménages (\$)	0,09	0,51	1,01	0,45	0,42	0,15	0,14
Revenu moyen des ménages après impôt (\$)	0,07	0,43	0,87	0,34	0,35	0,13	0,12

Caractéristique	Manitoba	Saskatchewan	Alberta	Colombie-Britannique	Yukon	Territoires du Nord-Ouest	Nunavut
Revenu du ménage							
Revenu du ménage en 2010 des ménages privés	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Revenu total moyen des ménages (\$)	0,34	0,48	0,33	0,24	1,57	0,66	0,39
Revenu moyen des ménages après impôt (\$)	0,28	0,40	0,29	0,20	1,38	0,59	0,36

... n'ayant pas lieu de figurer

Source : Statistique Canada, Enquête nationale auprès des ménages de 2011.

9. Conclusion

Le lancement de l'Enquête nationale auprès des ménages (ENM) de 2011 pour remplacer le questionnaire complet du recensement a posé de nombreux défis. La participation à l'ENM n'étant pas obligatoire, les taux de non-réponse étaient plus élevés que ceux observés antérieurement pour le questionnaire complet du recensement. Les types et les profils de non-réponse étaient également différents. Le plan d'échantillonnage, le suivi des cas de non-réponse ainsi que les procédures de pondération et d'estimation adoptés pour l'ENM ont donc nécessité une approche plus complexe.

Les changements comprenaient une augmentation de la taille de l'échantillon, qui est passée d'un ménage privé sur cinq à un ménage privé sur trois. Une stratégie améliorée de suivi des cas de non-réponse a été mise en œuvre afin de s'assurer de cibler certaines régions et populations considérées comme posant un risque de produire des taux de non-réponse plus élevés qu'à l'ordinaire. Bien qu'ils présentent plusieurs limites, des indicateurs de biais dû à la non-réponse totale ont été produits pour quelques caractéristiques. Enfin, le plan d'échantillonnage complexe utilisé pour l'ENM a mené à l'élaboration d'une nouvelle méthodologie novatrice de pondération et d'estimation de la variance assez différente de celle appliquée pour les recensements antérieurs.

Annexe A Glossaire

Les définitions des termes, variables et concepts du recensement sont présentées ici telles qu'elles figurent dans le *Dictionnaire du Recensement de 2011* (n° 98-301-X2011001 au catalogue). L'utilisateur doit se reporter au *Dictionnaire du Recensement de 2011* pour obtenir les définitions complètes et des observations additionnelles se rapportant aux concepts, comme des renseignements sur les variables directes et dérivées, ainsi que leur univers respectif.

Aire de diffusion (AD) : Une aire de diffusion (AD) est une petite unité géographique relativement stable formée d'un ou de plusieurs îlots de diffusion avoisinants. Il s'agit de la plus petite région géographique normalisée pour laquelle toutes les données du recensement sont diffusées. Les AD couvrent tout le territoire du Canada.

Division de recensement (DR) : Division de recensement (DR) est le terme général de régions créées en vertu des lois provinciales (comme les comtés, les municipalités régionales de comté et les *regional districts*) ou des régions équivalentes. Les divisions de recensement sont des régions géographiques intermédiaires entre la municipalité (subdivision de recensement) et la province/territoire.

Logement : Ensemble de pièces d'habitation qu'une personne ou un groupe de personnes habite ou pourrait habiter.

Logement collectif : Établissement commercial, institutionnel ou communautaire que le recenseur peut identifier comme tel grâce à une enseigne ou en s'adressant à la personne qui en a la charge, à un résident, à un voisin, etc. Sont inclus dans cette catégorie les pensions et maisons de chambres, les hôtels, motels et maisons de chambres pour touristes, les établissements de soins infirmiers, les hôpitaux, les résidences de personnel, les casernes (bases militaires), les campements de travailleurs, les prisons, les foyers collectifs, etc. Il peut s'agir d'un logement occupé par des résidents habituels ou uniquement par des résidents étrangers et/ou par des personnes présentes temporairement.

Logement privé : Ensemble distinct de pièces d'habitation ayant une entrée privée donnant sur l'extérieur ou sur un corridor, un hall, un vestibule ou un escalier commun à l'intérieur. L'entrée doit donner accès au logement sans que l'on ait à passer par les pièces d'habitation de quelqu'un d'autre. Le logement doit répondre aux deux conditions qui le rendent propre à l'habitation durant toute l'année :

1. Avoir une source de chauffage ou d'énergie (comme en atteste la présence d'une cheminée, de fils électriques, de tuyaux ou compteurs pour l'huile [mazout] ou le gaz, d'une génératrice, de bois de chauffage, d'ampoules électriques, d'une thermopompe, de panneaux solaires, etc.).
2. Fournir un espace clos permettant de s'abriter des intempéries (comme en atteste la présence de murs d'enceinte et d'un toit ainsi que de portes et fenêtres offrant une protection contre le vent, la pluie et la neige).

