
Rapport technique sur l'échantillonnage et la pondération

Recensement de la population, 2021



Date de diffusion : le 15 novembre 2023



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 suivants :

- | | |
|---|----------------|
| • Service de renseignements statistiques | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur | 1-514-283-9350 |

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 « Contactez-nous » > « [Normes de service à la clientèle](#) ».

Note de reconnaissance

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

Publication autorisée par le ministre responsable de Statistique Canada

© Sa Majesté le Roi du chef du Canada, représenté par le ministre de l'Industrie, 2023

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.

Date de diffusion : le 15 novembre 2023

N° 98-306-X au catalogue, numéro 2021001

ISBN 978-0-660-47836-4

Table des matières

Introduction	1
1. Collecte des données du recensement	2
1.1 Méthodes de livraison dans le cadre du recensement	2
1.2 Approche de collecte par vagues	2
1.3 Vérification de l'occupation et activités de suivi pour le Recensement de 2021	4
1.4 Questionnaires du Recensement de la population	4
1.5 Taux de réponse à la collecte	6
2. Échantillonnage	7
2.1 Univers pour l'échantillon du questionnaire détaillé	7
2.2 Plan d'échantillonnage pour le questionnaire détaillé	7
3. Traitement des données du recensement	8
3.1 Introduction	8
3.2 Réception et enregistrement	8
3.3 Imagerie et saisie à partir d'images	8
3.4 Contrôles de couverture, contrôles d'intégralité et suivi des questionnaires rejetés au contrôle	9
3.5 Codage	10
3.6 Classification et ajustements pour la non-réponse des logements inoccupés et des logements non répondants	11
3.7 Utilisation des données administratives	12
3.8 Vérification et imputation	13
3.9 Non-réponse	13
3.10 Pondération	14
3.11 Taux de réponse finaux	14
4. Estimation à l'aide de l'échantillon du questionnaire détaillé du recensement	16
4.1 Éléments à considérer pour choisir une méthode d'estimation	16
4.2 Régions de pondération	17
4.3 Poids de sondage	22
4.4 Ajustement pour la couverture et pour la non-réponse totale	23
4.5 Calage final	25
4.6 Détails sur la sélection des contraintes	27
5. Évaluation des procédures de pondération	30
5.1 Répartition des poids	30
5.2 Écarts entre les chiffres du recensement et les estimations pour le questionnaire détaillé, Canada	32

6. Estimation de la variance	42
6.1 Éléments à considérer pour choisir une méthode d'estimation de la variance.....	42
6.2 Estimateur de variance	44
6.3 Ajustement des poids de répliques	45
7. Inférence statistique	47
7.1 Intervalles de confiance et leur interprétation	47
7.2 Construction d'intervalles de confiance.....	49
7.3 Intervalle de confiance de Student.....	50
7.4 Intervalle de confiance de Wilson modifié pour les proportions	51
7.5 Intervalle de confiance de Wilson modifié pour les comptes	52
8. Conclusion	56
Annexe A – Glossaire	57
Annexe B – Historique de l'échantillonnage dans le cadre du recensement canadien	60
Annexe C – Contraintes utilisées ou exclues dans le processus de pondération	62
Bibliographie	75

Introduction

Lors du Programme du recensement de 2021, les ménages canadiens ont été dénombrés à l'aide de deux principaux types de questionnaires : le questionnaire abrégé et le questionnaire détaillé. En 2021, un échantillon de 25 % des ménages canadiens a reçu un questionnaire détaillé, lequel comprenait aussi les questions du questionnaire abrégé. Les autres ménages ont reçu un questionnaire abrégé.

Outre les questions du questionnaire abrégé, le questionnaire détaillé comprend un ensemble de questions visant à broser un portrait complet de la population et des ménages canadiens, selon leurs caractéristiques démographiques, sociales et économiques.

Les estimations produites à partir de réponses aux questions posées dans les deux types de questionnaires sont fondées sur l'ensemble de la population, et ont été obtenues au moyen d'un recensement. En effet, tous les ménages répondants aux deux types de questionnaires contribuent à un chiffre donné. C'est le cas, par exemple, du chiffre de population pour un groupe d'âge donné.

Les estimations produites à partir de réponses à au moins une question posée uniquement dans le questionnaire détaillé sont fondées sur l'échantillon. Dans ces cas, seuls les ménages répondants de l'échantillon du questionnaire détaillé contribuent à l'estimation. L'estimation selon le plus haut niveau de scolarité est un exemple d'une telle estimation.

L'échantillon du questionnaire détaillé est réparti uniformément au niveau géographique afin d'assurer une bonne fiabilité des estimations pour toutes les régions du pays et d'accorder la même importance relative à toutes les unités géographiques d'une taille donnée.

Le présent rapport technique porte sur la méthodologie utilisée dans le processus d'estimation à partir de l'échantillon du questionnaire détaillé du Recensement de la population de 2021. Le [chapitre 1](#) décrit en détail les méthodes de collecte utilisées dans le cadre du recensement et de l'échantillon du questionnaire détaillé. Le [chapitre 2](#) décrit comment l'échantillonnage a été appliqué au questionnaire détaillé. Le [chapitre 3](#) explique les procédures de traitement des données. Le [chapitre 4](#) décrit les procédures utilisées pour assigner un poids aux unités répondantes de l'échantillon du questionnaire détaillé afin d'obtenir des estimations pour la population. Le [chapitre 5](#) porte sur diverses évaluations des procédures de pondération, tandis que le [chapitre 6](#) décrit la méthodologie d'estimation de la variance utilisée pour l'échantillon du questionnaire détaillé du Recensement de 2021. Le [chapitre 7](#) présente le concept d'inférence statistique à l'aide d'intervalles de confiance. Une conclusion suit.

1. Collecte des données du recensement

Pendant l'étape de la collecte des données, l'objectif était de veiller à ce que des réponses soient obtenues de la part de tous les ménages au Canada. Les opérations sur le terrain comprenaient le listage des logements, l'envoi des lettres d'invitation, la détermination de l'occupation/inoccupation d'un logement et la tenue d'interviews avec les non-répondants.

1.1 Méthodes de livraison dans le cadre du recensement

Pour la plupart des logements privés, on a demandé aux répondants de remplir le questionnaire pour eux-mêmes et pour tous les membres de leur ménage.

Le 3 mai 2021, tous les logements privés dans les secteurs d'envoi par la poste (SEP) — c'est-à-dire environ 86 % des logements privés au Canada — ont reçu par la poste une lettre bilingue les invitant à remplir leur questionnaire en ligne. Comme en 2016, la lettre contenait un code d'accès sécurisé (CAS), l'adresse du site Web du Recensement de 2021 et un numéro de téléphone à composer pour demander un questionnaire papier, le cas échéant.

Dans les secteurs de listage/livraison (L/L), qui représentent 7 % des logements, les employés du recensement ont livré des lettres d'invitation. La livraison porte-à-porte dans les secteurs de L/L a eu lieu du 3 au 10 mai 2021. Au cours de cette opération, les employés du recensement ont dressé la liste de tous les logements privés de secteurs particuliers dans leur registre des visites. La lettre d'invitation comportait un CAS afin que les répondants puissent remplir le questionnaire en ligne. Il était possible d'obtenir un questionnaire papier sur demande, en composant un numéro sans frais. Dans les secteurs L/L, le répondant devait fournir une adresse postale à un téléphoniste pour recevoir le questionnaire papier par la poste.

En 2021, la méthode d'envoi par la poste et livraison à la porte (EPLP) a vu le jour. Les secteurs d'EPLP sont ceux où tous les logements disposent d'une adresse et dont la majorité peut recevoir du courrier. Dans ces secteurs mixtes, les logements ayant une adresse postale valide ont reçu les envois par la poste (EP) habituels (tout comme les SEP), tandis que les logements n'ayant pas d'adresse valide (dont l'adresse postale ne correspondait pas à l'adresse municipale) ont reçu une lettre d'invitation déposée à leur porte par un employé du recensement. Les secteurs d'EPLP ont été mis en place pour maximiser le nombre de logements d'EP pour le recensement. Les secteurs d'EPLP représentent plus de 6 % des logements et ont permis une augmentation de l'utilisation de la méthode d'EP pour atteindre environ 90 % des logements (par rapport à 82 % en 2016).

Traditionnellement, les logements restants, situés dans les communautés des Premières Nations, les établissements métis, les régions inuites et d'autres régions éloignées, sont dénombrés en personne au moyen de méthodes de recensement par interview. Toutefois, pour la première fois en 2021, toutes les communautés des Premières Nations, les établissements métis, les régions inuites et d'autres régions éloignées ont eu la possibilité de s'autodéclarer, tant que c'était possible sur le plan opérationnel (p. ex. la communauté avait accès à Internet). En fonction de la situation, la lettre d'invitation de la méthode EP, L/L ou EPLP a été utilisée (avec des changements mineurs, p. ex. questionnaire papier non offert) avant le suivi des cas de non-réponse (SCNR). Les ménages dans les régions où il n'était pas possible sur le plan opérationnel d'offrir l'autodéclaration ont rempli leur questionnaire du recensement avec l'aide d'un employé du recensement (en personne ou par téléphone). En 2021, les logements situés dans les communautés éloignées, nordiques ou autochtones représentent environ 1 % des logements au Canada.

1.2 Approche de collecte par vagues

Pour le Recensement de 2021, Statistique Canada a adopté une approche par vagues qui consistait, par divers moyens de communication à des moments précis tout au long de la période de collecte, à rappeler aux Canadiens et Canadiennes de remplir leur questionnaire. L'organisme a également encouragé les répondants à remplir leur questionnaire en ligne, mais il a diminué le risque d'une diminution du taux de réponse global en offrant également d'autres moyens de répondre, comme l'obtention d'un questionnaire papier. Le tableau suivant présente les principales dates des différentes vagues.

Rapport technique sur l'échantillonnage et la pondération

Tableau 1.2.1
Étapes et calendrier de la collecte des données du recensement

Étape de la collecte	Activité principale	Couverture	Date de début
Première vague — Lettre d'invitation	Les logements des secteurs d'EP reçoivent une lettre d'invitation contenant un CAS.	Tous les logements des secteurs d'EP, y compris ceux des secteurs d'EPLP (90 % de tous les logements)	3 mai 2021
	Les logements des secteurs de LL et les logements destinés à la livraison à la porte dans les secteurs d'EPLP reçoivent une lettre d'invitation contenant un CAS.	Tous les logements des secteurs de LL et les logements destinés à la livraison à la porte dans les secteurs d'EPLP (9 % de tous les logements)	3 mai 2021
Deuxième vague — Lettre ou carte de rappel	Les logements des secteurs d'EP reçoivent une lettre de rappel contenant un CAS.	Tous les logements non répondants des secteurs d'EP, y compris ceux des secteurs d'EPLP	12 mai 2021
	Les logements des secteurs de LL reçoivent une carte de rappel.	Tous les logements des secteurs de LL	12 mai 2021
Troisième vague — Deuxième lettre de rappel	Les logements des secteurs d'EP reçoivent une deuxième lettre de rappel contenant un CAS.	Tous les logements non répondants des secteurs d'EP, y compris ceux des secteurs d'EPLP	21 mai 2021
Message de rappel	Les logements des secteurs d'EP reçoivent un rappel par message texte (si un numéro de téléphone cellulaire est disponible), par message vocal enregistré (si un numéro de téléphone fixe est disponible) ou par courriel (si une adresse courriel est disponible).	Tous les logements non répondants des secteurs d'EP, y compris ceux des secteurs d'EPLP	30 mai 2021
Suivi des cas de non-réponse	Début du SCNR par téléphone ou en personne dans les secteurs de LL.	Tous les logements non répondants des secteurs de LL	21 mai 2021
	Début du SCNR par téléphone ou en personne dans les secteurs d'EP et d'EPLP.	Tous les logements non répondants des secteurs d'EP et d'EPLP	2 juin 2021
Lettre d'avis final	Les logements des secteurs d'EP reçoivent une lettre d'avis final contenant un CAS.	Tous les logements non répondants des secteurs d'EP, y compris ceux des secteurs d'EPLP	13 juillet 2021

LL = Listage/livraison

EP = Envoi par la poste

EPLP = Envoi par la poste et livraison à la porte

SCNR = Suivi des cas de non-réponse

CAS = Code d'accès sécurisé

Source : Statistique Canada, Recensement de la population, 2021.

Dans les communautés des Premières Nations, les établissements métis, les régions inuites et d'autres régions éloignées, selon la situation, une lettre d'invitation a été livrée, par la poste ou en personne, avant l'étape de suivi des cas de non-réponse, qui a commencé le 14 mai 2021. À partir du 3 août, une lettre de rappel a aussi été livrée

aux ménages non répondants dans les secteurs d'envoi par la poste (SEP). En l'absence d'un accès Internet, les questionnaires ont été remplis en personne avec l'aide d'un employé du recensement de Statistique Canada, et ce, à partir du 3 mai 2021.

Assistance téléphonique du recensement

L'Assistance téléphonique du recensement est un service gratuit multilingue qui est mis à la disposition de tous les répondants à l'échelle du pays. Le numéro sans frais figurait dans tous les documents de communication du recensement.

1.3 Vérification de l'occupation et activités de suivi pour le Recensement de 2021

Vérification de l'occupation des appartements (VOA) — La VOA vise à vérifier l'occupation/inoccupation de toutes les unités d'un immeuble d'appartements, par l'entremise d'une personne-ressource chargée de la gestion. Les renseignements ont été recueillis au moyen d'une interview téléphonique avec la personne-ressource. Cette dernière pouvait être, par exemple, le propriétaire, le concierge ou le gestionnaire de l'immeuble. La VOA est une activité importante, car elle aide à déterminer avec plus de précision l'occupation/inoccupation de ces types de logements et à réduire la charge de travail dans le cadre de l'activité de SCNR du recensement. La VOA a été effectuée par les téléphonistes de la Sous-section du soutien à la collecte du 10 au 18 mai 2021.

Vérification de l'occupation des logements — Pour un échantillon des logements des SEP, l'occupation/inoccupation a été vérifiée immédiatement avant le SCNR. La vérification de l'occupation des logements a été effectuée du 21 au 28 mai 2021 pour déterminer le plus grand nombre possible de logements inoccupés ou annulés autour du jour du recensement, soit le 11 mai 2021, afin d'éliminer ces logements de la charge de travail du SCNR. L'occupation/inoccupation des logements peut être déterminée avec une plus grande exactitude lorsque la vérification est effectuée plus près du jour du recensement. Cette opération est indépendante de la VOA décrite ci-dessus.

Suivi des cas de non-réponse — Le SCNR vise à obtenir un questionnaire rempli pour tous les ménages n'ayant pas retourné le leur. Le suivi a été effectué par téléphone ou en personne. Dans les secteurs de L/L, le suivi a été effectué du 21 mai au 13 août 2021, et dans les SEP et les secteurs d'EPLP, du 2 juin au 13 août 2021. Dans les secteurs de recensement par interview et les réserves, le SCNR s'est déroulé du 14 mai au 24 septembre 2021. En l'absence d'un accès Internet, les questionnaires ont été remplis en personne avec l'aide d'un employé du recensement, et ce, à partir du 3 mai 2021.

1.4 Questionnaires du Recensement de la population

La majeure partie de la population canadienne réside dans des logements privés. Les données du recensement des résidents des logements privés sont recueillies principalement en demandant à un membre adulte du ménage de remplir un questionnaire en ligne pour l'ensemble du ménage.

Le recensement constitue la principale source de données démographiques exhaustives au Canada. En 2021, le questionnaire du recensement a permis de recueillir les renseignements suivants :

Les renseignements recueillis à partir des questionnaires abrégé et détaillé comprennent les éléments suivants :

- adresse;
- noms des résidents habituels;
- date de naissance et âge;
- sexe à la naissance et genre;
- liens entre les membres du ménage (y compris l'état matrimonial ou l'union libre);
- connaissance des langues officielles;
- langues parlées régulièrement à la maison et langue parlée le plus souvent à la maison;

- première langue apprise à la maison dans l'enfance;
- instruction dans la langue officielle minoritaire;
- expérience militaire canadienne.

Les renseignements recueillis uniquement à partir du questionnaire détaillé comprennent les éléments suivants :

- activités de la vie quotidienne;
- lieu de naissance de la personne et des parents;
- citoyenneté;
- connaissances de langues non officielles;
- origines ethniques ou culturelles;
- identité : Première Nation, Métis ou Inuk (Inuit);
- groupes de population;
- statut d'Indien inscrit ou des traités;
- appartenance à une Première Nation ou à une bande indienne;
- appartenance à une organisation métisse ou à un établissement métis;
- inscription en vertu d'un accord sur les revendications territoriales des Inuit ou bénéficiaire d'un tel accord;
- religion;
- mobilité (un an et cinq ans);
- scolarité;
- activités sur le marché du travail;
- langue de travail;
- lieu de travail et navettage;
- dépenses (garde d'enfants, pensions alimentaires pour les enfants et le conjoint);
- logement.

La plupart des données du recensement ont été recueillies au moyen du questionnaire abrégé ou du questionnaire détaillé. Un échantillon de 25 % des ménages canadiens ont reçu un questionnaire détaillé en 2021.

1.4.1 Questionnaire abrégé (questionnaires 2A, 3A et 2C)

[Questionnaire 2A](#) :

Ce questionnaire abrégé sert à dénombrer les résidents habituels de tous les logements privés.

[Questionnaire 3A](#) :

Ce questionnaire abrégé pour les particuliers (comparable au questionnaire 2A) sert à dénombrer une seule personne. Il est fourni aux résidents habituels des logements privés qui souhaitent être dénombrés séparément des autres membres du ménage (p. ex. colocataires, pensionnaires). Il sert également à dénombrer les résidents de certains logements collectifs comme des maisons de chambres et pensions.

[Questionnaire 2C](#) :

Ce questionnaire abrégé pour les personnes vivant à l'étranger (comparable au questionnaire 2A) sert à dénombrer les personnes qui sont temporairement à l'étranger au moment du recensement. Pour 2021, cela comprend les fonctionnaires canadiens (fédéraux et provinciaux) et leur famille, et les membres des Forces armées canadiennes et leur famille.

1.4.2 Questionnaire détaillé (questionnaires 2A-L et 2A-R)

Le questionnaire détaillé sert à compléter les données recueillies au moyen du questionnaire abrégé et vise à fournir des renseignements plus détaillés sur les personnes au Canada en fonction de leurs caractéristiques démographiques, sociales et économiques.

[Formulaire 2A-L](#) :

Ce questionnaire détaillé est le plus souvent utilisé.

[Formulaire 2A-R](#) :

Ce questionnaire est comparable au questionnaire 2A-L, mais il est uniquement utilisé dans les communautés éloignées, nordiques et autochtones. Il comporte les questions du questionnaire détaillé de même que des exemples adaptés aux communautés des Premières Nations, aux établissements métis, aux régions inuites et à d'autres régions éloignées. Ce questionnaire comprend aussi deux questions supplémentaires sur le logement de bande. Pour 2021, il y a une nouvelle question sur les frais de logement de bande.

1.5 Taux de réponse à la collecte

Le taux global de réponse à la collecte du Recensement de la population de 2021 était de 98,0 %. Ce taux a été produit directement à partir des résultats de la collecte, c'est-à-dire avant que le traitement des données et la vérification de la qualité des données ne soient terminés. Il représente le nombre de logements privés pour lesquels un questionnaire a été retourné, divisé par le nombre de logements privés codés comme occupés par les agents recenseurs. Le taux de réponse à la collecte au sein de l'échantillon du questionnaire détaillé était de 97,4 % (pour obtenir plus de détails, voir les [taux de réponse de la collecte du Recensement de la population de 2021](#)).

2. Échantillonnage¹

Lorsqu'une enquête-échantillon est menée, il est important de bien planifier la sélection de l'échantillon. Durant le processus d'échantillonnage, un sous-ensemble de la population cible de l'enquête est choisi pour recevoir le questionnaire. Les réponses du sous-ensemble permettent de produire des inférences sur l'ensemble de la population. On distingue deux types d'échantillonnage : l'échantillonnage probabiliste et l'échantillonnage non probabiliste. L'échantillonnage probabiliste est préférable lorsqu'il est important de produire des inférences statistiques sur la population entière, car il est alors possible de calculer la probabilité de sélection des unités et d'estimer l'erreur d'échantillonnage. Le présent chapitre porte sur la sélection de l'échantillon qui a reçu le questionnaire détaillé du Recensement de 2021.

2.1 Univers pour l'échantillon du questionnaire détaillé

L'univers des ménages du recensement a été divisé en trois parties : les ménages privés, les ménages collectifs et les ménages situés à l'extérieur du Canada. L'univers pour l'échantillon du questionnaire détaillé est uniquement formé des ménages privés, y compris les ménages vivant dans des logements privés annexés aux logements collectifs du Canada. Cet univers exclut les réserves et établissements partiellement dénombrés. Sauf indication contraire, l'expression « dans le champ de l'enquête » sera utilisée pour indiquer qu'un ménage fait partie de l'univers pour l'échantillon du questionnaire détaillé (c.-à-d. les ménages privés qui ne sont pas dans des réserves et établissements partiellement dénombrés). L'expression « hors du champ de l'enquête », quant à elle, fait référence aux ménages non compris dans l'univers (c.-à-d. les ménages vivant dans des logements collectifs, ceux à l'extérieur du Canada et ceux vivant dans des réserves et établissements partiellement dénombrés).

2.2 Plan d'échantillonnage pour le questionnaire détaillé

Dans la plupart des cas, le questionnaire détaillé a été envoyé au quart des ménages de l'univers du questionnaire détaillé afin de recueillir des données démographiques et socioéconomiques à propos de la population canadienne. L'échantillon a été tiré de la liste des logements du Recensement de la population de 2021. Au moment de la sélection de l'échantillon, les adresses correspondant à des logements hors du champ de l'enquête étaient inconnues, ce qui signifie que certains logements ont reçu un questionnaire détaillé 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 plan d'échantillonnage utilisé pour la sélection des logements qui reçoivent le questionnaire détaillé était un plan stratifié et systématique. Pour définir le plan de sondage et améliorer l'efficacité des opérations sur le terrain, on divise le Canada en petites régions géographiques appelées unités de collecte (UC). Chaque UC se voit attribuer l'une des méthodes de livraison décrites au [chapitre 1](#). Les strates du plan d'échantillonnage sont les UC. Dans les UC où le mode de collecte est l'envoi par la poste, le listage et livraison ou l'envoi par la poste et livraison à la porte, l'échantillonnage était systématique et la fraction de sondage était d'un quart. Le point de départ de la sélection était aléatoire. Dans les UC situées dans les communautés des Premières Nations, les établissements métis, les régions inuites et les autres régions éloignées, tous les ménages ont été sélectionnés. Ces UC étaient des strates à tirage complet.