Les logements qui ne répondent pas aux conditions les rendant propres à l'habitation durant toute l'année sont des logements marginaux. Les logements privés sont classés comme logements privés ordinaires et logements marginaux occupés. Les logements privés ordinaires se subdivisent à leur tour en trois grandes catégories : les logements occupés (par des résidents habituels), les logements occupés par des résidents étrangers et/ou temporaires et les logements inoccupés. Les logements marginaux sont classés comme logements occupés par des résidents habituels ou comme logements occupés par des résidents étrangers et/ou temporaires. Les logements marginaux inoccupés le 10 mai 2011 ne font pas partie du parc immobilier.

Logement privé occupé par des résidents habituels : Logement privé occupé de façon permanente par une personne ou un groupe de personnes. Sont également inclus dans cette catégorie les logements privés dont les résidents habituels sont temporairement absents le 10 mai 2011. Sauf indication contraire, toutes les données présentées dans les produits sur le logement ont trait aux logements privés occupés et non aux logements privés inoccupés ou aux logements occupés uniquement par des résidents étrangers et/ou temporaires.

Ménage : Personne ou groupe de personnes (autres que des résidents étrangers) occupant un même logement et n'ayant pas de lieu habituel de résidence ailleurs au Canada. Il peut se composer d'un groupe familial (famille de recensement) avec ou sans autres personnes, de deux familles ou plus partageant le même logement, d'un groupe de personnes non apparentées ou d'une personne seule. Les membres d'un ménage qui sont temporairement absents le

Rapport technique sur l'échantillonnage et la pondération

10 mai 2011 (par exemple, qui résident temporairement ailleurs) sont considérés comme faisant partie de leur ménage habituel. Chaque personne est membre d'un seul et unique ménage. À moins d'indications contraires, toutes les données contenues dans les rapports sur les ménages se rapportent aux ménages privés seulement.

Les ménages sont classés en trois catégories : les ménages privés, les ménages collectifs et les ménages à l'extérieur du Canada.

Ménage privé : Personne ou groupe de personnes (autres que des résidents étrangers) occupant un logement privé et n'ayant pas de domicile habituel ailleurs au Canada.

Secteur de recensement (SR) : Les secteurs de recensement (SR) sont de petites régions géographiques relativement stables qui comptent habituellement entre 2 500 et 8 000 habitants. Ils sont créés au sein de régions métropolitaines de recensement et d'agglomérations de recensement dont le noyau compte 50 000 habitants ou plus d'après le recensement précédent.

Un comité de spécialistes locaux (par exemple, des planificateurs, des travailleurs sociaux, des travailleurs du secteur de la santé et des éducateurs) délimite initialement les secteurs de recensement de concert avec Statistique Canada. Une fois qu'une région métropolitaine de recensement (RMR) ou qu'une agglomération de recensement (AR) a été subdivisée en secteurs de recensement, les secteurs de recensement sont maintenus même si, ultérieurement, la population du noyau de la RMR ou de l'AR devient inférieure à 50 000 habitants.

Subdivision de recensement (SDR) : Subdivision de recensement (SDR) est un terme générique qui désigne les municipalités (telles que définies par les lois provinciales/territoriales) ou les territoires considérés comme étant des équivalents municipaux à des fins statistiques (par exemple, les réserves indiennes, les établissements indiens et les territoires non organisés).

Unité de collecte (UC) : Petite unité géographique utilisée pour la collecte des données du recensement et de l'ENM. Les unités de collecte englobent le territoire du Canada au complet.

Annexe B Historique de l'échantillonnage au recensement canadien

L'échantillonnage a été utilisé pour la première fois au cours du recensement canadien de 1941. Un questionnaire sur le logement a été remis à tous les 10 logements. Les réponses aux 27 questions de ce questionnaire ont été intégrées à celles du questionnaire sur la population et les ménages correspondant à ces logements, ce qui a permis de faire le recoupement des caractéristiques-échantillon et des caractéristiques de base. Lors de ce même recensement, on a utilisé l'échantillonnage au cours de l'étape du dépouillement pour obtenir des estimations provisoires du revenu des salariés, de la répartition de la population en âge de travailler et de la composition des familles au Canada. Dans ce cas, l'échantillon était constitué du dixième de tous les secteurs de dénombrement du Canada et tous les questionnaires sur la population de ces secteurs ont été dépouillés à l'avance.