Il y avait une exception au plan d'échantillonnage. En effet, les logements privés annexés à des logements collectifs ont été ajoutés à l'échantillon avec certitude. Cependant, ils n'ont rempli que le questionnaire abrégé. Les réponses au questionnaire détaillé ont par la suite été imputées pour ces ménages.

À l'exception des ménages privés annexés à des logements collectifs, on a demandé à tous les ménages sélectionnés dans l'échantillon de remplir le questionnaire détaillé du recensement. On a demandé aux ménages des logements privés qui ne faisaient pas partie de l'échantillon du questionnaire détaillé de remplir le questionnaire abrégé.

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

3. Traitement des données du recensement

3.1 Introduction

Le présent chapitre est consacré au traitement de tous les questionnaires remplis, quel que soit leur type, depuis la réception des questionnaires jusqu'à la création d'une base de données du recensement exacte et complète. On trouvera ci-après la description des étapes de l'enregistrement des questionnaires, de l'imagerie et de la saisie des données des questionnaires, du contrôle, de la correction des erreurs, du suivi des questionnaires rejetés au contrôle, du codage, de la classification des logements et des ajustements pour la non-réponse, du couplage des données administratives, de l'imputation, de la pondération et des taux de réponse finaux.

Les processus automatisés, mis en œuvre pour le Recensement de 2021, ont dû être surveillés de manière à ce que toutes les résidences canadiennes soient dénombrées une seule fois. Le Système de contrôle principal (SCP) a été mis au point 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 identificateur unique. Ce système était mis à jour régulièrement, en fonction des renseignements sur la situation de chaque logement dans le processus d'enchaînement des opérations du recensement (p. ex. livraison, réception ou traitement). Des rapports étaient produits quotidiennement et mis en ligne à la disposition des gestionnaires, afin que les opérations du recensement se déroulent de façon efficace et efficiente.

3.2 Réception et enregistrement

Les réponses reçues par Internet ou recueillies par interview téléphonique au moyen de l'Assistance téléphonique du recensement arrivaient directement au Centre des opérations des données (COD) et leur réception était enregistrée automatiquement.

Les répondants qui avaient rempli un questionnaire papier devaient le retourner au COD par la poste. Postes Canada enregistrait automatiquement leur réception dans plusieurs emplacements au Canada (dans le cadre du processus ordinaire de traitement du courrier) en utilisant la fenêtre transparente des enveloppes-réponses pour numériser le code à barres qui figurait à la première page du questionnaire. Les enveloppes étaient ensuite livrées au COD chaque jour ouvrable. En outre, Postes Canada envoyait quotidiennement un fichier qui énumérait tous les questionnaires du recensement reçus à chaque établissement régional de traitement du courrier, selon la date de réception.

L'enregistrement de chaque questionnaire retourné était signalé dans le SCP à Statistique Canada. Le SCP produisait quotidiennement une liste de tous les logements pour lesquels un questionnaire n'avait pas été reçu, et la transmettait aux opérations sur le terrain pour éviter d'effectuer un suivi des cas de non-réponse auprès des ménages qui avaient déjà rempli leur questionnaire.

3.3 Imagerie et saisie à partir d'images

En 2021, tous les questionnaires papier du recensement (2A, 2C, 2A-L, 2A-R, 3A) ont été convertis en images. 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 tout corps étranger (comme des trombones ou des agrafes) en vue de la numérisation. Les questionnaires ont été mis en lots selon le type de formulaire, et divisés en feuilles individuelles en supprimant le dos de la reliure.
- Numérisation : Les questionnaires étaient convertis en images numériques.
- 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 qui contenaient des anomalies ont été présentées à un opérateur chargé de les examiner. L'opérateur pouvait accepter l'image telle quelle et l'envoyer directement à la saisie des données par clavier (passant outre la reconnaissance automatisée), ou envoyer le questionnaire entier à l'étape de la vérification à la sortie pour être retiré. Les étapes de la saisie des données par clavier et la vérification à la sortie sont décrites ci-dessous.
- Reconnaissance automatisée : Cette étape visait à reconnaître automatiquement les réponses écrites à la main et les marques inscrites sur le questionnaire.
- Saisie des données par clavier : Des opérateurs ont entré les réponses que le système de reconnaissance automatisée ne pouvait pas saisir avec suffisamment de précision. Environ 12 % de toutes les réponses ont été transmises à la saisie.
- Vérification à la sortie : Après avoir franchi toutes les étapes susmentionnées, les questionnaires papier ont été retirés du système. La vérification à la sortie est un processus de contrôle de la qualité destiné à s'assurer que les images et les données saisies sont d'une qualité suffisante de façon à ce que les questionnaires papier ne nécessitent pas un traitement ultérieur. Les questionnaires dans lesquels on a décelé des erreurs ont été retirés à la sortie et traités à nouveau.

3.4 Contrôles de couverture, contrôles d'intégralité et suivi des questionnaires rejetés au contrôle

À cette étape, un certain nombre de contrôles automatisés ont été exécutés sur les données des répondants. Ces contrôles ont été conçus pour déceler les cas où le nombre de personnes dénombrées dans le ménage était incorrect en raison d'une erreur de collecte, d'une erreur du répondant ou d'une erreur de saisie des données. La plupart de ces erreurs se sont produites dans les questionnaires papier et comprennent :

- les données entrées par erreur dans la mauvaise colonne relative à la personne;
- les données barrées ou rayées qui ont été saisies par erreur;
- les données qui n'ont pas été fournies pour chaque membre du ménage figurant sur la liste au début du questionnaire.

Des erreurs qui peuvent se produire tant dans les questionnaires en ligne que dans les questionnaires papier comprennent :

- des données sont fournies pour la même personne dans plus d'un questionnaire (p. ex. une personne remplit son propre questionnaire 3A, mais est également inscrite sur le questionnaire 2A de son ménage);
- des questionnaires sont reçus en double (p. ex. une personne répond au questionnaire sur Internet et son conjoint ou sa conjointe remplit la version papier et l'envoie par la poste).

Environ 54 % des rejets au contrôle ont été résolus automatiquement par le système, lorsque la nature de l'erreur était telle que la solution était évidente. Les solutions comprenaient la suppression des données erronées sur les personnes provenant d'une erreur du répondant ou d'une erreur de saisie, et la suppression des réponses en double. Le reste des rejets au contrôle ont été transmis aux commis au dépouillement. Un système interactif permettait aux commis de comparer les données saisies d'un questionnaire à l'autre et d'examiner l'image des questionnaires papier pour détecter les erreurs de saisie et les erreurs des répondants. Les rejets au contrôle ont été résolus en supprimant les personnes non valides ou en double ou en ajoutant les personnes manquantes (c.-à-d. en créant des enregistrements de personne vides) lorsque nécessaire.

Les contrôles de couverture ont été suivis d'une autre série de contrôles automatisés, afin de déceler les questionnaires où les réponses manquantes étaient trop nombreuses, ou lorsque les données n'avaient pas été fournies pour tous les membres habituels du ménage, y compris les cas où des personnes manquantes avaient été ajoutées par les commis aux contrôles de couverture. Les ménages rejetés au contrôle ont fait l'objet d'un suivi. Un intervieweur téléphonait aux répondants pour régler les problèmes de couverture et obtenir les renseignements manquants, au moyen d'une application d'interview téléphonique assistée par ordinateur. Dans le cas des ménages qui avaient répondu au questionnaire détaillé, seules les données manquantes aux questions qui figuraient également dans le questionnaire abrégé ont fait l'objet d'un suivi. Les données obtenues lors du

suivi étaient intégrées au système en vue d'un traitement subséquent. Dans les cas où le suivi n'a pas permis d'obtenir les données manquantes, les données ont été imputées à l'étape de la vérification et de l'imputation (voir la [section 3.8](#)).

3.5 Codage

Les questionnaires du recensement contenaient des questions dont les réponses pouvaient être sélectionnées dans une liste, ainsi que des questions dont les réponses devaient être inscrites par le répondant. Dans la mesure du possible, les réponses écrites se sont vu attribuer automatiquement un code numérique au moyen des fichiers de référence, des ensembles de codes et des classifications normalisées de Statistique Canada. Les fichiers de référence utilisés pour le processus d'appariement automatisé ont été créés au moyen de réponses réelles recueillies lors de recensements antérieurs ou d'autres enquêtes mesurant les mêmes concepts, et à l'aide de fichiers administratifs. Dans les cas où l'attribution automatique d'un code n'était pas possible, l'attribution des codes a été effectuée au moyen de modèles d'apprentissage automatique élaborés à l'aide de l'algorithme de traitement du langage naturel « fastText »². Enfin, les enregistrements auxquels un code n'a pas été attribué automatiquement par un fichier de référence ou un modèle d'apprentissage automatique ont été codés par des codeurs spécialement formés et des spécialistes du domaine.

Les questions suivantes nécessitaient un codage tant dans le questionnaire détaillé que dans le questionnaire abrégé :

- genre;
- lien avec la Personne 1;
- langue parlée à la maison;
- langue maternelle;
- instruction dans la langue officielle minoritaire.

Les questions suivantes nécessitaient un codage pour l'échantillon du questionnaire détaillé seulement :

- lieu de naissance de la personne;
- lieu de naissance des parents;
- citoyenneté;
- connaissance de langues non officielles;
- origines ethniques ou culturelles;
- groupe de population;
- religion;
- Première Nation/bande indienne;
- lieu de résidence un an auparavant;
- lieu de résidence cinq ans auparavant;
- principal domaine d'études;
- lieu des études;
- industrie;
- profession;
- lieu de travail;

2. Joulin, A., E. Grave, P. Bojanowski et T. Mikolov. 2016. « Bag of Tricks for Efficient Text Classification », *arXiv preprint arXiv:1607.01759v3*.

- revendications territoriales des Inuit;
- raison principale pour travailler à temps partiel;
- raison principale pour ne pas avoir travaillé pendant toute l'année;
- appartenance à une organisation métisse;
- langue de travail.

Un total d'environ 85 millions de réponses écrites ont été codées à partir des questionnaires du Recensement de 2021. Dans l'ensemble, environ 88 % de ces réponses ont été codées automatiquement et environ 9 % ont été codées à l'aide de l'apprentissage automatique, mais le taux de codage variait considérablement d'une question à l'autre.

3.6 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 privés, hormis ceux dans les unités de collecte (UC) situées dans les communautés des Premières Nations, les établissements métis, les régions inuites et d'autres régions éloignées, et tous les logements privés annexés à un logement collectif, qu'ils soient occupés ou inoccupés. Ces renseignements ont permis d'apporter des ajustements à la base de données du recensement. L'ECL a été réalisée auprès d'un échantillon aléatoire de 1 903 UC visées par les méthodes d'envoi par la poste, de listage/livraison et d'envoi par la poste et livraison à la porte. Les agents recenseurs ont de nouveau visité ces UC en juin, juillet et août 2021 afin de réévaluer l'occupation/inoccupation le jour du recensement de tous les logements privés pour lesquels aucune réponse n'avait été reçue. L'ECL a montré que 17,3 % des 1 259 149 logements privés classés comme étant inoccupés étaient en fait des logements occupés, et que 38,5 % des 342 162 logements privés sans réponse qui avaient été classés comme étant occupés ou dont l'occupation/inoccupation était inconnue étaient en fait des logements inoccupés. Les estimations fondées sur l'échantillon de l'ECL ont servi à ajuster l'occupation/inoccupation des logements individuels. Cela s'est traduit par une augmentation de 3,0 % du nombre de logements privés occupés, et par une diminution de 6,8 % du nombre de logements inoccupés à l'échelle du Canada.

Le statut de non-réponse final est déterminé après l'ajustement du statut d'occupation par l'ECL. Les logements privés occupés avec non-réponse ont vu la taille de leur ménage imputée en fonction de la distribution estimée résultant de l'ECL et le reste de leurs données ont fait l'objet d'une imputation. Les réponses imputées provenaient d'un autre ménage répondant du recensement ou de données administratives et étaient généralement le voisin le plus rapproché géographiquement faisant partie d'un ménage de même taille. Ce processus, appelé imputation des ménages entiers, est expliqué aux sections [3.7](#) et [3.8](#).

Le processus d'imputation des ménages entiers comporte un autre élément, outre l'utilisation des estimations de l'ECL, pour ajuster la base de données du recensement. Les aires non couvertes par l'ECL — les UC situées dans les communautés des Premières Nations, les établissements métis, les régions inuites et d'autres régions éloignées, et tous les logements privés annexés à un logement collectif — nécessitent une autre stratégie d'imputation. Dans ces aires seulement, tous les logements privés inoccupés sont traités comme étant réellement inoccupés. En partant du principe que les logements inoccupés sont classés correctement, aucune imputation n'est effectuée. Tous les logements privés non répondants que les agents recenseurs ont classés comme étant occupés ont été traités comme s'ils étaient occupés et ont été imputés comme logements occupés. Comme dans les aires couvertes par l'ECL, pour les logements imputés comme logements occupés, la taille de leur ménage et leurs réponses ont été imputées. Les réponses imputées provenaient d'un autre ménage répondant au recensement ou des données administratives. Aucune restriction n'a été imposée à la taille du ménage pour ces imputations, contrairement aux aires couvertes par l'ECL.

Le résultat du processus d'imputation des ménages entiers est que tous les logements privés sont classés comme occupés ou comme inoccupés (c.-à-d. qu'il n'y a plus de logements non répondants). À l'échelle du Canada (tant pour les aires couvertes par l'ECL que celles non couvertes par l'ECL), 3,1 % des logements privés occupés ont été imputés par l'entremise du processus d'imputation des ménages entiers.

Des renseignements plus détaillés sur l'ECL et le processus d'imputation des ménages entiers seront disponibles dans le [Rapport technique sur la couverture, Recensement de la population, 2021](#), produit n° 98-303-X au catalogue de Statistique Canada.

3.7 Utilisation des données administratives

Le recours aux données administratives a augmenté pour le Recensement de 2021 comparativement au Recensement de 2016. En plus des données administratives utilisées pour le processus du revenu, elles ont été utilisées pour l'immigration, ainsi que dans le cadre du processus d'imputation des ménages entiers. Toutes ces utilisations ont bénéficié du couplage des données administratives.

Revenu

Comme c'était le cas en 2016, les données administratives ont constitué l'unique source de renseignements sur le revenu pour le Programme du recensement. Cela a non seulement allégé le fardeau de réponse, mais a aussi augmenté la qualité et la quantité des données disponibles sur le revenu. Les renseignements sur le revenu des particuliers ont été compilés à partir de données administratives pour l'ensemble de la population âgée de 15 ans et plus. Le fichier des déclarations de revenus et de prestations T1; les feuillets T3, T4, T4A, T4RIF, T4RSP, T5, T4A(P), T4A(OAS), T4E et T5007; les données de l'Allocation canadienne pour enfants; et les données sur le crédit pour la taxe sur les produits et services/taxe de vente harmonisée sont des exemples de sources de données administratives utilisées. Le revenu imposable et le revenu non imposable qui ont été reçus de façon régulière et récurrente pendant l'année civile 2020³ ont été inclus. Les revenus ponctuels, comme les retraits forfaitaires de régimes enregistrés d'épargne-retraite et d'autres régimes d'épargne, les règlements forfaitaires d'assurance ou de prestations de retraite, les gains ou pertes en capital, les héritages et les gains de loterie ont été exclus.

Immigration

Le processus d'immigration a succédé au processus de catégorie d'admission de 2016⁴, qui intègre également des éléments qui figuraient dans le processus ethnoculturel de 2016⁵. Pour la première fois, en 2021, les données administratives d'Immigration, Réfugiés et Citoyenneté Canada (IRCC) ont été la principale source de renseignements pour la plupart des variables traitées dans le processus d'immigration pour l'échantillon du questionnaire détaillé du recensement. En 2016, on a demandé aux répondants d'indiquer leur lieu de naissance, leur citoyenneté, leur statut d'immigrant et leur année d'immigration (le cas échéant). Pour 2021, les questions sur le statut d'immigrant et l'année d'immigration ont été remplacées par des données administratives. En plus des variables traitées en 2016, les données administratives d'IRCC ont fourni de nouvelles variables avec des renseignements sur les résidents non permanents, l'année d'arrivée, la province ou le territoire de destination et plus encore.

Imputation des ménages entiers

Au cours du processus d'imputation des ménages entiers, des données administratives à l'échelle des ménages et des particuliers ont été utilisées pour imputer certains ménages non répondants afin d'améliorer la qualité des données sur les chiffres de population et des logements. Lorsque les données administratives étaient de qualité suffisante, elles étaient utilisées pour imputer les variables de la taille du ménage, de la date de naissance et du sexe à la naissance.

3. L'année de référence pour le revenu est l'année civile 2020, sauf indication contraire. Le Recensement de 2021 est le premier recensement à avoir lié plus d'une année de données sur le revenu. Prenez note que le processus de pondération repose uniquement sur les données de revenu de 2020.

4. En 2016, on a demandé à Statistique Canada d'ajouter des variables de la catégorie d'admission au recensement. Les données ont été obtenues au moyen du couplage d'enregistrements du Fichier d'établissement des immigrants fourni par Immigration, Réfugiés et Citoyenneté Canada (IRCC) aux données du Recensement de 2016.

5. Le processus ethnoculturel comprenait cinq sous-thèmes en 2016 : lieu de naissance, citoyenneté et immigration, lieu de naissance des parents, ascendance autochtone, origine ethnique et minorité visible.

3.8 Vérification et imputation

Les données recueillies dans le cadre de tout recensement ou enquête contiennent des omissions ou des incohérences. Par exemple, un répondant peut éprouver de la réticence à répondre à une question, fournir une réponse qui contredit la réponse fournie à une question précédente ou fournir une réponse vide de sens. D'autres erreurs, comme le codage incorrect, peuvent aussi se produire.

L'épuration définitive des données, qui est effectuée à l'étape de la vérification et de l'imputation, a été entièrement automatisée en utilisant le Système canadien de vérification et d'imputation du recensement (SCANVIR) (Statistique Canada, 2020), pour l'ensemble des sujets du recensement. Deux méthodes d'imputation ont été appliquées. La première méthode, appelée « imputation déterministe », consistait à attribuer des valeurs précises sous certaines conditions quand la résolution du problème est claire et non ambiguë. Des règles de vérification détaillées ont été appliquées pour déterminer ces conditions, puis des valeurs prédéterminées sont attribuées aux variables intervenant dans les règles. La deuxième méthode, appelée « imputation par donneur à changements minimaux par le plus proche voisin », comprenait l'application d'une série de règles de vérification détaillées qui permettaient de déceler toute réponse manquante ou incohérente. Quand un enregistrement qui contenait des réponses manquantes ou incohérentes était repéré, un autre enregistrement qui répondait aux règles de vérification et dont un ensemble de caractéristiques définies étaient les plus proches de celles de l'enregistrement erroné était sélectionné comme l'enregistrement donneur. Les données qui provenaient de cet enregistrement donneur étaient empruntées et utilisées pour apporter le nombre minimal de changements aux variables afin de résoudre tous les problèmes de réponses manquantes ou incohérentes.

Le processus de vérification et d'imputation commence par l'application de la procédure d'imputation des ménages entiers aux non-répondants au recensement dans les UC ayant un taux de réponse inférieur à 90 %. Pour les non-répondants qui ont des dossiers de données administratives de bonne qualité, les variables relatives à la taille du ménage, à la date de naissance et au sexe à la naissance sont imputées à partir de leurs données administratives pour tous les membres du ménage dans un premier temps. Le reste des variables manquantes sont imputées aux étapes subséquentes. Les autres non-répondants au recensement sont imputés par la méthode du plus proche voisin géographique parmi l'ensemble des répondants complets ou partiels, ou l'ensemble des non-répondants maintenant imputés par les données administratives. Dans les aires couvertes par l'ECL, le donneur doit avoir la même taille de ménage.

Une fois la procédure d'imputation des ménages entiers terminée, le reste des renseignements manquants ou invalides sont imputés de manière déterministe ou au moyen de la méthode du plus proche voisin, module par module. Ces modules sont conçus pour traiter ensemble toutes les variables comportant un sujet commun.

3.9 Non-réponse

Le statut de non-réponse peut différer à l'étape de la collecte et à celle du traitement. Les différences principales surviennent parce que l'occupation/inoccupation peut changer entre la collecte et le traitement, et parce que le ménage doit répondre à un minimum de questions pour qu'il soit considéré comme répondant à l'étape du traitement. À moins d'avis contraire, le terme « non-réponse » désigne la non-réponse à l'étape du traitement des données. Il en sera de même si l'on fait référence à la réponse au lieu de la non-réponse.

Pour le questionnaire détaillé du Recensement de 2021, il y a deux genres de ménages qui sont considérés comme des ménages non répondants :

- les ménages de l'échantillon qui n'ont répondu qu'aux questions communes aux deux types de questionnaires, c.-à-d. uniquement aux questions du questionnaire abrégé;
- les ménages qui n'ont répondu à aucune question.

On parle ici de non-réponse totale. Celle-ci est traitée différemment selon le mode de collecte et le type de ménage.

3.10 Pondération

Le Programme du recensement canadien de 2021 était composé d'un Recensement de la population et d'une enquête-échantillon pour laquelle le quart des ménages privés canadiens avaient été sélectionnés. Les ménages non échantillonnés pour l'enquête ont reçu un questionnaire abrégé, tandis que les ménages échantillonnés ont reçu un questionnaire détaillé. En plus des questions du questionnaire abrégé, le questionnaire détaillé a permis de recueillir des renseignements socioculturels ainsi que des informations sur les activités de la vie quotidienne, sur la mobilité, sur le lieu de naissance, sur la scolarité et sur les activités sur le marché du travail, entre autres. On a fait appel à la pondération pour représenter l'ensemble de la population à partir des renseignements recueillis auprès de l'échantillon.

La première étape de la pondération consistait à attribuer à chaque ménage un poids de sondage qui reflétait la probabilité qu'il soit échantillonné. Dans la plupart des UC, la fraction de sondage était d'un quart. Par conséquent, les ménages de ces UC ont reçu un poids de sondage de 4. Par la suite, les poids de sondage de ces UC ont subi un premier ajustement pour tenir compte de la couverture et de la non-réponse totale. Cet ajustement a été appliqué aux poids des ménages répondants. Finalement, un deuxième ajustement, que l'on appelle le calage final, a été effectué afin d'établir une meilleure concordance entre les estimations obtenues à partir des ménages répondants de l'échantillon et les chiffres du recensement pour un certain nombre de caractéristiques provenant du questionnaire abrégé ou de sources de données administratives. La méthodologie de pondération est décrite en détail au [chapitre 4](#). Tous les ménages privés annexés à un logement collectif et tous les ménages privés dans les UC situées dans les communautés des Premières Nations, les établissements métis, les régions inuites et d'autres régions éloignées ont été sélectionnés pour l'échantillon du questionnaire détaillé et ont donc reçu un poids de sondage de 1. Par la suite, ils ont été exclus des processus d'ajustement pour la couverture et la non-réponse ainsi que du calage final.

Les ménages de l'échantillon du questionnaire détaillé ayant un poids non nul à la fin des étapes de pondération étaient les ménages répondants ainsi que les ménages ayant reçu un poids de sondage de 1, c'est-à-dire les ménages privés annexés à un logement collectif et tous les ménages privés dans les UC situées dans les communautés des Premières Nations, les établissements métis, les régions inuites et d'autres régions éloignées. Ces ménages constituaient l'ensemble des ménages qui contribuaient aux estimations du questionnaire détaillé.

3.11 Taux de réponse finaux

Le [tableau 3.11.1](#) présente les taux de réponse finaux des ménages privés du Recensement de la population de 2021 à l'échelle nationale et pour chaque province et territoire, suivis des taux de réponse non pondérés et pondérés pour l'échantillon du questionnaire détaillé en se basant sur la définition de la non-réponse précisée à la [section 3.9](#).

Le taux de réponse final est calculé par le ratio du numérateur au dénominateur où :

- le numérateur correspond au nombre de logements privés pour lesquels un questionnaire a été rempli⁶;
- le dénominateur correspond au nombre de logements privés classés comme occupés selon la base de données du recensement.

La classification finale d'occupation/inoccupation est fondée sur l'analyse des données recueillies par le personnel sur le terrain, les données fournies par les répondants et les résultats d'une étude de la qualité concernant l'occupation/inoccupation selon l'ECL (voir la [section 3.6](#)). Les taux de réponse du [tableau 3.11.1](#) diffèrent des taux de réponse à la collecte, lesquels avaient été publiés auparavant et dont il est fait mention à la [section 1.5](#), en ce sens qu'ils tiennent compte du traitement des données et de la vérification de l'occupation/inoccupation des logements pour cerner les ménages non répondants. Par conséquent, ces taux de réponse sont considérés comme finaux.

6. Les logements privés annexés à des logements collectifs, qui faisaient partie de l'échantillon du questionnaire détaillé mais n'ont reçu qu'un questionnaire abrégé, sont considérés comme non-répondants pour le calcul des taux de réponse du questionnaire détaillé.

Rapport technique sur l'échantillonnage et la pondération

Les taux de réponse pondérés ont été produits pour l'échantillon du questionnaire détaillé. Ils sont définis comme le ratio du numérateur au dénominateur où :

- le numérateur correspond au nombre pondéré de logements privés pour lesquels un questionnaire a été rempli;
- le dénominateur correspond au nombre pondéré de logements privés classés comme occupés selon la base de données du recensement.

Tableau 3.11.1

Taux de réponse finaux des ménages privés au Recensement de la population de 2021 et à l'échantillon du questionnaire détaillé

Région	Taux de réponse — questionnaire abrégé	Taux de réponse non pondéré — questionnaire détaillé seulement pourcentage	Taux de réponse pondéré — questionnaire détaillé seulement
Canada	96,9	94,9	95,7
Terre-Neuve-et-Labrador	97,0	95,0	95,6
Île-du-Prince-Édouard	97,6	96,5	96,8
Nouvelle-Écosse	97,1	95,6	96,1
Nouveau-Brunswick	96,8	94,8	95,7
Québec	97,1	95,7	96,3
Ontario	97,2	95,8	96,2
Manitoba	96,5	93,1	94,4
Saskatchewan	95,5	91,8	93,5
Alberta	96,5	93,4	94,4
Colombie-Britannique	96,5	94,0	95,1
Yukon	95,7	85,5	89,5
Territoires du Nord-Ouest	91,8	86,2	89,2
Nunavut	79,7	78,1	78,1

Note : Tous les ménages privés et les logements occupés sont inclus dans le calcul de ces taux de réponse, sans exception.

Sources : Statistique Canada, Recensement de la population de 2021 et échantillon du questionnaire détaillé du Recensement de 2021.

4. Estimation à l'aide de l'échantillon du questionnaire détaillé du recensement

Tout processus d'échantillonnage nécessite une méthode d'estimation connexe pour convertir les données-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 et 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 méthode d'estimation génère un ensemble de poids, et le poids de chaque unité d'échantillonnage correspond au nombre d'unités dans la population que représente l'unité d'échantillonnage. Ces poids sont appliqués aux données-échantillon pour produire des millions d'estimations à l'aide de l'échantillon du questionnaire détaillé du recensement. Les estimations sont des mesures agrégées telles que les totaux, les moyennes, les proportions et les médianes, qui sont calculés à partir de l'échantillon pour diverses caractéristiques d'intérêt.

4.1 Éléments à considérer pour choisir une méthode d'estimation

4.1.1 Considérations de nature opérationnelle

Sur le plan mathématique, une méthode d'estimation peut être décrite au moyen d'une formule algébrique, ou 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 ensemble 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'échantillon du questionnaire détaillé du recensement (et la plupart des enquêtes-échantillon) consistait à subdiviser la méthode d'estimation en deux étapes : a) le calcul des poids (appelé procédure de pondération); b) l'utilisation des poids pour produire les estimations, comme l'estimation d'un chiffre particulier de population par totalisation des poids appliqués aux personnes ou aux ménages qui possédaient la caractéristique d'intérêt. La complexité mathématique se situe principalement 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. Étant donné que le poids appliqué à chaque unité d'échantillonnage est le même pour toute totalisation faisant intervenir l'unité en question, la cohérence entre les différentes estimations fondées sur les données-échantillon est assurée.

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 inconnue de la population 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 à ce que les énoncés au sujet des valeurs inconnues de la 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 l'estimation contiendront la vraie valeur de la population pour approximativement 95 % de tous les échantillons possibles. Le [chapitre 7](#) décrit en détail les conditions et les méthodes utilisées pour le calcul des intervalles de confiance des estimations pour l'échantillon du questionnaire détaillé du recensement.

Outre la minimisation de l'erreur-type, un second objectif relatif au choix d'une méthode d'estimation pour l'échantillon du questionnaire détaillé est de garantir, autant que possible, que les estimations de l'échantillon des caractéristiques évaluées au recensement concordent avec les valeurs connues du recensement. Heureusement, ces deux objectifs sont habituellement complémentaires en ce sens que l'erreur d'échantillonnage a tendance à être réduite lorsqu'on veille à ce que les estimations de certaines caractéristiques de base de l'échantillon concordent avec les chiffres de population correspondants. Cependant, bien que cela soit vrai en général, forcer les estimations de l'échantillon du questionnaire détaillé pour des 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. Par exemple, si plusieurs aires de diffusion ne comptent qu'un petit nombre de sujets ayant une certaine caractéristique, par exemple un pays de naissance donné, le fait d'assurer la cohérence entre les estimations de l'échantillon et les chiffres du recensement pour ce pays de naissance ferait augmenter de manière injustifiée l'erreur-type pour le reste des caractéristiques.

Si l'on ne dispose d'aucun renseignement au sujet de la population échantillonnée à part celle recueillie pour les unités d'échantillonnage et en l'absence de non-réponse totale, la méthode d'estimation est limitée à la pondération des unités d'échantillonnage par l'inverse de leurs probabilités de sélection. Par exemple, si une unité avait 1 chance sur 4 d'être sélectionnée, elle recevrait un poids de 4. Lorsque la non-réponse totale est observée, le poids doit également être ajusté en fonction de l'estimation de la probabilité de réponse de l'unité, par exemple. En pratique, on dispose souvent de certains renseignements supplémentaires au sujet de la population (p. ex. sa taille totale et peut-être sa répartition en fonction d'une variable donnée — selon la province ou le territoire). Ce genre de renseignement peut être utilisé pour améliorer la formule d'estimation afin de produire des estimations ayant plus de chances d'être proches de la valeur inconnue de la population. Dans le cas de l'échantillon du questionnaire détaillé du recensement, on dispose d'un grand nombre de données très détaillées sur la population échantillonnée grâce aux données du questionnaire abrégé du recensement, à tous les niveaux géographiques. Dans le cadre des ajustements pour la couverture, la non-réponse et le calage, on peut profiter de cette quantité considérable de renseignements sur la population pour améliorer les estimations tirées de l'échantillon du questionnaire détaillé.

Néanmoins, on ne peut pas faire concorder les estimations de l'échantillon du questionnaire détaillé pour les caractéristiques du recensement avec tous les chiffres du recensement à chaque niveau géographique. On observe des différences entre les estimations de l'échantillon et les chiffres du recensement 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 le plus minime possible des poids afin d'obtenir une concordance parfaite entre les estimations du questionnaire détaillé et les chiffres du recensement pour certaines caractéristiques et certains sous-groupes porte le nom de « calage ».

4.2 Régions de pondération

Les divers ajustements effectués sur les poids de sondage ont été faits indépendamment selon la région de pondération. Les régions géographiques qui ont été utilisées à cette fin étaient les aires de diffusion agrégées (ADA) et les super aires de diffusion agrégées (SADA). L'ADA est une nouvelle région géographique de diffusion créée pour le Recensement de 2016. La SADA a été créée expressément pour les procédures de pondération par l'agrégation d'ADA.

4.2.1 Aires de diffusion agrégées

Au total, pour le Recensement de 2021, le Canada est divisé en 5 433 ADA. Des ménages ont été sélectionnés pour l'échantillon du questionnaire détaillé dans 5 191 ADA. Parmi les 242 ADA où il n'y a pas eu de ménages échantillonnés, 237 étaient formées uniquement de ménages hors du champ de l'enquête. Dans les cinq autres ADA, il n'y avait qu'un petit nombre de ménages dans le champ de l'enquête et aucun d'entre eux n'a été sélectionné.

Les ADA de 2021 ont été créées en apportant des modifications minimales aux ADA de 2016, de façon à tenir compte des changements au niveau des aires de diffusion (AD). L'objectif était d'établir des tendances historiques des ADA. Étant donné que les critères liés à la taille sont les plus pertinents pour le processus de pondération, les critères de délimitation ayant servi à définir les ADA de 2016 sont présentés ci-dessous.

Rapport technique sur l'échantillonnage et la pondération

Les ADA respectent les critères de délimitation suivants :

1. Les ADA couvrent l'ensemble du pays et, dans la mesure du possible, ont une population de 5 000 à 15 000 personnes (selon les chiffres de population du recensement précédent).
2. Les ADA respectent les limites provinciales et territoriales ainsi que les limites des divisions de recensement (DR), des régions métropolitaines de recensement (RMR) et des agglomérations de recensement (AR) qui ont été divisées en secteurs de recensement (SR) au Recensement de 2016.
3. Les ADA sont établies à partir de l'une des trois régions géographiques de diffusion du Recensement de 2016 : les AD, les subdivisions de recensement (SDR) ou les SR :
 - Dans les RMR et les AR qui comptent des SR, les SR adjacents sont regroupés afin de respecter le critère de population de l'ADA.
 - Dans les régions sans SR (soit les régions à l'extérieur des RMR et des plus grandes AR), où les SDR ont une population inférieure à 15 000 personnes, les SDR adjacentes sont regroupées afin de respecter le critère de population de l'ADA.
 - Dans les régions sans SR, où les SDR ont une population supérieure à 15 000 personnes, les AD adjacentes sont regroupées dans ces SDR afin de respecter le critère de population de l'ADA.
4. Chaque SDR formée d'une réserve indienne et d'un petit nombre d'autres régions où l'on a eu recours au recensement par interview constitue des ADA distinctes.

« Pour obtenir de plus amples renseignements sur les [aires de diffusion agrégée](#), consultez le *Dictionnaire, Recensement de la population, 2021*, produit n° 98-301-X au catalogue. »

Le [tableau 4.2.1.1](#) indique dans quelle mesure les ADA qui comptaient des ménages dans l'échantillon du questionnaire détaillé étaient bien ajustées aux SDR. Le premier scénario s'est produit dans la grande majorité des cas, car les ADA ont été formées d'abord et avant tout de sorte à respecter les limites géographiques des SR et des SDR. Le scénario 4 est le seul où les limites géographiques des SDR n'ont pas été respectées. Les SR ne figuraient pas dans le tableau, car ils figuraient dans le premier scénario, sauf un qui figurait dans le scénario 3.

Tableau 4.2.1.1

Nombre de subdivisions de recensement situées dans les limites des aires de diffusion agrégées comptant des ménages dans l'échantillon du questionnaire détaillé, Recensement de 2021

Scénario	Description	Subdivision de recensement	
		nombre	pourcentage
1	La SDR était de taille suffisamment petite pour être contenue entièrement dans une ADA, et cette même ADA ne comprenait que des SDR entières. Aucune SDR dans l'ADA ne faisait aussi partie d'une autre ADA.	4 526	93,26
2	La SDR était suffisamment petite pour être contenue entièrement dans une ADA, mais une autre SDR dans la même ADA faisait aussi partie d'une autre ADA.	39	0,80
3	La SDR était suffisamment grande pour contenir des ADA entières. Aucune des ADA ne faisait partie d'une autre SDR.	262	5,40
4	La SDR faisait partie de deux ADA ou plus.	26	0,54
Total		4 853	100,00

SDR = Subdivision de recensement

ADA = Aire de diffusion agrégée

Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2021.

Le [tableau 4.2.1.2](#) montre la répartition du nombre d'ADA qui comptent des ménages dans l'échantillon du questionnaire détaillé selon la province ou le territoire.

Rapport technique sur l'échantillonnage et la pondération

Tableau 4.2.1.2

Nombre d'aires de diffusion agrégées comptant des ménages dans l'échantillon du questionnaire détaillé, selon la province ou le territoire

Région	Nombre d'ADA
Terre-Neuve-et-Labrador	83
Île-du-Prince-Édouard	23
Nouvelle-Écosse	148
Nouveau-Brunswick	129
Québec	1 144
Ontario	1 659
Manitoba	222
Saskatchewan	263
Alberta	515
Colombie-Britannique	912
Yukon	29
Territoires du Nord-Ouest	38
Nunavut	26
Canada	5 191

ADA = Aire de diffusion agrégée

Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2021.

Le [tableau 4.2.1.3](#) montre le nombre d'ADA selon le nombre de ménages du recensement dans le champ de l'enquête. La majorité des ADA qui comptaient des ménages dans l'échantillon du questionnaire détaillé comptaient de 2 000 à 4 999 ménages. Un nombre important d'ADA étaient peu peuplées.

Tableau 4.2.1.3

Répartition des aires de diffusion agrégées comptant des ménages dans l'échantillon du questionnaire détaillé, selon le nombre de ménages dans le champ de l'enquête

Nombre de ménages dans le champ de l'enquête	Nombre d'ADA	Pourcentage
0 à 499	996	19,19
500 à 999	118	2,27
1 000 à 1 999	359	6,92
2 000 à 2 999	1 190	22,92
3 000 à 3 999	1 189	22,91
4 000 à 4 999	733	14,12
5 000 à 5 999	356	6,86
6 000 à 6 999	143	2,75
7 000 à 7 999	46	0,89
8 000 à 8 999	29	0,56
9 000 à 9 999	13	0,25
10 000 et plus	19	0,37
Total	5 191	100,00

ADA = Aire de diffusion agrégée

Source : Statistique Canada, Recensement de la population de 2021.

Le [tableau 4.2.1.4](#) présente le nombre d'ADA selon les intervalles du nombre de ménages répondants au questionnaire détaillé du Recensement de 2021. Pour les ADA qui comptaient le moins de répondants, un traitement particulier a été appliqué afin d'avoir suffisamment de ménages pour la pondération (voir la [section 4.5](#)).

Tableau 4.2.1.4

Répartition des aires de diffusion agrégées comptant des ménages dans l'échantillon du questionnaire détaillé, selon le nombre de ménages répondant au questionnaire détaillé

Nombre de répondants	Nombre d'ADA	Pourcentage
0 à 99	690	13,29
100 à 199	276	5,32
200 à 299	132	2,54
300 à 399	128	2,47
400 à 499	272	5,24
500 à 599	478	9,21
600 à 699	559	10,77
700 à 799	583	11,23
800 à 899	499	9,61
900 à 999	411	7,92
1 000 à 1 099	322	6,20
1 100 à 1 199	246	4,74
1 200 à 1 299	189	3,64
1 300 à 1 399	128	2,47
1 400 à 1 499	98	1,89
1 500 et plus	180	3,47
Total	5 191	100,00

ADA = Aire de diffusion agrégée

Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2021.

4.2.2 Super aires de diffusion agrégées

Les SADA ont été créées expressément pour la pondération des données du Recensement de 2016 afin de mener certaines procédures de pondération pour lesquelles il est souhaitable d'avoir un grand nombre d'observations.

Les SADA de 2021 ont été créées en apportant des modifications minimales aux SADA de 2016, de façon à tenir compte des changements au niveau des ADA. Étant donné que les critères liés à la taille présentent un intérêt particulier pour le processus de pondération, les critères de délimitation ayant servi à définir les SADA de 2016 sont présentés ci-dessous.

Les SADA ont été établies de sorte à respecter les règles suivantes (en ordre de priorité) :

1. Les SADA sont créées en regroupant des ADA (obligatoire).
2. Les SADA respectent les limites provinciales et territoriales (obligatoire).
3. Les SADA ont une population de 50 000 à 150 000 personnes (sauf pour les DR qui ont une population de 40 000 à 50 000 personnes qui forment leur propre SADA) excluant les personnes vivant dans des UC de recensement par interview.
4. Les SADA respectent les limites des DR.
5. Les SADA respectent les limites des RMR et des AR.
6. Les SADA respectent les limites des SDR.
7. Les SADA sont en un seul morceau, donc contiguës.
8. Les SADA sont aussi compactes que possible.

Les deux premières règles étaient obligatoires et les règles 3 à 9 ont été respectées autant que possible. Au total, 409 SADA ont été formées.

Rapport technique sur l'échantillonnage et la pondération

Le [tableau 4.2.2.1](#) montre la répartition du nombre de SADA selon la province ou le territoire.

Tableau 4.2.2.1
Nombre de super aires de diffusion agrégées, selon la province ou le territoire

Région	Nombre de SADA
Terre-Neuve-et-Labrador	8
Île-du-Prince-Édouard	2
Nouvelle-Écosse	13
Nouveau-Brunswick	8
Québec	97
Ontario	150
Manitoba	15
Saskatchewan	14
Alberta	44
Colombie-Britannique	55
Yukon	1
Territoires du Nord-Ouest	1
Nunavut	1
Total	409

SADA = Super aire de diffusion agrégée

Note : Pour les trois territoires, la SADA correspond au territoire.

Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2021.

Le [tableau 4.2.2.2](#) indique dans quelle mesure les SADA sont bien ajustées aux DR et aux RMR. Les SADA respectent les limites géographiques de la grande majorité des DR (scénarios 1 et 3) et respectent celles des trois quarts des RMR. Les autres RMR font partie d'au moins deux SADA (scénario 4).

Tableau 4.2.2.2
Nombre de divisions de recensement et de régions métropolitaines de recensement situées dans les limites des super aires de diffusion agrégées, Recensement de 2021

Scénario	Description	Divisions de recensement		Régions métropolitaines de recensement	
		nombre	pourcentage	nombre	pourcentage
1	La DR ou la RMR était de taille suffisamment petite pour être contenue entièrement dans une SADA, et la même SADA ne comprenait que des DR ou des RMR entières. Aucune DR ou RMR contenue dans la SADA ne faisait aussi partie d'une autre SADA.	249	84,98	6	14,63
2	La DR ou la RMR était de taille suffisamment petite pour être contenue entièrement dans une SADA, mais une autre DR ou RMR contenue dans la même SADA faisait aussi partie d'une autre SADA.	2	0,68	0	0,00

Rapport technique sur l'échantillonnage et la pondération

Tableau 4.2.2.2

Nombre de divisions de recensement et de régions métropolitaines de recensement situées dans les limites des super aires de diffusion agrégées, Recensement de 2021

Scénario	Description	Divisions de recensement		Régions métropolitaines de recensement	
		nombre	pourcentage	nombre	pourcentage
3	La DR ou la RMR était de taille suffisamment grande pour contenir des SADA entières. Aucune des SADA ne faisait aussi partie d'une autre DR ou RMR.	40	13,65	26	63,41
4	La DR ou la RMR faisait partie de deux SADA ou plus.	2	0,68	9	21,95
Total		293	100,00	41	100,00

DR = Division de recensement

RMR = Région métropolitaine de recensement

SADA = Super aire de diffusion agrégée

Source : Statistique Canada, Recensement de la population de 2021.

Le [tableau 4.2.2.3](#) montre le nombre de SADA, selon le nombre de personnes dans le champ de l'enquête.

Tableau 4.2.2.3

Répartition des super aires de diffusion agrégées comptant des ménages dans l'échantillon du questionnaire détaillé, selon le nombre de personnes dans le champ de l'enquête

Nombre de personnes dans le champ de l'enquête	Nombre de SADA	Pourcentage
30 000 à 39 999	3	0,73
40 000 à 49 999	20	4,89
50 000 à 59 999	23	5,62
60 000 à 69 999	29	7,09
70 000 à 79 999	101	24,69
80 000 à 89 999	66	16,14
90 000 à 99 999	46	11,25
100 000 à 149 999	114	27,87
150 000 et plus	7	1,71
Total	409	100,00

SADA = Super aire de diffusion agrégée

Source : Statistique Canada, Recensement de la population de 2021.

4.3 Poids de sondage

Le poids de sondage de chaque ménage de l'échantillon du questionnaire détaillé a été calculé différemment selon le mode de collecte dans le secteur où se trouvait le logement correspondant.