L'échantillonnage a également été utilisé pour le Recensement du logement de 1951. Cette fois, on a choisi chaque cinquième logement (ceux dont le numéro d'identification se terminait par 2 ou par 7) et le questionnaire sur le logement comportait 24 questions. Au cours du Recensement de 1961, on a prélevé un échantillon de 20 % des ménages privés et on a demandé aux personnes de 15 ans et plus de ces ménages de remplir un questionnaire-échantillon sur la population comportant des questions sur la migration interne, la fécondité et le revenu. L'échantillonnage n'a pas été utilisé au cours des recensements de moindre envergure de 1956 et de 1966.

Plusieurs innovations importantes ont été apportées aux méthodes de dénombrement lors du Recensement de 1971. La première a été l'utilisation de l'autodénombrement pour la majorité de la population au lieu de la méthode traditionnelle de recensement par interview. Ce changement découlait des résultats de plusieurs études effectuées au Canada et ailleurs (Fellegi, 1964; Hansen et al., 1959) selon lesquels la présence d'un agent recenseur avait un effet important sur la variance des chiffres du recensement lors d'un recensement par interview. On prévoyait que l'autodénombrement permettrait de réduire la variance des chiffres du recensement attribuable à la présence de l'agent recenseur et, en même temps, donnerait au répondant le temps voulu pour remplir le questionnaire en toute confidentialité; par conséquent, on s'attendait à ce que les réponses soient plus exactes que par le passé.

Un second changement survenu au Recensement de 1971 et qui le différencie de tous les recensements précédents a été la modification du contenu. Le nombre de sujets traités et de questions posées était plus élevé qu'aux recensements précédents. Une étude des coûts, du fardeau du répondant et des délais de production par rapport à la qualité des données dans le cadre d'un recensement par échantillonnage et par autodénombrement a présidé à la décision de recueillir toutes les données, sauf celles portant sur quelques caractéristiques de base, auprès d'un échantillon d'un tiers de la population au Recensement de 1971. Sauf dans les régions les plus isolées du Canada, le tiers des ménages privés a reçu le questionnaire complet comportant toutes les questions du recensement, tandis que le reste des ménages privés a reçu le questionnaire abrégé qui ne contenait que les questions de base soit le nom, le lien avec le chef de ménage, le sexe, la date de naissance, l'état matrimonial, la langue maternelle, le type de logement, le mode d'occupation, le nombre de pièces, l'alimentation en eau, les toilettes ainsi que certaines questions en rapport avec la couverture du recensement. Tous les ménages vivant dans des régions éloignées prédésignées et dans des logements collectifs ont reçu le questionnaire complet. Le rapport statistique « *Sampling in the Census* » (Bureau fédéral de la statistique, 1968) donne une description détaillée des projets d'application de cette méthode au Recensement de 1971.

Le contenu du Recensement de 1976 était beaucoup moindre que celui du Recensement de 1971. En outre, le questionnaire de 1976 ne contenait pas les questions qui posent le plus de difficulté de collecte (p. ex., les questions sur le revenu) ni celles pour lesquelles les frais de codage sont le plus élevés (p. ex., celles sur la profession, l'industrie et le lieu de travail). Par conséquent, les avantages de l'échantillonnage en termes de coûts et d'allègement du fardeau du répondant étaient moins évidents qu'au Recensement de 1971. Néanmoins, après avoir évalué les économies éventuelles reliées à divers taux d'échantillonnage et étudié les conséquences sur le plan des relations publiques du retour au dénombrement intégral (l'échantillonnage ayant fait ses preuves en 1971), il a été décidé d'appliquer en 1976 la même méthode d'échantillonnage qu'en 1971.

Au Recensement de 1981, on a utilisé la plupart des mêmes méthodes qu'en 1971 et 1976, sauf que le taux d'échantillonnage est passé du tiers au cinquième des ménages. Les études réalisées à cette époque ont révélé que la diminution de la qualité des données (mesurée au moyen de la variance) qui en résultait serait acceptable et qu'elle ne serait pas suffisamment importante pour annuler les avantages découlant de la réduction des coûts, du fardeau de

Rapport technique sur l'échantillonnage et la pondération

réponse et des délais de production (voir Royce, 1983). L'échantillonnage du cinquième des ménages a été la méthode retenue pour tous les recensements entre 1981 et 2006.