- Si le mode de collecte était l'envoi par la poste (EP), le listage/livraison (L/L) ou l'envoi par la poste et livraison à la porte (EPLP), le poids de sondage correspondait à l'inverse de la fraction de sondage, c.-à-d. un poids de 4.
- Les ménages situés dans les communautés des Premières Nations, les établissements métis, les régions inuites et d'autres régions éloignées ont reçu un poids de sondage de 1.

Les ménages qui vivent dans des logements privés annexés à un logement collectif faisaient exception. Comme mentionné à la [section 2.2](#), tous ces ménages ont été ajoutés à l'échantillon. Ils étaient considérés comme étant à tirage complet et ont par le fait même reçu un poids de sondage de 1.

4.3.1 Poids des ménages sélectionnés dans l'échantillon

Les ménages échantillonnés qui ont un poids de sondage de 1 n'ont reçu aucun ajustement de poids. Ces ménages ont conservé leur poids de 1 une fois les procédures de pondération terminées (couverture et non-réponse, ainsi que calage aux totaux du recensement). Ils provenaient d'UC de recensement par interview ou étaient des ménages privés annexés à des logements collectifs.

Pour ces ménages, la non-réponse totale et la non-réponse partielle ont été traitées par imputation. Une fois les données manquantes imputées, ces ménages étaient considérés comme répondants aux fins d'estimation (bien qu'ils soient non-répondants pour le calcul des taux de réponse présentés à la [section 3.11](#)).

4.4 Ajustement pour la couverture et pour la non-réponse totale

Bien qu'il existe plusieurs manières de traiter la non-réponse aux enquêtes, elles peuvent être classées en deux catégories principales : l'imputation et la repondération. La première est généralement employée pour traiter les valeurs manquantes isolées, et la deuxième, pour traiter les cas de non-réponse totale. Un ménage était traité comme un ménage répondant au questionnaire détaillé lorsqu'il répondait à au moins une question du questionnaire détaillé. Compte tenu du taux de réponse élevé au questionnaire détaillé, toute méthode d'ajustement aux cas de non-réponse aurait eu un effet généralement modeste sur les estimations et les poids définitifs de l'enquête. La couverture et la non-réponse totale des ménages vivant dans les UC situées dans les communautés des Premières Nations, les établissements métis, les régions inuites et d'autres régions éloignées ont été compensées par des procédures d'imputation, principalement par l'imputation de ménages entiers telle que décrite à la [section 3.6](#). Dans le reste du pays, des procédures de repondération ont été employées. Le reste du présent chapitre décrit ces procédures de pondération.

Les ajustements pour la couverture et pour la non-réponse visent principalement à réduire au minimum l'incidence de tout biais causé par une couverture incomplète (ou des questionnaires reçus en double) et par la non-réponse totale. Pour que l'ajustement réduise réellement le biais potentiel, il est très utile de disposer d'un vaste ensemble de renseignements sur les non-répondants. En l'absence de tels renseignements, l'ajustement pour la non-réponse que l'on peut appliquer sera limité et ne pourra pas grandement réduire le biais potentiel. Seuls les renseignements géographiques de tous les ménages non répondants étaient connus. Les renseignements sur les non-répondants étaient donc plutôt limités. Heureusement, avant que ne soient appliqués les ajustements pour la couverture et pour la non-réponse, le processus d'imputation de ménages entiers a été effectué. Une part importante de l'imputation de ménages entiers consiste à imputer les caractéristiques du questionnaire abrégé pour tous les non-répondants à ce questionnaire. Cela comprenait les non-répondants de l'échantillon du questionnaire détaillé qui n'ont pas répondu aux questions du questionnaire abrégé. Ces renseignements supplémentaires ont servi de base à l'ajustement pour la non-réponse pour l'échantillon du questionnaire détaillé.

Aux fins d'ajustement pour la couverture et pour la non-réponse totale dans l'échantillon du questionnaire détaillé, on a appliqué une procédure de repondération fondée sur le calage aux poids de sondage. Cette procédure peut être divisée en quatre étapes principales :

1. sélection des contraintes de calage pour les étapes 2 et 3;
2. ajustement pour la couverture selon un calage non linéaire;
3. estimation de la propension à la non-réponse en fonction du calage non linéaire de la non-réponse;
4. application d'une méthode de pointage fondée sur la propension à l'étape 3.

Les étapes 1 à 4 ont été appliquées de manière indépendante dans chaque SADA. Autrement dit, l'ajustement pour la non-réponse a été appliqué par SADA. La [section 4.2](#) présente des définitions et des renseignements au sujet des ADA et des SADA.

La première étape consistait en une sélection ascendante des contraintes de calage dans la SADA, qui a été effectuée comme suit :

- L'ensemble de contraintes potentielles a été tiré des variables communes au questionnaire abrégé et au questionnaire détaillé, ainsi que de données administratives obtenues au moyen de stratégies de couplage des enregistrements (au cours desquelles toutes les unités de la population du questionnaire détaillé sont soumises aux procédures de couplage). En raison des exigences de la méthode de calage non linéaire utilisée à la deuxième et à la troisième étape, seules les contraintes au niveau de la SADA, ainsi que le nombre de ménages et de personnes dans chaque ADA de la SADA, ont été prises en considération.
- Dans chaque SADA, deux contraintes obligatoires ont été sélectionnées en premier : le nombre de ménages dans la SADA (TOTHLD) et le nombre de personnes dans la SADA (TOTPERS).
- Les contraintes propres aux ADA — le nombre de ménages (HHADA) et le nombre de personnes (PPADA) — ont été évaluées pour la sélection.
- Toutes les autres contraintes potentielles de la SADA ont été évaluées, en donnant priorité à celles qui divisaient le plus possible la population de la SADA en deux parts égales.

Le processus de sélection excluait les contraintes qui visaient moins de 250 ménages dans la SADA et les contraintes jugées redondantes ou presque redondantes au chapitre de la colinéarité avec ces contraintes ou avec des contraintes déjà sélectionnées. Les contraintes redondantes à des contraintes déjà sélectionnées ont été exclues parce qu'elles n'apportaient aucun nouveau renseignement. Compte tenu de ces filtres, l'ordre de priorité employé dans l'évaluation des contraintes a fait en sorte que les contraintes sélectionnées se complétaient mutuellement et compensaient tout écart potentiel de couverture entre le questionnaire détaillé et le questionnaire abrégé, ainsi que pour la non-réponse totale au recensement.

La deuxième étape consistait à appliquer un ajustement pour la couverture selon un calage non linéaire à l'échantillon entier dans la SADA (c.-à-d. les répondants comme les non-répondants). Pour les contraintes sélectionnées à la première étape, les comptes pondérés de l'échantillon du questionnaire détaillé ont été harmonisés avec les chiffres de population correspondants. Cette étape avait pour but de corriger tout écart potentiel de la couverture entre l'échantillon du questionnaire détaillé et son complément (c.-à-d. l'ensemble des ménages recevant uniquement le questionnaire abrégé). Le surdénombrement peut se produire lorsque des personnes sont comptées dans deux ménages différents. La couverture pour les deux populations pourrait également être différente si, par exemple, les logements occupés étaient plus susceptibles d'être incorrectement répertoriés comme étant inoccupés selon le questionnaire détaillé que selon le questionnaire abrégé. Un autre objectif de cette étape consistait à isoler autant que possible l'erreur d'échantillonnage. Sans cette étape, le calage de la non-réponse à l'étape suivante confondrait l'erreur due à la non-réponse avec l'erreur d'échantillonnage. Cette étape fait concorder les estimations de l'échantillon avec les estimations de population. En outre, les mêmes totaux de contrôle sont utilisés dans les deux procédures de calage. Par conséquent, l'estimation de la propension à la non-réponse effectuée à l'étape suivante n'a pas besoin de corriger (directement ou indirectement) l'erreur d'échantillonnage. La combinaison d'une correction pour l'erreur d'échantillonnage et pour l'erreur due à la non-réponse à l'étape suivante aurait été inappropriée. La procédure de calage aurait échoué si le poids de tout répondant avait dû être réduit en vue de concorder avec les chiffres du recensement, car la propension estimée aurait alors dépassé 1. De plus, la méthode de pointage appliquée à la dernière étape nécessitait une estimation de la propension de réponse toute seule. Dans la mesure où la variable d'intérêt était liée aux contraintes sélectionnées, la variance d'échantillonnage a également été réduite lors de cette étape.

Après ces deux étapes, le principal ajustement pour la non-réponse a été effectué. Le poids des non-répondants, ajusté à l'étape précédente, a été mis à 0 tandis que le poids des répondants a été augmenté pour faire en sorte que les sommes pondérées de la SADA concordent avec les chiffres de population correspondants pour les contraintes sélectionnées. Une fonction de lien logistique entre la propension de réponse et les caractéristiques utilisées lors du calage a permis une estimation implicite de la propension de réponse. Folsom et Singh (2000) ont proposé cette méthode de calage non linéaire comme méthode d'ajustement pour la non-réponse tout en s'assurant que les deux estimations correspondaient aux chiffres de population sélectionnés et que les probabilités de réponse estimées se trouvaient entre 0 et 1. Cette dernière condition ne reste pas nécessairement valable

lorsqu'un calage linéaire est utilisé pour l'ajustement pour la non-réponse. Dans la mesure où la propension de réponse était liée aux contraintes sélectionnées, le biais de non-réponse potentiel a été réduit sans augmenter la variance lors de cette étape.

L'inverse des probabilités de réponse estimées obtenues lors de l'étape précédente pourrait être utilisé directement pour ajuster le poids de la non-réponse. Cependant, la méthode de pointage a été utilisée pour la dernière étape de l'ajustement pour la non-réponse afin de lisser les probabilités estimées de l'étape précédente. Cela a permis d'assurer la qualité de l'ajustement pour la non-réponse et d'éviter les ajustements extrêmes. Pour chaque ADA, des classes de pondération homogènes ont été constituées en fonction des probabilités de réponse estimées. Dans chaque classe, la moyenne harmonique pondérée des probabilités de réponse a été calculée. La moyenne harmonique a été utilisée, car les valeurs aberrantes des probabilités de réponse estimées ont une moins grande incidence sur elle. L'inverse de cette moyenne a été appliqué au poids des répondants de la classe à titre d'ajustement pour la non-réponse. Il s'agit d'une méthode équivalente à l'application de la moyenne arithmétique pondérée des facteurs d'ajustement à l'intérieur de chacune des classes de pondération homogènes, où les facteurs d'ajustement correspondraient à l'inverse des propensions à répondre estimées.

En bref, l'ajustement pour la couverture et pour la non-réponse totale était un produit de deux quantités : l'ajustement pour la couverture et l'inverse de la moyenne harmonique obtenue selon la méthode de pointage.

4.5 Calage final

Le calage final est un calage linéaire et il a été effectué dans le but de réduire au minimum la variabilité d'échantillonnage des estimations dérivées des réponses au questionnaire détaillé tout en assurant la cohérence entre les totaux estimés et les totaux du Recensement de la population. Cette étape de pondération était nécessaire, car il était important d'assurer la cohérence entre les totaux estimés et les totaux du Recensement de la population pour un grand nombre de variables et de régions géographiques, c'est-à-dire de respecter des contraintes de calage.

Seuls les poids des ménages dont le mode de collecte était l'EP, le L/L ou l'EPLP ont été calés, car ces ménages ont été échantillonnés. Les ménages faisant exception à cette règle étaient ceux dans ces secteurs qui vivaient dans des logements privés annexés à un logement collectif. Parce que tous ces ménages font partie de l'échantillon du questionnaire détaillé et que toutes les réponses au questionnaire détaillé de ces ménages ont été imputées, aucun calage n'a été effectué. Les poids finaux de ces ménages équivalaient donc à 1. Les poids produits par le processus de calage étaient les poids finaux qui servaient à calculer les estimations provenant du questionnaire détaillé et ces poids s'appliquaient autant aux ménages qu'aux familles et aux personnes. Autrement dit, toutes les familles et personnes d'un même ménage reçoivent le poids du ménage. Pour ce dernier ajustement, il était important de limiter la variabilité des poids calés afin d'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 de 1 à 20.

Les contraintes de calage ont été définies aux niveaux de la personne, du ménage et de la famille de recensement. De plus, les contraintes peuvent être choisies à deux niveaux géographiques différents, soit au niveau des ADA et au niveau des SADA. Ces deux niveaux permettent de maximiser de façon globale la cohérence entre les totaux estimés et les totaux du Recensement de la population tout en minimisant le nombre de contraintes de calage, ce qui devrait permettre de réduire la variabilité des estimations. L'[annexe C](#) énumère toutes les contraintes qui ont été prises en considération durant le processus de calage au niveau des ADA et au niveau des SADA. Les caractéristiques pour lesquelles des données provenant du recensement, de sources administratives et du questionnaire détaillé étaient disponibles et pour lesquelles on s'est efforcé d'établir une concordance comprennent, entre autres, l'âge, le genre, l'état matrimonial, l'union libre, la taille du ménage, le type de logement, la langue officielle parlée, l'année d'immigration et le lieu de naissance.

Le processus de sélection des contraintes est appliqué simultanément à une SADA et à ses ADA, mais de manière indépendante dans chaque SADA. Par la suite, le calage a été exécuté en utilisant toutes les contraintes sélectionnées. Trois nouvelles contraintes ont été ajoutées au processus de calage de 2021. Il s'agit du nombre de personnes qui vivent dans un appartement dans un immeuble de cinq étages ou plus (APT5PLUS) et de deux contraintes liées au nombre de personnes qui ont immigré de 2016 à 2021 (YRIMD_2016 et YRIMG1_2016).

De plus, les contraintes précédemment basées sur le concept de sexe du Recensement de 2016 sont désormais basées sur la variable sur le genre à deux catégories du Recensement de 2021⁷. Au niveau des SADA, 203 contraintes ont été définies, tandis que 271 contraintes ont été définies au niveau des ADA. Différentes raisons ont justifié le choix du niveau géographique des contraintes de calage. Ce choix a été fait en collaboration avec les spécialistes du sujet. Par exemple, certaines contraintes n'ont été définies qu'au niveau des SADA, car elles n'auraient pas été assez populeuses au niveau des ADA. Pour d'autres contraintes telles que les groupes d'âge, en plus de chercher des contraintes assez populeuses, le choix a été fait de façon à ce que les contraintes ne soient pas trop semblables lorsqu'elles sont évaluées par le processus de sélection.

Pour faciliter le calage des petites ADA, celles-ci ont été regroupées avant d'effectuer la sélection des contraintes de calage afin d'avoir un minimum de 60 ménages répondants au questionnaire détaillé par ADA. Les petites ADA qui appartenaient entièrement à une SDR ont premièrement été regroupées avec d'autres ADA de la même SADA. Par la suite, les petites ADA des DR ont été regroupées avec d'autres ADA de la même SADA. Enfin, les petites ADA restantes ont été regroupées à une ADA appartenant à une SADA voisine. La procédure de groupement des ADA a produit 4 207 groupements d'ADA de 60 ménages répondants ou plus.

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

Contraintes obligatoires : Ces contraintes devaient être utilisées dans le calage parce qu'il devait y avoir concordance entre les chiffres du recensement et les estimations provenant du questionnaire détaillé aux niveaux géographiques qui sont des agrégats habituels des ADA et des SADA (p. ex. Canada, provinces et territoires). Le nombre de personnes et le nombre de ménages dans les ADA et les SADA étaient les deux contraintes obligatoires.

Contraintes à faible réponse : Les contraintes évaluées pour une population de 200 ménages ou moins n'ont pas été utilisées dans le calage parce qu'elles peuvent rendre les estimations de l'enquête instables.

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 choisissant itérativement la contrainte qui divisait le plus également possible la population de la SADA ou de l'ADA en deux. Les contraintes qui étaient trop linéairement dépendantes ont été exclues. De plus, afin d'éviter d'introduire un biais dans les estimations ponctuelles et d'éviter d'augmenter leur variance, le nombre de contraintes sélectionnées a été limité. À la suite d'évaluations, il a donc été déterminé que ce nombre devait être plus petit que la racine carrée du nombre de ménages répondants participant à la contrainte.

Après avoir sélectionné les contraintes de calage à utiliser, une dernière vérification a été faite afin de vérifier que l'ensemble des contraintes choisi au niveau des ADA et au niveau des SADA soit exempt de colinéarité.

Le calage proprement dit a ensuite été exécuté pour l'ensemble final de contraintes de la deuxième étape en modifiant les poids ajustés pour la couverture et la non-réponse aussi peu que possible de manière à ce que les estimations pondérées soient égales aux totaux du recensement pour ces contraintes. Le Système généralisé d'estimation de Statistique Canada a été utilisé pour effectuer le calage.

7. Pour obtenir plus de renseignements sur la variable sur le genre à deux catégories, voir le [Guide de référence sur l'âge, le sexe à la naissance et le genre, Recensement de la population, 2021](#), produit n° 98-500-X2021014 au catalogue.

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. Quelques-unes de ces raisons sont énumérées ci-dessous.

- Contraintes exclues durant la sélection des contraintes : Comme il est décrit plus haut, des contraintes potentielles pouvaient être exclues parce que le nombre de réponses était faible, parce qu'elles étaient linéairement dépendantes d'autres contraintes choisies (ou trop dépendantes) ou parce qu'elles étaient linéairement dépendantes de contraintes pour lesquelles le nombre de réponses était faible (ou trop dépendantes). Cela a donné lieu à des différences entre les chiffres du recensement et les estimations provenant du questionnaire détaillé pour ces variables lorsqu'il ne s'agissait pas de dépendance linéaire parfaite avec des contraintes choisies.
- Sous-région de pondération : L'ADA était la plus petite région de pondération pour laquelle on a essayé d'obtenir une concordance entre les chiffres du recensement et les estimations provenant du questionnaire détaillé. Toute entité plus petite qu'une ADA, comme la plupart des AD, est appelée sous-région de pondération. Des écarts pouvaient exister entre les chiffres du recensement et les estimations provenant du questionnaire détaillé pour ces sous-régions de pondération.

4.6 Détails sur la sélection des contraintes

Au cours du processus de pondération, la sélection de contraintes a été effectuée à deux reprises : d'abord lors de l'ajustement pour la couverture et la non-réponse, abordé à la [section 4.4](#), et lors du calage final, abordé à la [section 4.5](#). Les variables qui constituaient les contraintes étaient essentiellement les mêmes, mais l'inclusion ou l'exclusion de contraintes variaient légèrement entre les deux étapes de pondération afin de mieux atteindre l'objectif de chaque étape. La présente section explique de quelle façon les contraintes ont été choisies durant ces étapes de pondération.

Le processus de sélection des contraintes, pour les deux ajustements, a commencé à partir d'un ensemble de contraintes obligatoires décrites dans les sections qui précèdent, puis a évalué l'ajout de toutes les autres contraintes candidates une par une. L'ordre de l'évaluation des contraintes candidates était le même pour toutes les SADA. Lors de l'ajout d'une contrainte, les critères **Population nulle** et **Faible population** étaient évalués et la contrainte était exclue si l'un ou l'autre des critères n'était pas satisfait. Si une contrainte répondait aux deux critères, le nouvel ensemble de contraintes qui la comprenait était alors évalué pour les critères de **dépendance linéaire**, de **grande colinéarité** et de **redondance explicative**. Si l'un de ces critères n'était pas respecté, la contrainte était rejetée. Sinon, la contrainte était ajoutée à l'ensemble des contraintes incluses et le processus de sélection était itéré jusqu'à la prochaine contrainte candidate de la liste. Le [tableau 4.6.1](#) résume ces cinq critères, s'ils ont été appliqués à chacun des deux processus, de même que les différences dans la paramétrisation des critères entre les deux processus d'ajustement des poids.

Pour chaque processus d'ajustement des poids, la sélection des contraintes a été effectuée indépendamment pour chacune des 408 SADA qui comportaient des ménages échantillonnés dont le poids était ajusté.

La liste des contraintes de même que la fréquence d'inclusion ou d'exclusion de chaque contrainte, et ce pour chacun des deux processus de pondération, peut être consultée dans l'[annexe C](#).

Rapport technique sur l'échantillonnage et la pondération

Tableau 4.6.1

Critères appliqués dans la sélection des contraintes d'ajustement pour la couverture, la non-réponse et le calage final

Critère	Ajustement pour la couverture et la non-réponse	Calage final
<p>Population nulle selon les chiffres du recensement : Si la contrainte ne comportait aucune population dans l'aire de pondération, l'estimation après ajustement doit aussi être égale à 0 pour cette contrainte. Ces contraintes ne sont pas classées comme étant exclues; elles sont plutôt classées comme étant inadmissibles au processus d'ajustement.</p>	<p>Appliqué au niveau de la SADA/ADA.</p>	<p>Appliqué au niveau de la SADA/ADA.</p>
<p>Faible population selon les chiffres du recensement : Si une contrainte concerne moins qu'un certain nombre de ménages dans la population de la région de pondération, on considère que cette contrainte est négligeable et on l'exclut. Le fait d'inclure ce type de contrainte augmenterait exagérément la variance. Toutefois, les contraintes relatives à une faible population peuvent être implicitement calées et, dans ce cas, sont comprises dans le nombre total de contraintes calées.</p>	<p>Appliqué au niveau de la SADA/ADA. Le nombre de ménages dans la population de la région de pondération est supérieur à 0 et inférieur à 250.</p>	<p>Appliqué au niveau de la SADA/ADA. Le nombre de ménages dans la population de la région de pondération est supérieur à 0 et inférieur à 200.</p>
<p>Dépendance linéaire : S'il est possible de calculer la valeur d'une contrainte en combinant les valeurs d'autres contraintes, l'une de ces contraintes n'est pas nécessaire et doit être supprimée pendant le processus d'ajustement en raison de sa dépendance linéaire. Toutefois, les contraintes exclues en raison de leur dépendance linéaire sont implicitement calées. Par conséquent, celles-ci sont comprises dans le nombre total de contraintes calées.</p>	<p>Appliqué au niveau de la SADA. La sélection des contraintes peut être comparée à la sélection des variables explicatives dans une régression linéaire. Par conséquent, le FIV¹ et l'indice de conditionnement² sont utilisés pour détecter une colinéarité élevée.</p>	<p>Appliqué au niveau de la SADA/ADA. Deux vérifications de dépendance sont effectuées pour déterminer les contraintes linéairement dépendantes. La première vérification est effectuée pendant la sélection des contraintes au niveau de la SADA/ADA, et la deuxième vérification comprend l'ensemble des contraintes choisies aux deux niveaux de la hiérarchie géographique (SADA et ADA).</p>

Tableau 4.6.1

Critères appliqués dans la sélection des contraintes d'ajustement pour la couverture, la non-réponse et le calage final

Critère	Ajustement pour la couverture et la non-réponse	Calage final
Grande colinéarité : Si la valeur d'une contrainte peut être pratiquement calculée en la combinant à d'autres valeurs de contrainte, au moins l'une de ces contraintes doit être supprimée du processus d'ajustement. Une contrainte de ce type n'est pas parfaitement calée.	Appliqué au niveau de la SADA. La sélection des contraintes peut être comparée à la sélection des variables explicatives dans une régression linéaire. Par conséquent, le FIV ¹ et l'indice de conditionnement ² sont utilisés pour détecter une colinéarité élevée.	Appliqué au niveau de la SADA/ADA. Deux vérifications de dépendance linéaire sont effectuées pour déterminer les contraintes presque linéairement dépendantes. La première vérification est effectuée pendant la sélection des contraintes au niveau de la SADA et de l'ADA, et la deuxième vérification comprend l'ensemble des contraintes choisies simultanément aux deux niveaux de la hiérarchie (SADA et ADA).
Redondance explicative : Si une contrainte explique la non-réponse (presque) au même titre que d'autres contraintes déjà sélectionnées, la procédure de calage de la non-réponse échouerait. Autrement dit, si une contrainte n'ajoute rien de plus aux explications fournies par les contraintes déjà sélectionnées au sujet du mécanisme de non-réponse, il ne faudrait pas l'inclure.	Appliqué au niveau de la SADA. On applique une procédure séquentielle (une forme de régression logistique) pour tester la convergence de la régression logistique.	S.O.