L'information obtenue antérieurement au moyen du questionnaire complet obligatoire du recensement a été recueillie dans le cadre de l'Enquête nationale auprès des ménages en 2011. En raison de ce changement, chaque ménage a dû répondre aux dix questions qui figuraient sur le questionnaire du recensement.

Rapport technique sur l'échantillonnage et la pondération

Annexe C Contraintes utilisées dans le processus de pondération

Voici la liste des 60 contraintes utilisées pour produire les poids des ménages pour l'ENM de 2011. Les 58 premières contraintes sont appliquées au niveau de la région de pondération (RP) seulement.

N°	Contrainte	Niveau	Description
1	TOTPERS	RP (Personne)	Nombre total de personnes
2	TPERGE15	RP (Personne)	Nombre total de personnes de 15 ans et plus
3	MALE	RP (Personne)	Hommes
4	MALEGE15	RP (Personne)	Hommes de 15 ans et plus
5	AGE4	RP (Personne)	Personnes de 0 à 4 ans
6	AGE9	RP (Personne)	Personnes de 5 à 9 ans
7	AGE14	RP (Personne)	Personnes de 10 à 14 ans
8	AGE19	RP (Personne)	Personnes de 15 à 19 ans
9	AGE24	RP (Personne)	Personnes de 20 à 24 ans
10	AGE29	RP (Personne)	Personnes de 25 à 29 ans
11	AGE34	RP (Personne)	Personnes de 30 à 34 ans
12	AGE39	RP (Personne)	Personnes de 35 à 39 ans
13	AGE44	RP (Personne)	Personnes de 40 à 44 ans
14	AGE49	RP (Personne)	Personnes de 45 à 49 ans
15	AGE54	RP (Personne)	Personnes de 50 à 54 ans
16	AGE59	RP (Personne)	Personnes de 55 à 59 ans
17	AGE64	RP (Personne)	Personnes de 60 à 64 ans
18	AGE74	RP (Personne)	Personnes de 65 à 74 ans
19	AGE75PL	RP (Personne)	Personnes de 75 ans et plus
20	SINGLE	RP (Personne)	Personnes célibataires (jamais mariées)
21	MARRIED	RP (Personne)	Personnes mariées
22	SEPARATED	RP (Personne)	Personnes séparées
23	DIVORCED	RP (Personne)	Personnes divorcées
24	WIDOWED	RP (Personne)	Personnes veuves
25	COMLAW	RP (Personne)	Personnes vivant en union libre
26	COUPLE	RP (Personne)	Personnes faisant partie d'un couple (mariage, partenaire)
27	CHILD	RP (Personne)	Nombre d'enfants
28	LONEPAR	RP (Personne)	Parents seuls
29	INEFAM	RP (Personne)	Personnes dans une famille économique
30	NOINEFAM	RP (Personne)	Personnes dans un ménage ne faisant pas partie d'une famille économique
31	NOTINFAM	RP (Personne)	Personnes dans un ménage ne faisant pas partie d'une famille de recensement
32	EN_EN_BI	RP (Personne)	LM = anglais, LPM = anglais, LO = bilingue *
33	EN_EN_EN	RP (Personne)	LM = anglais, LPM = anglais, LO = anglais *
34	FR_EN_BI	RP (Personne)	LM = français, LPM = anglais, LO = bilingue *
35	FR_FR_BI	RP (Personne)	LM = français, LPM = français, LO = bilingue *
36	FR_FR_FR	RP (Personne)	LM = français, LPM = français, LO = français *

Rapport technique sur l'échantillonnage et la pondération

Annexe C Contraintes utilisées dans le processus de pondération (suite)