SADA = Super aire de diffusion agrégée

ADA = Aire de diffusion agrégée

FIV = Facteur d'inflation de la variance

S.O. = Sans objet / non disponible

1. Le FIV quantifie l'accroissement de la variance des coefficients de régression attribuable à la colinéarité.

2. L'indice de conditionnement quantifie le degré selon lequel une matrice est proche de la singularité.

Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2021.

5. Évaluation des procédures de pondération

Comme il est décrit au [chapitre 4](#), la première étape de la pondération de l'échantillon du questionnaire détaillé était l'attribution de poids de sondage aux ménages. L'attribution des poids s'est faite différemment selon le mode de collecte propre au secteur dans lequel se trouvait le ménage. Les ménages privés annexés à un logement collectif ou situés dans des communautés des Premières Nations, des établissements métis, des régions inuites ou d'autres régions éloignées ont un poids de sondage égal à 1. Les poids finaux de ces ménages privés correspondent donc aux poids de sondage et conservent leur valeur initiale de 1. Tous les autres ménages privés ont un poids de sondage supérieur à 1. Tous les résultats présentés dans le reste du présent chapitre ont été calculés pour le sous-ensemble de ménages ayant un poids de sondage supérieur à 1.

En bref, un poids de sondage a été attribué à chaque ménage et celui-ci était déterminé par le plan de sondage du questionnaire détaillé. Certains ajustements ont ensuite été nécessaires pour tenir compte de la couverture et de la non-réponse totale. Un calage non linéaire a été effectué lors de cet ajustement pour estimer les paramètres des modèles de non-réponse. Les poids, après avoir été ajustés pour la couverture et la non-réponse totale, ont par la suite été ajustés durant le processus de calage final afin de produire les poids finaux. Les poids finaux ont permis d'obtenir une concordance généralement meilleure entre les chiffres du recensement et les estimations correspondantes provenant du questionnaire détaillé.

Les prochaines sections examinent la répartition des poids et, pour diverses caractéristiques, les écarts entre les chiffres du recensement et les estimations de l'échantillon.

5.1 Répartition des poids

Le [graphique 5.1.1](#) et les tableaux [5.1.1](#) et [5.1.2](#) montrent la répartition des poids de sondage, des poids ajustés pour la couverture et la non-réponse et des poids ajustés lors du calage final. Les poids sont groupés par intervalles de longueur de 0,5, sauf pour les premier et dernier intervalles. Le graphique montre le pourcentage de fois que les poids se trouvent dans chaque intervalle. Tous les poids de sondage se situaient de 3,75 à 4,25. Cela est dû au plan de sondage utilisé pour tirer l'échantillon du questionnaire détaillé, selon lequel environ 1 ménage sur 4 a reçu un questionnaire détaillé dans la plupart des régions. Il est également possible de constater l'effet qu'ont eu les ajustements pour la couverture et la non-réponse ainsi que pour le calage final sur les poids. Une différence très apparente est présente dans l'intervalle de 3,75 à 4,25. En fait, 55 % des ménages avaient des poids ajustés pour la couverture et la non-réponse situés de 3,75 à 4,25 comparativement à 100 % des poids de sondage se trouvant dans cet intervalle. Après l'ajustement du calage final, 30 % des ménages avaient des poids finaux se situant dans cet intervalle. Les poids finaux étaient répartis plus uniformément entre l'ensemble des catégories comparativement aux poids de sondage et aux poids ajustés pour la couverture et la non-réponse.

Logiquement, le processus d'ajustement pour la non-réponse aurait tendance à accroître les poids pour compenser les unités non répondantes. Cet accroissement s'est produit dans la plupart des cas. Les changements entre les poids de sondage et les poids ajustés pour la couverture et la non-réponse sont présentés dans le [tableau 5.1.1](#). Ce tableau montre que la plupart des unités ayant quitté l'intervalle [3,75; 4,25) se sont déplacées vers les intervalles [4,25; 4,75) ou [4,75; 5,25). Cependant, le processus d'ajustement pour la couverture et la non-réponse a aussi déplacé certains poids de l'intervalle [3,75; 4,25) aux intervalles [1,00; 2,75), [2,75; 3,25) ou [3,25; 3,75) pour certaines unités. La raison principale est que la procédure comprenait un ajustement pour la sur-couverture et la sous-couverture. Lorsque certains groupes de population ont été surdénombrés, les poids ont été réduits dans ces aires.

Un élément important à noter dans le [tableau 5.1.1](#) est le fait que les unités non répondantes avaient à l'origine un poids positif, puisqu'elles avaient été sélectionnées pour l'échantillon. Le processus d'ajustement pour la non-réponse leur a attribué un poids de 0 et a redistribué leur poids original parmi les unités répondantes. Les unités non répondantes correspondent à la ligne « 0 (non-répondants) » dans le [tableau 5.1.1](#) et ont été retirées du [tableau 5.1.2](#), puisqu'elles n'ont pas été utilisées dans le processus de calage. Le [tableau 5.1.2](#) présente les changements entre les poids ajustés pour la couverture et la non-réponse et les poids calés.

Selon le [tableau 5.1.2](#), la plupart des poids n'ont connu qu'une modification mineure lors du processus de calage. En fait, 78,1 % des cas sont demeurés dans le même intervalle ou ne sont montés ou n'ont baissé que d'un seul intervalle. L'intervalle le plus stable était celui allant de 1,00 à 2,75 : en effet, 64 % des ménages dont le poids

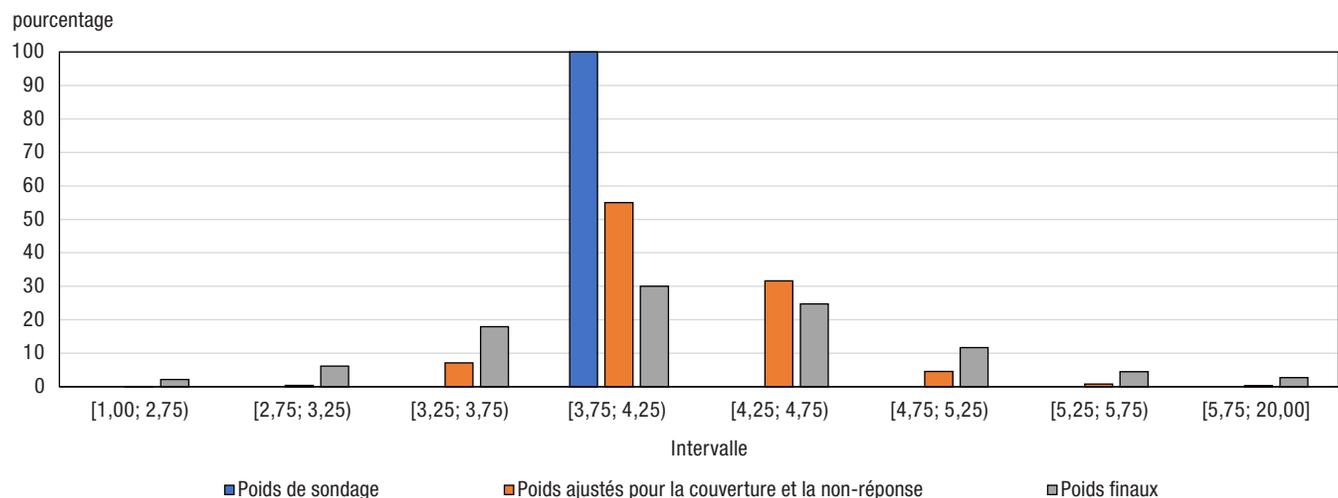
Rapport technique sur l'échantillonnage et la pondération

ajusté pour la couverture et la non-réponse se trouvait dans cet intervalle y sont demeurés après le calage. La deuxième catégorie la plus stable était l'intervalle allant de 5,75 à 12,25 : ici, 61,9 % des ménages dont le poids ajusté pour la couverture et la non-réponse se trouvait dans cet intervalle y sont demeurés (même si l'intervalle de calage va jusqu'à 20,00, et non jusqu'à 12,25).

Enfin, tandis que les poids ajustés pour la couverture et la non-réponse se situaient de 1,00 à 12,25, les poids finaux se situaient de 1,00 à 20,00.

Graphique 5.1.1

Répartition des poids de sondage, des poids ajustés pour la couverture et la non-réponse et des poids finaux



Notes : Tous les ménages ayant un poids de sondage de 1 ont été exclus du processus de pondération. Ces ménages étaient situés dans des communautés des Premières Nations, des établissements métis, des régions inuites ou d'autres régions éloignées, ou étaient des ménages privés annexés à un logement collectif.

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

Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2021.

Tableau 5.1.1

Répartition des poids de sondage et des poids ajustés pour la couverture et la non-réponse

Poids ajustés pour la couverture et la non-réponse	Poids de sondage	
	[3,75; 4,25)	Total
0 (non-répondants)	149 945	149 945
[1,00; 2,75)	1 555	1 555
[2,75; 3,25)	14 575	14 575
[3,25; 3,75)	252 583	252 583
[3,75; 4,25)	1 949 875	1 949 875
[4,25; 4,75)	1 121 300	1 121 300
[4,75; 5,25)	161 359	161 359
[5,25; 5,75)	29 512	29 512
[5,75; 20,00]	11 441	11 441
Total	3 692 145	3 692 145

Notes : Tous les ménages ayant un poids de sondage de 1 ont été exclus du processus de pondération. Ces ménages étaient situés dans des communautés des Premières Nations, des établissements métis, des régions inuites ou d'autres régions éloignées, ou étaient des ménages privés annexés à un logement collectif.

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

Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2021.

Tableau 5.1.2

Répartition des poids ajustés pour la couverture et la non-réponse et des poids finaux

Poids finaux	Poids ajustés pour la couverture et la non-réponse							Total	
	[1,00; 2,75)	[2,75; 3,25)	[3,25; 3,75)	[3,75; 4,25)	[4,25; 4,75)	[4,75; 5,25)	[5,25; 5,75)		
[1,00; 2,75)	999	4 045	19 737	38 100	11 595	1 706	364	174	76 720
[2,75; 3,25)	350	4 559	47 337	131 632	31 612	2 910	395	115	218 910
[3,25; 3,75)	144	3 710	84 405	426 528	111 790	7 757	909	193	635 436
[3,75; 4,25)	42	1 527	65 228	690 146	286 261	18 837	1 895	394	1 064 330
[4,25; 4,75)	13	514	25 637	451 501	359 080	36 037	3 507	678	876 967
[4,75; 5,25)	6	151	7 430	154 852	205 139	41 044	5 725	1 166	415 513
[5,25; 5,75)	1	45	2 020	42 165	77 392	28 209	6 127	1 642	157 601
[5,75; 20,00]	0	24	789	14 951	38 431	24 859	10 590	7 079	96 723
Total	1 555	14 575	252 583	1 949 875	1 121 300	161 359	29 512	11 441	3 542 200

Notes : Tous les ménages ayant un poids de sondage de 1 ont été exclus du processus de pondération. Ces ménages étaient situés dans des communautés des Premières Nations, des établissements métis, des régions inuites ou d'autres régions éloignées, ou étaient des ménages privés annexés à un logement collectif.

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

Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2021.

5.2 Écarts entre les chiffres du recensement et les estimations pour le questionnaire détaillé, Canada

Le [chapitre 4](#) décrit les méthodes utilisées pour calculer les poids finaux des ménages et la [section 5.1](#) montre certaines relations entre les poids de sondage, les poids ajustés pour la couverture et la non-réponse et les poids finaux. L'ajustement pour la couverture et la non-réponse a réduit les écarts entre les chiffres du recensement et les estimations correspondantes pour le questionnaire détaillé pour les contraintes prises en considération (voir l'[annexe C](#)). À la suite de ces ajustements, le calage a davantage réduit, voire éliminé ces écarts pour certaines variables (contraintes). Cependant, certains écarts persistent, puisque des contraintes sont parfois exclues. La différence relative entre les chiffres du recensement et les estimations pour le questionnaire détaillé, aussi appelée écart, est définie comme suit :

$$\text{Écart} = \frac{(\text{estimation pour le questionnaire détaillé} - \text{chiffre du recensement})}{\text{chiffre du recensement}} \times 100 \%$$

Ce ratio représente le pourcentage associé à la caractéristique qui a été surestimée (valeur positive) ou sous-estimée (valeur négative). Aux fins de comparaison, il est également utile d'examiner les valeurs absolues de l'écart et de la différence, ci-après appelées l'écart absolu et la différence absolue respectivement.

Le [tableau 5.2.1](#) montre les différences à l'échelle du Canada entre les chiffres du Recensement de 2021 et les estimations pour le questionnaire détaillé pour les contraintes prises en considération pour les poids de sondage, les poids ajustés pour la couverture et la non-réponse et les poids finaux.

Le [tableau 5.2.1](#) montre également les écarts pour les estimations en fonction des poids finaux. L'analyse de ces écarts permet de mieux comprendre les différences. Plus de 94 % des cas dans le [tableau 5.2.1](#) avaient un écart de -1 % à 1 %, et plus de 99,5 % avaient un écart de -5 % à 5 %.

Le [graphique 5.2.2](#) montre, pour toutes les contraintes, la différence entre les chiffres du recensement et chacune des trois estimations : différence selon le poids de sondage (en bleu), différence ajustée pour la couverture et la non-réponse (en orange) et différence selon le poids final (en vert). L'axe des X représente la taille de la population de la contrainte, en milliers, et les trois lignes pointillées montrent, pour chaque contrainte :

Rapport technique sur l'échantillonnage et la pondération

- la différence entre la somme des poids de sondage et le chiffre du recensement;
- la différence entre la somme des poids ajustés pour la couverture et la non-réponse et le chiffre du recensement;
- la différence entre la somme des poids finaux et le chiffre du recensement.

Les contraintes sont classées de gauche à droite par ordre croissant de la taille de la population.

Le [graphique 5.2.3](#) montre les écarts en pourcentage entre les chiffres du recensement et les estimations finales de toutes les contraintes selon la taille de la population. Pour les contraintes de taille de population moyenne ou grande, les écarts sont tous faibles. Seules certaines contraintes pour les populations de petite taille ont des écarts relativement importants.

L'observation la plus importante à tirer du [graphique 5.2.2](#) est le fait que l'ajustement pour la couverture et la non-réponse améliore grandement les estimations selon le poids de sondage pour ce qui est de la réduction des écarts. Le graphique ne l'illustre pas de manière apparente, mais c'est l'ajustement pour la couverture qui joue le plus grand rôle. La différence entre les chiffres du recensement et les estimations pour le questionnaire détaillé pour les poids de sondage avait tendance à être (beaucoup) plus grande que la différence entre les chiffres du recensement et les estimations du questionnaire détaillé pour les poids ajustés pour la couverture et la non-réponse. Cette dernière différence, pour sa part, avait tendance à être plus grande que la différence correspondante utilisant les poids finaux. Cela illustre l'importance des processus d'ajustement pour la non-réponse et de calage. Une différence entre le chiffre du recensement et l'estimation pour le questionnaire détaillé pourrait avoir lieu dans une SADA ou une ADA pour une caractéristique si la contrainte est exclue durant le calage. Autrement dit, le processus n'a pas exercé un contrôle sur la contrainte exclue pour une aire donnée. Si la contrainte est exclue dans de nombreuses aires, les différences pourraient s'annuler partiellement, ou elles pourraient s'additionner pour créer une différence plus importante à l'échelle du Canada. Le nombre total de personnes (TOTPERS) et le nombre total de ménages (TOTHLD) étaient les deux seules contraintes obligatoires pour lesquelles la concordance entre les chiffres du recensement et les estimations pour le questionnaire détaillé devait être garantie pour toutes les ADA. Par conséquent, la différence ainsi que l'écart pour les poids finaux pour ces caractéristiques étaient nuls. Cependant, toutes les autres contraintes ont dû être exclues pour certaines aires.

L'[annexe C](#) et le [tableau 5.2.1](#) illustrent que les contraintes qui avaient fréquemment été exclues avaient souvent tendance à présenter des différences ou des écarts élevés. En examinant les contraintes définies uniquement au niveau des SADA, il y avait une relation positive entre le nombre de fois qu'une contrainte a été exclue et la différence absolue entre les chiffres du recensement et les estimations du questionnaire détaillé. La contrainte « Personnes vivant en couple (mariées ou en union libre) » (COUPLE) présentait la plus grande différence absolue, soit 8 642, tout en étant la quatrième contrainte la plus fréquemment exclue, ayant été exclue 627 fois entre l'ajustement pour la couverture et la non-réponse et le calage final. Enfin, parmi les contraintes qui visaient uniquement les SADA et qui présentaient les 10 différences relatives les plus importantes, 7 contraintes faisaient également partie des 10 contraintes ayant été exclues le plus fréquemment.

Parmi toutes les contraintes potentielles, les 10 différences absolues les plus importantes se situaient de 2 466 à 8 642. Cependant, étant donné les chiffres élevés du recensement, les écarts pour ces contraintes étaient faibles (allant de -0,45 % à 0,28 %). Les écarts les plus grands ont été observés pour certaines catégories de lieux de naissance. Comme de nombreuses catégories de lieux de naissance ne représentent pas un nombre élevé au Canada, ces variables ont souvent été exclues durant le calage. Cela a donné lieu à de grandes différences absolues et à des écarts absolus particulièrement grands.

Rapport technique sur l'échantillonnage et la pondération

Tableau 5.2.1

Différences et écarts entre les chiffres du recensement et les estimations du questionnaire détaillé, Canada

Caractéristique	Chiffre du recensement	Poids de sondage		Poids ajustés pour la couverture et la non-réponse		Poids finaux		
		estimation	différence	estimation	différence	estimation	différence	écart (%)
ADULTCF	18 645 248	18 561 008	-84 240	18 646 762	1 514	18 646 559	1 311	0,01
AGE00_14	5 863 246	5 825 360	-37 886	5 862 200	-1 046	5 863 461	215	0,00
AGE14	2 074 620	2 061 284	-13 336	2 075 061	441	2 074 101	-519	-0,03
AGE15_24	4 110 940	4 060 056	-50 884	4 109 533	-1 407	4 111 450	510	0,01
AGE15_29	6 479 017	6 395 416	-83 601	6 475 962	-3 055	6 479 062	45	0,00
AGE19	1 960 356	1 942 036	-18 320	1 965 980	5 624	1 965 564	5 208	0,27
AGE24	2 150 584	2 118 020	-32 564	2 143 553	-7 031	2 145 886	-4 698	-0,22
AGE25_34	4 835 713	4 774 696	-61 017	4 833 889	-1 824	4 835 382	-331	-0,01
AGE29	2 368 077	2 335 360	-32 717	2 366 429	-1 648	2 367 612	-465	-0,02
AGE30_49	9 549 900	9 477 840	-72 060	9 547 899	-2 001	9 550 076	176	0,00
AGE34	2 467 636	2 439 336	-28 300	2 467 460	-176	2 467 770	134	0,01
AGE35_44	4 820 439	4 784 580	-35 859	4 819 469	-970	4 820 173	-266	-0,01
AGE39	2 464 600	2 441 428	-23 172	2 464 015	-585	2 464 706	106	0,00
AGE4	1 783 383	1 769 160	-14 223	1 782 623	-760	1 783 298	-85	0,00
AGE44	2 355 839	2 343 152	-12 687	2 355 454	-385	2 355 468	-371	-0,02
AGE45_54	4 582 978	4 553 408	-29 570	4 581 504	-1 474	4 582 969	-9	0,00
AGE49	2 261 825	2 253 924	-7 901	2 260 970	-855	2 262 132	307	0,01
AGE50_64	7 432 010	7 390 072	-41 938	7 431 294	-716	7 431 838	-172	0,00
AGE54	2 321 153	2 299 484	-21 669	2 320 534	-619	2 320 837	-316	-0,01
AGE55_64	5 110 857	5 090 588	-20 269	5 110 759	-98	5 111 001	144	0,00
AGE59	2 593 703	2 581 512	-12 191	2 593 232	-471	2 593 861	158	0,01
AGE64	2 517 154	2 509 076	-8 078	2 517 527	373	2 517 141	-13	0,00
AGE65PL	6 534 621	6 522 260	-12 361	6 535 110	489	6 534 356	-265	0,00
AGE74	3 958 758	3 953 664	-5 094	3 959 122	364	3 957 893	-865	-0,02
AGE75PL	2 575 863	2 568 596	-7 267	2 575 987	124	2 576 463	600	0,02
AGE9	2 005 243	1 994 916	-10 327	2 004 516	-727	2 006 063	820	0,04
APT5PLUS	1 593 869	1 585 684	-8 185	1 593 702	-167	1 593 648	-221	-0,01
APTLT5	2 728 558	2 713 976	-14 582	2 727 869	-689	2 728 625	67	0,00
CHILD	10 442 414	10 361 584	-80 830	10 442 904	490	10 443 251	837	0,01
CHILDFAM	5 884 937	5 846 660	-38 277	5 883 853	-1 084	5 884 257	-680	-0,01

Rapport technique sur l'échantillonnage et la pondération

Tableau 5.2.1

Différences et écarts entre les chiffres du recensement et les estimations du questionnaire détaillé, Canada

Caractéristique	Chiffre du recensement	Poids de sondage		Poids ajustés pour la couverture et la non-réponse		Poids finaux		
		estimation	différence	estimation	différence	estimation	différence	écart (%)
COMLAWNO_DIV	1 859 786	1 849 532	-10 254	1 859 958	172	1 859 981	195	0,01
COMLAWNO_OTHERS	4 008 592	3 990 936	-17 656	4 008 187	-405	4 008 588	-4	0,00
COMLAWNO_SEP	716 514	711 864	-4 650	716 241	-273	716 757	243	0,03
COMLAWNO_SINGLE	14 508 945	14 350 816	-158 129	14 504 993	-3 952	14 508 422	-523	0,00
COMLAWNO_SINGLE_GE15	8 645 699	8 525 456	-120 243	8 642 793	-2 906	8 644 960	-739	-0,01
COMLAWNO_SINGLE_LT15	5 863 246	5 825 360	-37 886	5 862 200	-1 046	5 863 461	215	0,00
COMLAWNO_WID	1 432 292	1 429 540	-2 752	1 431 989	-303	1 431 849	-443	-0,03
COMLAWYE_MARRIED	17 341 257	17 269 196	-72 061	17 339 283	-1 974	17 341 785	528	0,00
COMLAW_YE	3 836 830	3 812 728	-24 102	3 835 614	-1 216	3 835 160	-1 670	-0,04
COUPLE	17 005 592	16 940 552	-65 040	17 009 739	4 147	17 014 234	8 642	0,05
EMPIN_GT50	10 466 983	10 418 360	-48 623	10 468 477	1 494	10 466 948	-35	0,00
EMPIN_LE50	10 470 465	10 370 160	-100 305	10 466 820	-3 645	10 470 574	109	0,00
EMPIN_P0	14 921 346	14 822 428	-98 918	14 917 167	-4 179	14 921 272	-74	0,00
EMPIN_P0_GE15	9 058 100	8 997 068	-61 032	9 054 967	-3 133	9 057 810	-290	0,00
EMPIN_P0_LT15	5 863 246	5 825 360	-37 886	5 862 200	-1 046	5 863 461	215	0,00
EMPIN_P100	5 232 212	5 212 280	-19 932	5 231 362	-850	5 231 417	-795	-0,02
EMPIN_P25	5 237 240	5 184 568	-52 672	5 233 499	-3 741	5 238 205	965	0,02
EMPIN_P50	5 233 225	5 185 592	-47 633	5 233 321	96	5 232 369	-856	-0,02
EMPIN_P75	5 234 771	5 206 080	-28 691	5 237 115	2 344	5 235 531	760	0,01
EMPIN_SADA_GT50	10 468 377	10 421 676	-46 701	10 468 529	152	10 468 160	-217	0,00
EMPIN_SADA_LE50	10 469 071	10 366 844	-102 227	10 466 767	-2 304	10 469 363	292	0,00
EMPIN_SADA_P0	14 921 346	14 822 428	-98 918	14 917 167	-4 179	14 921 272	-74	0,00
EMPIN_SADA_P0_GE15	9 058 100	8 997 068	-61 032	9 054 967	-3 133	9 057 810	-290	0,00
EMPIN_SADA_P0_LT15	5 863 246	5 825 360	-37 886	5 862 200	-1 046	5 863 461	215	0,00
EMPIN_SADA_P100	5 234 004	5 216 312	-17 692	5 234 300	296	5 234 269	265	0,01
EMPIN_SADA_P25	5 235 162	5 181 240	-53 922	5 231 079	-4 083	5 235 190	28	0,00
EMPIN_SADA_P50	5 233 909	5 185 604	-48 305	5 235 688	1 779	5 234 173	264	0,01
EMPIN_SADA_P75	5 234 373	5 205 364	-29 009	5 234 229	-144	5 233 890	-483	-0,01
FEMALE	18 157 552	18 051 988	-105 564	18 155 596	-1 956	18 157 618	66	0,00
FEMALEGE15	15 303 878	15 217 160	-86 718	15 302 343	-1 535	15 303 972	94	0,00

Rapport technique sur l'échantillonnage et la pondération

Tableau 5.2.1

Différences et écarts entre les chiffres du recensement et les estimations du questionnaire détaillé, Canada

Caractéristique	Chiffre du recensement	Poids de sondage		Poids ajustés pour la couverture et la non-réponse		Poids finaux		écart (%)
		estimation	différence	estimation	différence	estimation	différence	
FEMALELT15	2 853 674	2 834 828	-18 846	2 853 253	-421	2 853 646	-28	0,00
HHADA	14 826 894	14 768 580	-58 314	14 826 894	0	14 826 894	0	0,00
HHADACSD	14 826 894	14 768 580	-58 314	14 826 894	0	14 826 894	0	0,00
HHINC_GT50	7 412 136	7 372 056	-40 080	7 414 714	2 578	7 412 314	178	0,00
HHINC_LE50	7 414 758	7 396 524	-18 234	7 412 180	-2 578	7 414 580	-178	0,00
HHINC_P100	3 704 978	3 678 564	-26 414	3 703 026	-1 952	3 704 942	-36	0,00
HHINC_P25	3 708 673	3 704 396	-4 277	3 708 008	-665	3 708 664	-9	0,00
HHINC_P50	3 706 085	3 692 128	-13 957	3 704 172	-1 913	3 705 915	-170	0,00
HHINC_P75	3 707 158	3 693 492	-13 666	3 711 687	4 529	3 707 372	214	0,01
HHINC_SADA_GT50	7 413 255	7 373 352	-39 903	7 413 008	-247	7 413 093	-162	0,00
HHINC_SADA_LE50	7 413 639	7 395 228	-18 411	7 413 886	247	7 413 801	162	0,00
HHINC_SADA_P100	3 706 522	3 682 152	-24 370	3 706 067	-455	3 706 454	-68	0,00
HHINC_SADA_P25	3 706 961	3 702 248	-4 713	3 707 028	67	3 706 238	-723	-0,02
HHINC_SADA_P50	3 706 678	3 692 980	-13 698	3 706 857	179	3 707 564	886	0,02
HHINC_SADA_P75	3 706 733	3 691 200	-15 533	3 706 941	208	3 706 638	-95	0,00
HHSIZE1	4 356 317	4 363 644	7 327	4 356 599	282	4 354 798	-1 519	-0,03
HHSIZE2	5 086 584	5 065 324	-21 260	5 087 377	793	5 086 686	102	0,00
HHSIZE3	2 172 884	2 160 960	-11 924	2 172 802	-82	2 172 690	-194	-0,01
HHSIZE4	1 983 642	1 970 564	-13 078	1 984 050	408	1 984 337	695	0,04
HHSIZE5	777 142	770 908	-6 234	779 026	1 884	779 459	2 317	0,30
HHSIZEGE5	1 227 467	1 208 088	-19 379	1 226 066	-1 401	1 228 382	915	0,07
HHSIZEGE6	450 325	437 180	-13 145	447 040	-3 285	448 924	-1 401	-0,31
INEFAM	29 985 843	29 806 124	-179 719	29 989 178	3 335	29 991 606	5 763	0,02
IR_LINK_NO	35 456 960	35 217 972	-238 988	35 451 968	-4 992	35 457 267	307	0,00
IR_LINK_YE	401 834	392 976	-8 858	400 496	-1 338	401 527	-307	-0,08
LIM_NO	31 967 342	31 741 028	-226 314	31 962 884	-4 458	31 966 263	-1 079	0,00
LIM_YE	3 891 452	3 869 920	-21 532	3 889 579	-1 873	3 892 531	1 079	0,03
LONEPAR	1 639 656	1 620 456	-19 200	1 637 023	-2 633	1 632 325	-7 331	-0,45
MALE	17 701 242	17 558 960	-142 282	17 696 868	-4 374	17 701 176	-66	0,00
MALEGE15	14 691 670	14 568 428	-123 242	14 687 920	-3 750	14 691 361	-309	0,00

Rapport technique sur l'échantillonnage et la pondération

Tableau 5.2.1

Différences et écarts entre les chiffres du recensement et les estimations du questionnaire détaillé, Canada

Caractéristique	Chiffre du recensement	Poids de sondage		Poids ajustés pour la couverture et la non-réponse		Poids finaux		écart (%)
		estimation	différence	estimation	différence	estimation	différence	
MALELT15	3 009 572	2 990 532	-19 040	3 008 948	-624	3 009 815	243	0,01
MARRIED	13 504 427	13 456 468	-47 959	13 503 668	-759	13 506 625	2 198	0,02
NB_NOTINCF	6 771 132	6 688 356	-82 776	6 762 799	-8 333	6 768 984	-2 148	-0,03
NOCLDFAM	4 257 515	4 244 072	-13 443	4 258 039	524	4 255 185	-2 330	-0,05
NOINEFAM	5 872 951	5 804 824	-68 127	5 863 285	-9 666	5 867 188	-5 763	-0,10
NOINEFAMHHSIZEEQ1	4 356 317	4 363 644	7 327	4 356 599	282	4 354 798	-1 519	-0,03
NOINEFAMHHSIZEGT1	1 516 634	1 441 180	-75 454	1 506 687	-9 947	1 512 390	-4 244	-0,28
NOTINFAM	6 771 132	6 688 356	-82 776	6 762 799	-8 333	6 768 984	-2 148	-0,03
NOTINFAMHHSIZEEQ1	4 356 317	4 363 644	7 327	4 356 599	282	4 354 798	-1 519	-0,03
NOTINFAMHHSIZEGT1	2 414 815	2 324 712	-90 103	2 406 200	-8 615	2 414 186	-629	-0,03
OLN_BI	6 523 298	6 482 368	-40 930	6 523 158	-140	6 524 917	1 619	0,02
OLN_EN	24 660 168	24 472 152	-188 016	24 656 197	-3 971	24 658 926	-1 242	-0,01
OLN_FR	3 996 708	3 982 912	-13 796	3 996 280	-428	3 996 783	75	0,00
OLN_NO	678 620	673 516	-5 104	676 829	-1 791	678 168	-452	-0,07
OTHERDTYPE	2 742 490	2 728 540	-13 950	2 742 407	-83	2 742 753	263	0,01
POBG2_1	18 188	18 436	248	18 747	559	18 398	210	1,15
POBG2_10	69 328	68 104	-1 224	68 893	-435	69 005	-323	-0,47
POBG2_11	100 824	100 972	148	100 654	-170	100 771	-53	-0,05
POBG2_16	39 401	39 044	-357	39 043	-358	39 386	-15	-0,04
POBG2_17	37 600	36 964	-636	36 661	-939	37 076	-524	-1,39
POBG2_18	4 705	4 568	-137	4 571	-134	4 660	-45	-0,96
POBG2_19	26 175	25 972	-203	26 157	-18	26 087	-88	-0,34
POBG2_20	27 622	27 504	-118	27 542	-80	27 646	24	0,09
POBG2_21	973 117	971 244	-1 873	974 259	1 142	973 913	796	0,08
POBG2_22	138 803	138 148	-655	140 247	1 444	140 100	1 297	0,93
POBG2_24	45 496	44 620	-876	44 556	-940	44 886	-610	-1,34
POBG2_25	24 018	23 724	-294	23 849	-169	23 699	-319	-1,33
POBG2_26	90 748	90 332	-416	91 084	336	91 009	261	0,29
POBG2_27	70 378	69 804	-574	70 429	51	70 554	176	0,25
POBG2_28	159 482	157 392	-2 090	158 758	-724	159 151	-331	-0,21

Rapport technique sur l'échantillonnage et la pondération

Tableau 5.2.1

Différences et écarts entre les chiffres du recensement et les estimations du questionnaire détaillé, Canada

Caractéristique	Chiffre du recensement	Poids de sondage		Poids ajustés pour la couverture et la non-réponse		Poids finaux		écart (%)
		estimation	différence	estimation	différence	estimation	différence	
POBG2_29	87 715	88 556	841	88 327	612	88 550	835	0,95
POBG2_3	39 318	39 292	-26	39 516	198	39 455	137	0,35
POBG2_30	43 855	43 428	-427	44 366	511	44 233	378	0,86
POBG2_31	145 516	145 352	-164	146 258	742	145 899	383	0,26
POBG2_32	208 157	207 616	-541	208 788	631	208 633	476	0,23
POBG2_33	133 732	133 192	-540	133 176	-556	133 053	-679	-0,51
POBG2_34	24 611	24 160	-451	24 537	-74	24 468	-143	-0,58
POBG2_35	75 006	74 360	-646	75 190	184	75 623	617	0,82
POBG2_36	82 061	80 744	-1 317	81 632	-429	82 056	-5	-0,01
POBG2_37	135 441	133 692	-1 749	134 832	-609	135 038	-403	-0,30
POBG2_38	97 604	97 816	212	98 419	815	98 205	601	0,62
POBG2_39	33 558	33 732	174	33 517	-41	33 587	29	0,09
POBG2_4	69 501	69 680	179	69 323	-178	69 454	-47	-0,07
POBG2_40	147 988	148 988	1 000	147 968	-20	147 920	-68	-0,05
POBG2_41	2 707	2 676	-31	2 736	29	2 713	6	0,21
POBG2_42	207 251	203 500	-3 751	206 111	-1 140	207 270	19	0,01
POBG2_43	759 124	760 536	1 412	761 959	2 835	761 199	2 075	0,27
POBG2_45	93 239	92 184	-1 055	93 552	313	93 699	460	0,49
POBG2_46	73 961	73 020	-941	73 263	-698	73 343	-618	-0,84
POBG2_47	182 036	179 428	-2 608	180 255	-1 781	180 423	-1 613	-0,89
POBG2_48	453 197	451 084	-2 113	454 014	817	453 697	500	0,11
POBG2_50	1 117	1 104	-13	1 121	4	1 091	-26	-2,29
POBG2_51	20 746	20 264	-482	20 457	-289	20 452	-294	-1,42
POBG2_54	114 417	114 916	499	114 900	483	114 799	382	0,33
POBG2_55	21 665	21 680	15	22 065	400	21 929	264	1,22
POBG2_56	87 057	86 636	-421	87 125	68	87 376	319	0,37
POBG2_57	31 161	30 388	-773	30 862	-299	30 901	-260	-0,84
POBG2_59	50 363	51 048	685	51 538	1 175	51 260	897	1,78
POBG2_6	38 714	38 148	-566	38 503	-211	38 242	-472	-1,22
POBG2_60	129 348	128 936	-412	129 813	465	129 745	397	0,31

Rapport technique sur l'échantillonnage et la pondération

Tableau 5.2.1

Différences et écarts entre les chiffres du recensement et les estimations du questionnaire détaillé, Canada

Caractéristique	Chiffre du recensement	Poids de sondage		Poids ajustés pour la couverture et la non-réponse		Poids finaux		écart (%)
		estimation	différence	estimation	différence	estimation	différence	
POBG2_63	12 109	12 044	-65	11 964	-145	12 044	-65	-0,54
POBG2_64	1 282 470	1 274 528	-7 942	1 284 077	1 607	1 281 244	-1 226	-0,10
POBG2_65	33 796	33 840	44	33 947	151	33 820	24	0,07
POBG2_66	340 472	340 244	-228	340 373	-99	340 034	-438	-0,13
POBG2_67	225 939	224 856	-1 083	225 835	-104	226 274	335	0,15
POBG2_68	30 781	29 976	-805	30 454	-327	30 502	-279	-0,91
POBG2_69	132 839	131 732	-1 107	133 079	240	132 934	95	0,07
POBG2_7	77 130	76 736	-394	76 908	-222	76 934	-196	-0,25
POBG2_70	62 339	61 204	-1 135	61 677	-662	61 229	-1 110	-1,78
POBG2_71	108 119	107 468	-651	108 495	376	108 324	205	0,19
POBG2_9	1 800	1 896	96	1 906	106	1 942	142	7,91
POBG3_10	356 921	356 844	-77	356 937	16	356 600	-321	-0,09
POBG3_12	60 983	60 972	-11	61 581	598	61 384	401	0,66
POBG3_14	372 701	368 480	-4 221	373 187	486	373 777	1 076	0,29
POBG3_15	991 957	992 892	935	995 727	3 770	994 720	2 763	0,28
POBG3_16	52 163	52 944	781	53 444	1 281	53 202	1 039	1,99
POBG3_17	1 436 629	1 427 828	-8 801	1 438 445	1 816	1 435 473	-1 156	-0,08
POBG3_18	333 700	330 792	-2 908	333 977	277	334 217	517	0,15
POBG3_19	238 714	236 492	-2 222	238 111	-603	238 401	-313	-0,13
POBG3_2	371 246	369 500	-1 746	371 278	32	371 108	-138	-0,04
POBG3_20	225 939	224 856	-1 083	225 835	-104	226 274	335	0,15
POBG3_21	197 885	195 612	-2 273	197 493	-392	196 876	-1 009	-0,51
POBG3_22	269 123	268 664	-459	269 484	361	268 968	-155	-0,06
POBG3_3	81 706	80 576	-1 130	80 276	-1 430	81 123	-583	-0,71
POBG3_4	188 533	187 180	-1 353	188 954	421	188 989	456	0,24
POBG3_5	230 116	228 392	-1 724	230 849	733	230 586	470	0,20
POBG3_6	1 156 486	1 155 964	-522	1 157 691	1 205	1 157 397	911	0,08
POBG3_7	260 737	260 508	-229	260 109	-628	260 456	-281	-0,11
POBG3_8	709 194	703 532	-5 662	707 532	-1 662	707 463	-1 731	-0,24
POBG3_9	294 966	292 056	-2 910	294 438	-528	295 820	854	0,29

Rapport technique sur l'échantillonnage et la pondération

Tableau 5.2.1

Différences et écarts entre les chiffres du recensement et les estimations du questionnaire détaillé, Canada

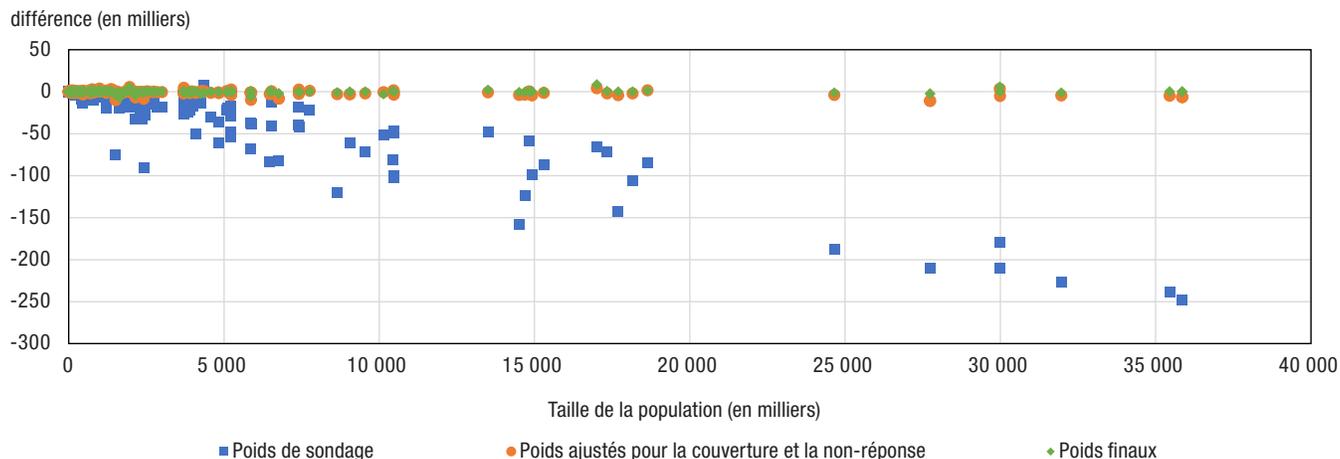
Caractéristique	Chiffre du recensement	Poids de sondage		Poids ajustés pour la couverture et la non-réponse		Poids finaux		
		estimation	différence	estimation	différence	estimation	différence	écart (%)
PPADA	35 858 794	35 610 948	-247 846	35 852 464	-6 330	35 858 794	0	0,00
PPADACSD	35 858 794	35 610 948	-247 846	35 852 464	-6 330	35 858 794	0	0,00
SNGLDET	7 761 977	7 740 380	-21 597	7 762 916	939	7 761 868	-109	0,00
TOTCFAM	10 142 452	10 090 732	-51 720	10 141 892	-560	10 139 442	-3 010	-0,03
TOTHHLD	14 826 894	14 768 580	-58 314	14 826 894	0	14 826 894	0	0,00
TOTPERS	35 858 794	35 610 948	-247 846	35 852 464	-6 330	35 858 794	0	0,00
TPERGE15	29 995 548	29 785 588	-209 960	29 990 263	-5 285	29 995 333	-215	0,00
TPERLT15	5 863 246	5 825 360	-37 886	5 862 200	-1 046	5 863 461	215	0,00
YRIMD_1900	1 057 784	1 051 708	-6 076	1 057 998	214	1 057 899	115	0,01
YRIMD_1980	336 657	335 244	-1 413	335 330	-1 327	336 180	-477	-0,14
YRIMD_1986	478 539	477 856	-683	479 476	937	479 823	1 284	0,27
YRIMD_1991	704 709	705 148	439	706 313	1 604	705 588	879	0,12
YRIMD_1996	664 518	662 160	-2 358	665 920	1 402	665 189	671	0,10
YRIMD_2001	834 315	830 188	-4 127	833 798	-517	833 822	-493	-0,06
YRIMD_2006	938 128	935 132	-2 996	938 038	-90	938 252	124	0,01
YRIMD_2011	1 045 048	1 038 256	-6 792	1 045 197	149	1 044 095	-953	-0,09
YRIMD_2016	1 265 550	1 261 424	-4 126	1 265 880	330	1 265 578	28	0,00
YRIMD_M3	27 740 538	27 530 424	-210 114	27 729 416	-11 122	27 738 072	-2 466	-0,01
YRIMD_M5	793 008	783 408	-9 600	795 099	2 091	794 295	1 287	0,16
YRIMG1_1900	1 057 784	1 051 708	-6 076	1 057 998	214	1 057 899	115	0,01
YRIMG1_1980	815 196	813 100	-2 096	814 806	-390	816 003	807	0,10
YRIMG1_1991	1 369 227	1 367 308	-1 919	1 372 232	3 005	1 370 777	1 550	0,11
YRIMG1_2001	1 772 443	1 765 320	-7 123	1 771 836	-607	1 772 075	-368	-0,02
YRIMG1_2011	1 045 048	1 038 256	-6 792	1 045 197	149	1 044 095	-953	-0,09
YRIMG1_2016	1 265 550	1 261 424	-4 126	1 265 880	330	1 265 578	28	0,00
YRIMG1_M3	27 740 538	27 530 424	-210 114	27 729 416	-11 122	27 738 072	-2 466	-0,01
YRIMG1_M5	793 008	783 408	-9 600	795 099	2 091	794 295	1 287	0,16

Note : Tous les ménages ayant un poids de sondage de 1 ont été exclus du processus de pondération. Ces ménages étaient situés dans des communautés des Premières Nations, des établissements métis, des régions inuites ou d'autres régions éloignées, ou étaient des ménages privés annexés à un logement collectif.

Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2021.

Graphique 5.2.2

Différences entre les chiffres du recensement et les estimations pondérées à partir des poids de sondage, des poids ajustés pour la couverture et la non-réponse, et des poids finaux

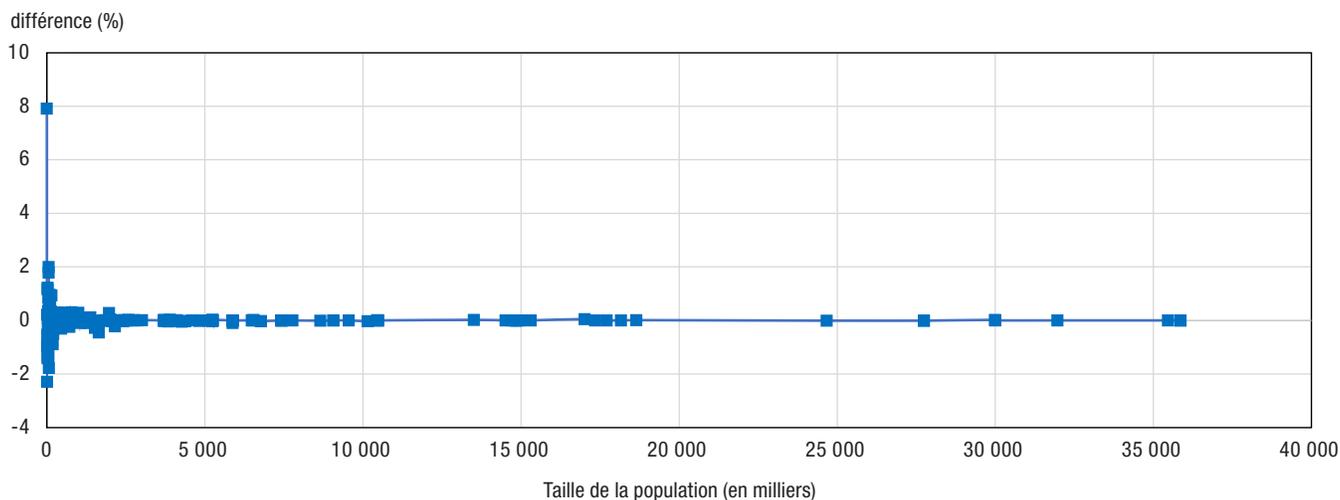


Note : Tous les ménages ayant un poids de sondage de 1 ont été exclus du processus de pondération. Ces ménages étaient situés dans des communautés des Premières Nations, des établissements métis, des régions inuites ou d'autres régions éloignées, ou étaient des ménages privés annexés à un logement collectif.

Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2021.

Graphique 5.2.3

Écart entre les chiffres du recensement et les estimations finales, en pourcentage du chiffre du recensement



Note : Tous les ménages ayant un poids de sondage de 1 ont été exclus du processus de pondération. Ces ménages étaient situés dans des communautés des Premières Nations, des établissements métis, des régions inuites ou d'autres régions éloignées, ou étaient des ménages privés annexés à un logement collectif.

Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2021.

6. Estimation de la variance

L'erreur d'une estimation est la différence entre l'estimation et la valeur réelle de ce que l'on vise à estimer. Toutes les estimations des questionnaires du recensement sont sujettes aux erreurs non dues à l'échantillonnage comme l'erreur de non-réponse totale. Les estimations du questionnaire détaillé sont aussi sujettes à l'erreur d'échantillonnage. L'erreur d'échantillonnage provient du fait que les estimations sont produites à partir d'observations provenant d'un échantillon et non du Recensement de la population. L'erreur de non-réponse totale survient quand des ménages sélectionnés dans l'échantillon ne répondent pas à l'enquête.

L'erreur d'une estimation a une composante aléatoire, mesurée par la variance, et une composante systématique, mesurée par 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 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. Les méthodes d'échantillonnage et d'estimation utilisées dans l'enquête-échantillon du questionnaire détaillé visent toutes à réduire au minimum le biais.

Certaines méthodes d'estimation sont plus précises que d'autres pour estimer une caractéristique donnée de la population. Elles ont donc une incidence sur l'erreur. La variance estimée peut être utilisée pour produire plusieurs indicateurs de qualité couramment utilisés pour mesurer la précision d'une estimation. Par exemple, elle peut être utilisée pour calculer des erreurs-types, des intervalles de confiance et des coefficients de variation. L'intervalle de confiance a été choisi comme indicateur de la qualité fondé sur la variance pour appuyer les estimations du questionnaire détaillé du Recensement de la population de 2021 parce qu'il permet aux utilisateurs de réaliser facilement une inférence statistique. Par conséquent, les intervalles de confiance accompagnent généralement les estimations du questionnaire détaillé dans les produits de données du Recensement de 2021.

Il est important de bien distinguer ces mesures de variabilité des autres mesures de la qualité qui ne sont pas des mesures de variabilité proprement dites. Des exemples de telles mesures sont les taux de réponse finaux présentés à la [section 3.11](#) et le taux d'imputation d'une réponse manquante. Pour obtenir de plus amples renseignements, veuillez consulter le [chapitre 9](#) du *Guide du Recensement de la population, 2021*, produit n° 98-304-X au catalogue de Statistique Canada, et les [Lignes directrices sur la qualité des données du Recensement de 2021](#), produit n° 98-26-0006 au catalogue de Statistique Canada.

Comme l'échantillon du questionnaire détaillé est stratifié géographiquement selon des strates à tirage partiel (les UC d'envoi par la poste, de listage/livraison et d'envoi par la poste et livraison à la porte) et des strates à tirage complet (les UC dans les communautés des Premières Nations, les établissements métis, les régions inuites et d'autres régions éloignées), deux estimateurs de variance sont utilisés. Un premier estimateur de variance est utilisé pour estimer la variance dans les aires géographiques à tirage partiel (voir la [section 6.3.1](#)) et un deuxième estimateur est utilisé pour estimer la variance due à la non-réponse totale dans les aires à tirage complet (voir la [section 6.3.2](#)). Pour le reste du présent chapitre, le terme variance est employé pour désigner la variance d'échantillonnage et de non-réponse totale dans les aires géographiques à tirage partiel ou la variance de non-réponse totale dans les aires à tirage complet.

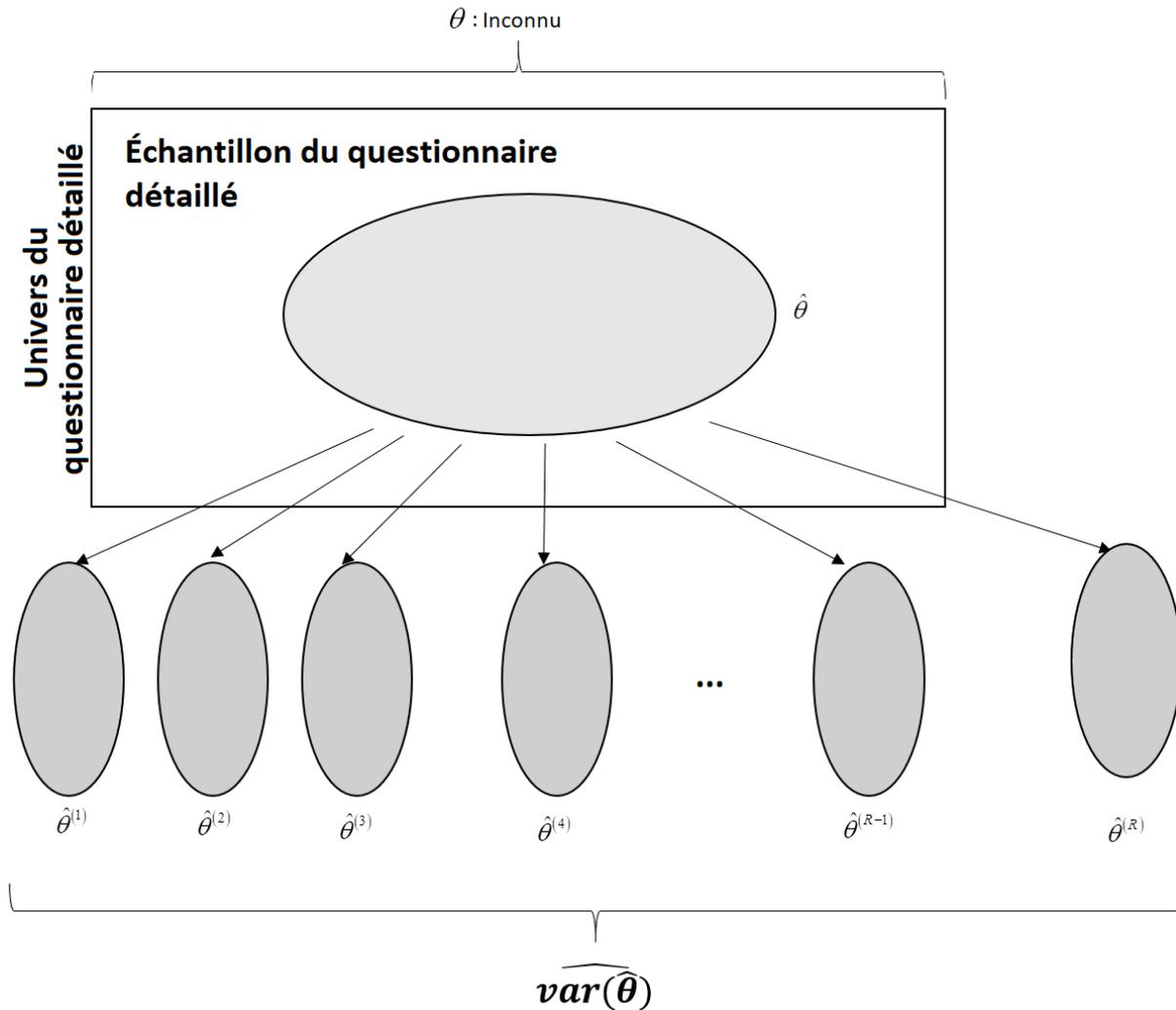
6.1 Éléments à considérer pour choisir une méthode d'estimation de la variance

Un très grand nombre d'estimations diverses ont été produites, et des indicateurs de qualité pour ces mêmes estimations devaient être établis dans un délai raisonnable. Par conséquent, un estimateur de variance par rééchantillonnage, dérivé de la méthode modifiée par répliques répétées partiellement équilibrées, a été utilisé (Judkins, 1990). Cette méthode a été utilisée pour la première fois lors du Recensement de 2016 et a été largement maintenue pour le Recensement de 2021. La méthode consiste à tirer des échantillons (ou répliques) à partir de l'échantillon original. Des poids sont calculés pour chacune des répliques et ceux-ci subissent les mêmes ajustements de couverture, de non-réponse et de calage que l'échantillon original. Dorénavant, les poids établis à partir de l'échantillon original sont appelés poids principaux. Les poids résultant de chacune des répliques sont appelés poids de répliques. Des estimations sont ensuite produites pour chaque réplique, puis la variance est estimée à l'aide des estimations des répliques et des poids principaux.

Rapport technique sur l'échantillonnage et la pondération

La figure 6.1.1 donne un aperçu du concept d'estimation de la variance par répliques lorsque R échantillons sont utilisés.

Figure 6.1.1
Aperçu de l'estimation de la variance par répliques



Source : Statistique Canada.

Description complète :

La figure 6.1.1 donne un aperçu de la méthodologie d'estimation de la variance par répliques utilisée dans le cadre du Recensement de 2021. La méthode d'estimation de la variance par répliques simule la sélection de plusieurs échantillons afin d'estimer la variance d'échantillonnage.

Plus précisément, la figure montre l'univers du questionnaire détaillé qui représente la population d'intérêt ainsi que l'échantillon du questionnaire détaillé. L'échantillon est situé à l'intérieur de l'univers pour indiquer qu'il correspond à un sous-ensemble de la population d'intérêt. Cet échantillon permet d'estimer une caractéristique de la population d'intérêt, par exemple « le nombre de personnes faisant partie d'une minorité visible »⁸. Le symbole θ est utilisé pour représenter la valeur réelle de cette caractéristique. Un accent circonflexe sur le θ indique que la valeur représente une estimation de cette caractéristique. Cette valeur est appelée $\hat{\theta}$ chapeau.

Les R autres échantillons placés hors de l'univers sont reliés à l'échantillon du questionnaire détaillé par des flèches. Les flèches indiquent que ces échantillons sont tirés de l'échantillon du questionnaire détaillé. La caractéristique d'intérêt est réestimée à partir de ces R sous-échantillons. Les R valeurs de $\hat{\theta}$ chapeau, nommées $\hat{\theta}$ chapeau un, $\hat{\theta}$ chapeau deux, jusqu'à $\hat{\theta}$ chapeau R sont utilisées pour calculer la variance estimée de $\hat{\theta}$ chapeau.

Nous définissons :

- θ , la valeur réelle de la caractéristique dans la population, qui peut être un total, une moyenne, un quantile, etc.;
- $\hat{\theta}$, la valeur de θ estimée à partir des poids principaux;
- $\hat{\theta}^{(r)}$, la valeur de θ estimée à partir des poids de la réplique r , $r = 1, \dots, R$;
- $\bar{\hat{\theta}}$, la valeur moyenne des R estimations de répliques $\hat{\theta}^{(r)}$;
- $\widehat{\text{var}}(\hat{\theta})$, la valeur estimée de la variance de $\hat{\theta}$.

6.2 Estimateur de variance

L'estimateur par répliques choisi pour l'enquête-échantillon du questionnaire détaillé a été dérivé de la méthode des demi-échantillons équilibrés de Fay (Judkins, 1990). La méthode détermine la création des répliques, le calcul des poids de répliques ainsi que le facteur multiplicatif utilisé pour estimer la variance.

Afin de produire les estimations de variance des estimations de l'échantillon du questionnaire détaillé, deux ensembles de poids de répliques ont été créés : un premier ensemble pour 32 poids de répliques et un deuxième ensemble pour 100 poids de répliques. L'ensemble de 32 poids de répliques a été produit pour estimer les erreurs-types de produits normalisés dont le calcul est sujet à des contraintes opérationnelles (c.-à-d. le besoin de diffuser un grand nombre d'intervalles de confiance dans un délai raisonnable). L'ensemble de 100 poids de répliques a été mis à la disposition des analystes de Statistique Canada et des analystes des centres de données de recherche qui ont accès aux microdonnées, et ce, afin de leur fournir des estimateurs de variance plus précis.

8. Dans les produits d'analyse et de communication du Recensement de 2021, le terme « minorité visible » a été remplacé par les termes « population racisée » ou « groupes racisés », ce qui reflète l'utilisation accrue de ces termes dans la sphère publique.

L'estimateur de variance par répliques peut être calculé de deux façons, dont une plus conservatrice que l'autre. La première méthode consiste à additionner les différences au carré entre les estimations de répliques, $\hat{\theta}^{(r)}$, et la moyenne des estimations des répliques, $\bar{\theta}$. La seconde méthode consiste à additionner les différences au carré entre les estimations des répliques, $\hat{\theta}^{(r)}$, et l'estimation provenant de l'échantillon principal, $\hat{\theta}$. Pour les deux méthodes, la somme des différences au carré est multipliée par un certain facteur. La deuxième méthode, qui utilise l'estimation provenant de l'échantillon principal, est plus conservatrice. Dans le système informatique utilisé pour la diffusion de statistiques, l'estimateur de variance est calculé à partir de l'estimation provenant de l'échantillon principal.

Par exemple, deux estimateurs de variance pour un estimateur \hat{T} d'un total T calculé à partir d'un ensemble de R répliques sont donnés par les équations suivantes :

$$\widehat{\text{Var}}_1(\hat{T}) = \frac{1}{R/2} \sum_{r=1}^R (\hat{T}^{(r)} - \bar{\hat{T}})^2,$$

$$\widehat{\text{Var}}_2(\hat{T}) = \frac{1}{R/2} \sum_{r=1}^R (\hat{T}^{(r)} - \hat{T})^2$$

où

$$\hat{T} = \sum_{k \in S} w_k y_k,$$

$$\hat{T}^{(r)} = \sum_{k \in S} w_k^{(r)} y_k, \text{ et}$$

$$\bar{\hat{T}} = \sum_{r=1}^R \hat{T}^{(r)} / R.$$

Le poids final de l'échantillon est représenté par w_k , $w_k^{(r)}$ est le poids final de la réplique r , y_k est la valeur de la caractéristique y pour l'unité k et S est l'échantillon du questionnaire détaillé.

Le nombre de degrés de liberté de l'estimateur de variance est approximé par le nombre de différences au carré $(\hat{T}^{(r)} - \bar{\hat{T}})^2$, soit 32 ou 100, de l'estimateur de variance. Le nombre de degrés de liberté donne un aperçu de la précision de l'estimateur de variance et il est utilisé lors du calcul d'intervalles de confiance pour les estimations provenant du questionnaire détaillé. Voir le [chapitre 7](#) pour obtenir plus de détails.

6.3 Ajustement des poids de répliques

6.3.1 Unités de collecte d'envoi par la poste et de listage/livraison

Comme mentionné à la [section 6.2](#), les poids de répliques ont été calculés pour tous les ménages de l'échantillon du questionnaire détaillé. Les répliques ont été partiellement équilibrées. Elles ont été équilibrées par strate de rééchantillonnage, celles-ci étant créées en combinant des UC de façon à obtenir de 600 à 1 800 ménages par strate de rééchantillonnage.

La méthode modifiée des demi-échantillons équilibrés de Fay, décrite par Rao et Shao (1999), exige une valeur « epsilon » dans le calcul des poids de répliques afin de contrôler la perturbation des poids de répliques. Cette perturbation fait en sorte que tous les ménages échantillonnés participent à chacune des répliques, contrairement à d'autres méthodes d'estimation par répliques plus courantes. Cela facilite le calage des poids de répliques et, à l'occasion, le calcul des estimations ponctuelles sur chacune des répliques (p. ex. le dénominateur d'un estimateur de ratio pour une réplique donnée ne sera pas nul si le dénominateur correspondant n'était pas nul avec le poids

final). L'ajout d'un facteur epsilon dans le calcul des poids de répliques a aussi permis de tenir compte de la grande fraction de sondage utilisée pour sélectionner l'échantillon du questionnaire détaillé. Les détails techniques relatifs au processus de l'estimation de la variance sont donnés dans Devin et Verret (2016).

Les poids de répliques ont été assujettis aux mêmes ajustements que le poids de sondage de l'échantillon principal. Ils ont été ajustés pour la couverture et la non-réponse totale en suivant la même méthodologie que pour le poids de l'échantillon principal (voir la [section 4.4](#)). Ensuite, les poids de répliques résultants ont été calés aux chiffres du recensement, toujours en suivant la même méthodologie que pour le poids principal (voir la [section 4.5](#)).

6.3.2 Unités de collecte dans les communautés des Premières Nations, les établissements métis, les régions inuites et d'autres régions éloignées⁹

Comme le décrit le [chapitre 2](#), tous les ménages situés dans des UC dans les communautés des Premières Nations, les établissements métis, les régions inuites et d'autres régions éloignées ont été sélectionnés avec certitude. Par conséquent, ils avaient à l'origine un poids de sondage de 1. Aucun ajustement pour la couverture n'était nécessaire. Tous ces ménages ont été sélectionnés pour le questionnaire détaillé; il ne pouvait donc pas y avoir de différence de couverture entre le questionnaire abrégé et le questionnaire détaillé. La non-réponse totale dans ces aires a été traitée par le processus d'imputation des ménages entiers, décrit au [chapitre 3](#). En d'autres mots, les données d'un ménage non répondant ont été remplacées par les données d'un ménage répondant de la même UC (à l'exception des variables géographiques des non-répondants, que l'on connaissait déjà). La repondération n'a donc pas été nécessaire pour les ménages situés dans ces UC.

Le calage n'a pas non plus été nécessaire pour ces aires, car le questionnaire détaillé était un recensement. Par conséquent, tous les ménages situés dans les UC dans les communautés des Premières Nations, les établissements métis, les régions inuites et d'autres régions éloignées ont conservé leur poids original de 1 dans le plan de pondération final.

Bien qu'il n'y ait pas eu de variabilité d'échantillonnage dans les ménages situés dans les UC dans les communautés des Premières Nations, les établissements métis, les régions inuites et d'autres régions éloignées, il y a eu de la variabilité liée à l'imputation des ménages entiers. L'estimation de la variance dans ces aires a été calculée selon une méthode semblable à celle pour le reste du pays, à quelques exceptions près. Premièrement, la probabilité de réponse selon la combinaison de la taille du ménage dans chaque division de recensement a été estimée en divisant le nombre de ménages répondants par le nombre de ménages dans le champ de l'enquête. Ensuite, les poids de répliques de base ont été créés comme pour le reste du pays, à l'exception du fait que tous les répondants dont la probabilité de réponse était égale à 1 ont été placés dans chaque réplique. Les répondants dont la probabilité de réponse estimée était inférieure à 1 n'ont pas été considérés comme des certitudes et ont été traités comme des éléments échantillonnés (c.-à-d. qu'ils ont été répartis aléatoirement entre les répliques). Les ménages non répondants imputés par l'imputation des ménages entiers ont également été répartis entre les répliques, et chacun s'est vu attribuer l'indicateur d'inclusion dans la réplique correspondant à son donneur d'une manière semblable à celle de Shao et Tang (2011). Ainsi, les poids ont varié d'une réplique à l'autre. Enfin, les poids de répliques ont été calés en fonction du nombre de ménages et du nombre de personnes dans la SADA. Par conséquent, la variance estimée de ces deux quantités était égale à 0 au niveau de la SADA et aux niveaux plus agrégés, par exemple au niveau du Canada (puisque ces deux contraintes sont obligatoires dans le reste du pays).