N°	Contrainte	Niveau	Description
37	LAN_ABOR	RP (Personne)	Personnes dont la langue maternelle est une langue autochtone
38	LAN_AFR	RP (Personne)	Personnes dont la langue maternelle est une langue africaine
39	LAN_ARAB	RP (Personne)	Personnes dont la langue maternelle est l'arabe
40	LAN_CHIN	RP (Personne)	Personnes dont la langue maternelle est une langue chinoise
41	LAN_GER	RP (Personne)	Personnes dont la langue maternelle est une langue germanique
42	LAN_INDO	RP (Personne)	Personnes dont la langue maternelle est une langue indienne
43	LAN_IRAN	RP (Personne)	Personnes dont la langue maternelle est une langue iranienne
44	LAN_KOR	RP (Personne)	Personnes dont la langue maternelle est une langue coréenne
45	LAN_PILI	RP (Personne)	Personnes dont la langue maternelle est le Pilipino**
46	LAN_ROM	RP (Personne)	Personnes dont la langue maternelle est une langue romaine
47	LAN_SLAV	RP (Personne)	Personnes dont la langue maternelle est une langue slave
48	CHILDFAM	RP (Famille de recensement)	Familles de recensement avec enfants
49	NOCLDFAM	RP (Famille de recensement)	Familles de recensement sans enfant
50	TOTHHLD	RP (Ménage)	Nombre total de ménages
51	HHSIZE1	RP (Ménage)	Nombre de ménages de 1 personne
52	HHSIZE2	RP (Ménage)	Nombre de ménages de 2 personnes
53	HHSIZE3	RP (Ménage)	Nombre de ménages de 3 personnes
54	HHSIZE4	RP (Ménage)	Nombre de ménages de 4 personnes
55	HHSIZE5	RP (Ménage)	Nombre de ménages de 5 personnes
56	HHSIZEG6	RP (Ménage)	Nombre de ménages de 6 personnes et plus
57	SINGLDET	RP (Ménage)	Nombre de ménages vivant dans un logement individuel non attenant
58	APTLT5	RP (Ménage)	Nombre de ménages vivant en appartement dans un immeuble de moins de 5 étages
59	HHDACT	AD	Nombre de ménages dans chaque aire de diffusion de la RP
60	PPDACT	AD	Nombre de personnes dans chaque aire de diffusion de la RP

Notes :

* LM signifie langue maternelle, LPM signifie langue parlée à la maison et LO signifie langue officielle. Bilingue fait référence à la capacité de parler l'anglais et le français.

** Le tagalog n'était pas inclus dans la contrainte Pilipino.

Rapport technique sur l'échantillonnage et la pondération

Bibliographie

CHEPITA, R. 2012. *Non-Response Follow-up Strategy and Sampling Design for the 2011 National Household Survey*, rapport interne, Statistique Canada, Ottawa, Ontario.

BUREAU FÉDÉRAL DE LA STATISTIQUE. 1968. *Sampling in the Census*, rapport interne, Statistique Canada, Ottawa, Ontario.

FELLEGI, I.P. 1964. « Response Variance and its Estimation », *Journal of the American Statistical Association*, vol. 59, décembre, p. 1016 à 1041.

HANSEN, M.H. et W.N. HURWITZ. 1946. « The Problem of Non-Response in Sample Surveys ». *Journal of the American Statistical Association*, vol. 41, p. 517 à 529.

HANSEN, M.H., W.N. HURWITZ et M.A. BERSHAD. 1959. « Measurement Errors in Censuses and Surveys ». *Bulletin of the International Statistical Institute*, vol. 38, p. 359 à 374.

KISH, L. 1965. *Survey Sampling*, New York, Wiley.

RODRIGUE, J.-F., P. MATHIEU, J.-P. MORIN et J. TAYLOR. 2012. « La méthodologie de collecte du Recensement et de l'Enquête nationale auprès des ménages de 2011 au Canada », Journées de méthodologie statistique, Paris, France.

ROYCE, D. 1983. *The Use of Sampling in the 1981 Canadian Census*, rapport interne, Statistique Canada, Ottawa, Ontario.

STATISTIQUE CANADA. 2009. *Rapport technique du Recensement de 2006 : Échantillonnage et pondération*, produit n° 92-568-X au catalogue de Statistique Canada, Ottawa, Ontario.

—. 2012. *Dictionnaire du recensement*, produit n° 98-301-X2011001 au catalogue de Statistique Canada, Ottawa, Ontario.

—. 2013. *Guide de l'utilisateur de l'Enquête nationale auprès des ménages*, produit n° 99-001-XWF2011001 au catalogue de Statistique Canada, Ottawa, Ontario.

—. 2015. *Rapport technique sur la couverture*, Recensement de 2011, produit n° 98-303-X2011001 au catalogue de Statistique Canada, Ottawa, Ontario.

—. 2014. *SCANCIR, version 5.2*, Guide de base de l'utilisateur. Ottawa, Ontario.

VERRET, F. 2013. « The estimation methodology of the 2011 National Household Survey », dans *JSM Proceedings*, Survey Research Methods Section, Montréal, Québec, Canada, American Statistical Association.