9. L'exception à cette caractéristique correspondait aux unités situées dans des réserves et des établissements partiellement dénombrés, qui ont été exclues de la population cible et dont le poids a été fixé à 0, sans autre modification à l'ensemble de données ou aux poids. Pour obtenir de plus amples renseignements sur les réserves et les établissements partiellement dénombrés, voir l'[annexe 1.5](#) du *Guide du Recensement de la population, 2021*, n° 98-304-X au catalogue de Statistique Canada.

7. Inférence statistique

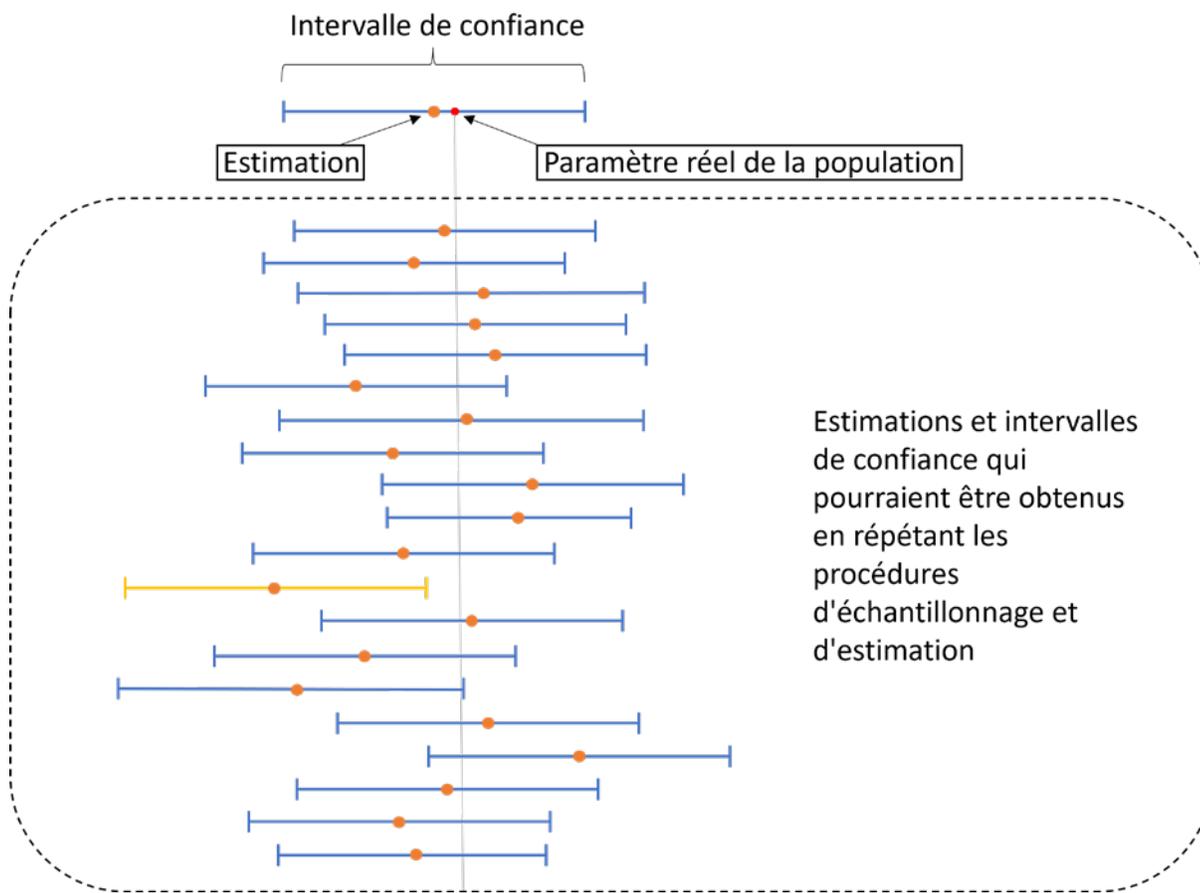
L'inférence statistique dans le contexte de l'échantillonnage d'enquête est le processus visant à tirer des conclusions sur la population à partir des données recueillies auprès des répondants à l'enquête. Pour tirer des conclusions sur la population cible de l'enquête-échantillon du questionnaire détaillé à l'aide des estimations produites à partir de l'échantillon, l'incertitude associée à ces estimations doit être prise en considération. Comme décrit au [chapitre 6](#), les estimations produites à partir de l'échantillon du questionnaire détaillé sont sujettes à la variabilité due à l'échantillonnage et à la non-réponse. La variance pour chaque estimation est une mesure qui permet de quantifier cette variabilité. Les estimations de la variance d'une statistique issue d'une enquête peuvent être utilisées pour produire d'autres mesures de la qualité des statistiques qui reflètent leur variabilité, mais qui sont plus faciles à interpréter que les estimations de la variance elles-mêmes. Ces mesures comprennent les erreurs-types, les coefficients de variation et les intervalles de confiance. Parmi ces mesures, les intervalles de confiance ont l'avantage de permettre aux utilisateurs de réaliser facilement des inférences statistiques. Pour cette raison, l'intervalle de confiance a été choisi comme indicateur de qualité fondé sur la variance pour accompagner les estimations du questionnaire détaillé du Recensement de 2021.

7.1 Intervalles de confiance et leur interprétation

Un intervalle de confiance associé à une estimation est un intervalle calculé autour de l'estimation pour refléter l'incertitude de l'estimation. Un intervalle de confiance est associé à un niveau de confiance, exprimé en pourcentage. Le niveau de confiance indique dans quelle mesure on peut être certain que le paramètre réel de population se trouve à l'intérieur de l'intervalle de confiance. Un niveau de confiance par défaut est généralement établi pour une enquête ou dans un domaine d'études en fonction des besoins des utilisateurs. Un niveau de confiance couramment utilisé est 95 %, et il s'agit du niveau de confiance par défaut pour le système de diffusion des données du recensement. Lorsqu'un utilisateur dispose d'un intervalle de confiance de 95 % pour une estimation, il peut dire qu'il est confiant à 95 % que le paramètre réel de population se trouve à l'intérieur de l'intervalle.

Une interprétation rigoureuse des intervalles de confiance repose sur des répétitions hypothétiques des méthodes d'échantillonnage et d'estimation. Cette interprétation est illustrée à la figure 7.1.1.

Figure 7.1.1
Interprétation des intervalles de confiance



Source : Statistique Canada, Recensement de la population de 2021.

Description complète :

La partie supérieure de la figure 7.1.1 représente un intervalle de confiance associé à l'estimation d'un paramètre de population. Le paramètre réel de population se trouve à l'intérieur de cet intervalle de confiance.

La figure montre également les différentes estimations qui peuvent être produites en répétant hypothétiquement les procédures d'échantillonnage et d'estimation plusieurs fois avec leurs intervalles de confiance correspondants. Dans le cas du questionnaire détaillé, cet échantillonnage et cette estimation répétés seraient effectués en tirant un très grand nombre d'échantillons à partir de l'univers pour l'échantillon du questionnaire détaillé conformément au plan d'échantillonnage décrit au [chapitre 2](#). Chacun des échantillons subirait les mêmes étapes de traitement, de pondération et d'estimation que l'échantillon réel. En général, les estimations et les intervalles de confiance produits pour une caractéristique donnée varieraient selon les échantillons. Cependant, si les hypothèses sous-jacentes de la méthode d'intervalle de confiance donnée sont valides, le pourcentage d'intervalles de confiance qui comprennent la valeur réelle de la population serait approximativement égal au niveau de confiance.

Dans l'exemple illustré à la figure, tous les intervalles de confiance sauf un renferment le paramètre réel de population.

La largeur de l'intervalle de confiance associé à une estimation indique le degré d'incertitude de l'estimation. Si deux estimations sont égales, mais que l'une possède un intervalle de confiance de 95 % qui est plus large que l'autre, l'estimation avec l'intervalle plus large comporte une plus grande incertitude.

Les types d'incertitude reflétés dans les intervalles de confiance pour les estimations du formulaire détaillé diffèrent selon le type de strate. Étant donné que les intervalles de confiance sont fondés sur des estimations de variance, ils reflètent les mêmes types d'incertitude que les estimations de variance sous-jacentes. Dans les unités de collecte (UC) d'envoi par la poste, de listage/livraison et d'envoi par la poste et livraison à la porte, où la fraction de sondage est égale au quart, l'incertitude mesurée est due à l'échantillonnage et à la non-réponse totale. Dans les UC situées dans les communautés des Premières Nations, les établissements métis, les régions inuites et d'autres régions éloignées, où tous les ménages font partie de l'échantillon, l'incertitude mesurée tient compte uniquement de la non-réponse totale.

7.2 Construction d'intervalles de confiance

Les intervalles de confiance sont généralement basés sur les propriétés d'une expression mathématique appelée un pivot. Selon le pivot utilisé et les hypothèses faites sur les propriétés du pivot, différents types d'intervalles de confiance en résulteront. Lors de la construction d'un intervalle de confiance pour un paramètre de population θ , le pivot utilisé fera généralement intervenir l'estimation de θ , désignée par $\hat{\theta}$, et l'erreur-type de $\hat{\theta}$.

L'erreur-type d'une estimation est définie comme étant la racine carrée de la variance estimée de l'estimation.

L'erreur-type de $\hat{\theta}$ est désignée par $\widehat{SE}(\hat{\theta})$. Le pivot peut également faire intervenir d'autres quantités. La méthode de l'intervalle de Wald, qui est la méthode d'estimation d'intervalles de confiance la plus répandue, est basée sur le pivot suivant :

$$\frac{\hat{\theta} - \theta}{\widehat{SE}(\hat{\theta})}$$

L'intervalle de Wald suppose que la distribution de ce pivot suit approximativement une distribution normale standard, c'est-à-dire une distribution normale avec une moyenne de 0 et un écart-type de 1. La borne inférieure (LB) et la borne supérieure (UB) de l'intervalle de confiance de Wald de 95 % pour le paramètre de population θ sont données par :

$$LB = \hat{\theta} - z \times \widehat{SE}(\hat{\theta}),$$
$$UB = \hat{\theta} + z \times \widehat{SE}(\hat{\theta}),$$

où z est le 97,5^e centile de la distribution normale standard.

Dans de nombreuses situations, l'hypothèse selon laquelle le pivot de Wald est approximativement normal n'est pas respectée. Lorsque les hypothèses qui sous-tendent la construction d'un intervalle de confiance ne sont pas respectées, cela peut donner lieu à une sous-couverture de l'intervalle de confiance. En d'autres mots, si on répétait les procédures d'échantillonnage et d'estimation un très grand nombre de fois et que l'on construisait les intervalles de confiance correspondants pour chaque estimation en utilisant la même méthode, la proportion de ces intervalles de confiance qui contiendraient la valeur réelle de la population pourrait être inférieure au niveau de confiance déclaré.

Pour réduire le risque de sous-couverture, des méthodes plus élaborées que la méthode de Wald ont été utilisées pour produire les intervalles de confiance des estimations du questionnaire détaillé. Ces méthodes offrent une couverture plus proche du taux nominal. Cependant, toutes les méthodes de construction d'intervalles de confiance reposent sur des hypothèses qui ne peuvent être vérifiées explicitement dans des cas d'utilisation particuliers. Lorsqu'ils travaillent avec des intervalles de confiance, les utilisateurs de données doivent garder à

l'esprit les cas susceptibles de conduire à ce que les hypothèses sous-jacentes à la construction des intervalles de confiance soient invalides. Les différentes méthodes utilisées pour produire les intervalles de confiance pour les estimations du formulaire détaillé, ainsi que les hypothèses requises, sont décrites en détail dans les sections suivantes.

7.3 Intervalle de confiance de Student

L'intervalle de confiance de Student est utilisé pour toutes les statistiques, sauf les proportions et les comptes. Étant donné que les proportions et les comptes constituent la plupart des estimations diffusées pour le questionnaire détaillé, cette méthode n'est utilisée que pour une minorité des intervalles de confiance diffusés.

L'intervalle de confiance de Student est fondé sur le même pivot que l'intervalle de Wald. Cependant, plutôt que de supposer que la distribution de ce pivot peut être approximée par une distribution normale, la distribution d'approximation est présumée suivre une distribution t de Student. Cette distribution est connue pour être une distribution d'approximation plus appropriée pour le pivot dans les cas où la taille de l'échantillon est petite. La distribution t de Student est définie par un seul paramètre appelé « degrés de liberté ». Le nombre de degrés de liberté de la distribution t de Student est influencé par le plan d'échantillonnage, le nombre d'unités échantillonnées et la méthode d'estimation de la variance. Le nombre de degrés de liberté a une incidence sur la largeur de l'intervalle de confiance. Dans le cas du Recensement de 2021, les degrés de liberté ont été approximés par le nombre de répliques utilisées pour l'estimation de la variance (voir la [section 6.2](#)) et désigné par R .

La borne inférieure et la borne supérieure de l'intervalle de confiance de Student de 95 % pour un paramètre de population d'intérêt θ sont données par :

$$LB = \hat{\theta} - t \times \widehat{SE}(\hat{\theta}),$$
$$UB = \hat{\theta} + t \times \widehat{SE}(\hat{\theta}),$$

où

- $\hat{\theta}$ est l'estimation de θ ;
- t est le 97,5^e centile de la distribution t de Student à R degrés de liberté;
- $\widehat{SE}(\hat{\theta})$ est l'erreur-type de $\hat{\theta}$.

7.3.1 Propriétés de l'intervalle de confiance de Student

La distribution t de Student est presque identique à une distribution normale standard lorsque le nombre de degrés de liberté est très élevé. Lorsque le nombre de degrés de liberté est petit, la distribution t de Student est plus large que la distribution normale standard. Par conséquent, l'intervalle de confiance de Student est plus large que l'intervalle de Wald pour la même estimation. Les intervalles de Wald font souvent l'objet d'une sous-couverture lorsque la taille de l'échantillon est petite. Bien que le questionnaire détaillé du recensement ait une grande taille d'échantillon pour l'ensemble du pays, la taille de l'échantillon dans les petites régions géographiques ou les petits domaines d'intérêt peut être petite. Dans ces cas, l'intervalle de confiance de Student aura généralement une meilleure couverture que l'intervalle de Wald.

En pratique, l'intervalle de confiance de Student peut tout de même faire l'objet d'une sous-couverture dans les cas où les tailles d'échantillon sont très petites. Cela est dû aux insuffisances des hypothèses lorsque la taille de l'échantillon est très petite. Par exemple, la distribution du pivot peut ne pas être approximée adéquatement par une distribution t de Student, ou l'approximation du nombre de degrés de liberté par le nombre de répliques peut surestimer considérablement les degrés de liberté réels de la distribution. La décomposition de ces hypothèses lorsque la taille de l'échantillon est très petite entraînera généralement une sous-couverture des intervalles de confiance de Student.

7.4 Intervalle de confiance de Wilson modifié pour les proportions

Il existe plusieurs méthodes différentes de construction d'intervalles de confiance pour les proportions. Pour le Recensement de 2021, la méthode d'intervalle de confiance de Wilson modifié a été choisie en raison de sa couverture généralement supérieure et des aspects pratiques de sa mise en œuvre. Cette méthode est utilisée pour toutes les statistiques de proportion. La méthode est basée sur l'intervalle de confiance de Wilson pour un plan d'échantillonnage aléatoire simple avec remise (EASAR) (Wilson 1927). Pour le recensement, une version modifiée de cet intervalle de confiance est utilisée afin de tenir compte du plan de sondage complexe (Kott et Carr 1997). De nombreuses études par simulation ont montré que cette méthode donne de meilleurs résultats que les intervalles de confiance de Wald et de Student dans les situations où ces intervalles de confiance produisent une sous-couverture pour les proportions (Neusy et Mantel 2016; Statistique Canada 2023).

Pour les proportions, l'hypothèse selon laquelle le pivot utilisé pour construire l'intervalle de confiance de Student suit une distribution t de Student ne tient pas dans le cas de petites tailles d'échantillon et lorsque la statistique prend des valeurs près de 0 ou de 1. L'intervalle de confiance de Wilson modifié pour une proportion p est plutôt basé sur le pivot suivant :

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n_e}}$$

où

- \hat{p} est l'estimation de p ;
- $n_e = \min(n / \text{deff}(\hat{p}), n)$ est la taille effective de l'échantillon;
- $\text{deff}(\hat{p}) = \frac{\hat{V}(\hat{p})}{\hat{p}(1-\hat{p})/n}$ est l'effet de plan estimé associé à \hat{p} par rapport à un plan de sondage EASAR;
- n est la taille de l'échantillon dans le champ de l'enquête;
- $\hat{V}(\hat{p})$ est la variance estimée de \hat{p} .

La taille de l'échantillon dans le champ de l'enquête est définie comme étant le nombre d'unités échantillonnées qui sont dans le champ d'enquête pour la question correspondant à la proportion p , c'est-à-dire le nombre d'unités échantillonnées auxquelles la question s'applique et qui appartiennent à la population d'intérêt liée à la question. L'expression sous la racine carrée dans le dénominateur du pivot correspond à la variance de l'estimation de la proportion sous un plan de sondage EASAR, mais dont la taille de l'échantillon a été remplacée par la taille effective de l'échantillon. En utilisant la taille effective de l'échantillon dans cette expression, la variance est ajustée pour tenir compte du plan d'échantillonnage complexe du formulaire détaillé du recensement. L'intervalle de confiance de Wilson modifié est basé sur l'hypothèse que ce pivot suit approximativement une distribution t de Student. Comme pour l'intervalle de confiance de Student, pour le Recensement de 2021, le nombre de degrés de liberté de la distribution t de Student est approximé par R , le nombre de répliques utilisées pour l'estimation de la variance.

La borne inférieure et la borne supérieure de l'intervalle de confiance à 95 % de Wilson modifié pour la proportion p sont données par :

$$LB = \frac{\hat{p} + t^2 / 2n_e}{1 + t^2 / n_e} - \frac{t \sqrt{\hat{p}(1 - \hat{p}) + t^2 / 4n_e}}{\sqrt{n_e}(1 + t^2 / n_e)},$$

$$UB = \frac{\hat{p} + t^2 / 2n_e}{1 + t^2 / n_e} + \frac{t \sqrt{\hat{p}(1 - \hat{p}) + t^2 / 4n_e}}{\sqrt{n_e}(1 + t^2 / n_e)},$$

où t est le 97,5^e centile de la distribution t de Student à R degrés de liberté et les autres termes sont définis comme ci-dessus.

7.4.1 Propriétés de l'intervalle de confiance de Wilson modifié pour les proportions

En plus d'obtenir une meilleure couverture que les intervalles de confiance de Wald et de Student pour les petites tailles d'échantillons, et lorsque le paramètre de population est près de 0 ou de 1, l'intervalle de confiance de Wilson modifié pour une proportion donnée possède la propriété désirable d'avoir une borne inférieure qui n'est jamais inférieure à 0 et une borne supérieure qui n'est jamais supérieure à 1. Étant donné que les proportions ne peuvent prendre de valeurs en dehors de l'intervalle de 0 à 1, on peut s'attendre à ce que les intervalles de confiance pour les proportions excluent les valeurs négatives et les valeurs supérieures à 1.

Il convient également de noter que, contrairement aux intervalles de Wald et de Student, l'intervalle de confiance de Wilson modifié pour les proportions est asymétrique, ce qui signifie que l'estimation ne se situera pas exactement au centre de l'intervalle. L'asymétrie est faible lorsque la taille d'échantillon effective est grande ou lorsque la proportion estimée est près de 0,5.

Tout comme les intervalles de confiance de Wald et de Student, l'intervalle de confiance de Wilson modifié pour les proportions peut faire l'objet d'une certaine sous-couverture, particulièrement lorsque la taille d'échantillon est très petite, que la valeur de la proportion est près de 0 ou de 1 ou qu'il existe une forte corrélation entre les membres d'un même ménage. Néanmoins, la méthode de Wilson modifié atteint généralement le taux de couverture nominal dans des situations extrêmes comparativement aux méthodes de Wald et de Student. En général, elle maintient une couverture aussi bonne sinon meilleure que les méthodes de Wald et de Student dans ces situations.

7.5 Intervalle de confiance de Wilson modifié pour les comptes

Pour les estimations des comptes du questionnaire détaillé, la méthode de l'intervalle de confiance utilisée est une méthode de Wilson modifiée semblable à la méthode utilisée pour les statistiques de proportion. Pour les comptes, les intervalles de confiance de Wald et Student n'atteignent pas la couverture nominale lorsque la taille de l'échantillon est petite et lorsque la valeur de la variable d'intérêt est nulle pour presque toutes les unités échantillonnées ou lorsque la valeur est de 1 pour presque toutes les unités échantillonnées. Dans ces situations, la distribution du pivot utilisé pour construire les intervalles de confiance de Wald et de Student n'est généralement pas bien approximée par une distribution normale ou une distribution t de Student.

L'intervalle de confiance de Wilson modifié pour les comptes a été élaboré pour le Recensement de 2021 en tant que solution de rechange aux intervalles de confiance de Wald et de Student pour offrir une couverture plus proche du taux nominal. Il a été testé dans un environnement simulé semblable à celui du formulaire détaillé du recensement et s'est avéré en mesure d'atteindre une bonne couverture (Neusy et coll., 2021).

7.5.1 Intervalle de confiance de Wilson modifié pour les comptes : forme théorique

La version de l'intervalle de confiance de Wilson modifié pour les comptes utilisée dans le cadre du Recensement de 2021 est une approximation d'une forme théorique de l'intervalle. La forme théorique peut être dérivée à l'aide d'une méthode comparable à celle qui a été employée pour l'intervalle de confiance de Wilson modifié pour les proportions. Dans le cas de l'intervalle de confiance de Wilson modifié pour les comptes, la formulation est fondée sur la notion de groupe de calage d'intérêt. Un **groupe de calage** est un ensemble d'unités pour lesquelles les poids de sondage sont calés pour respecter les totaux de contrôle (pour le questionnaire détaillé du recensement, les groupes de calage correspondent aux ADA et aux SADA), et un **groupe de calage d'intérêt** pour un compte Y correspond à un groupe de calage qui pourrait potentiellement renfermer des unités qui présentent la caractéristique d'intérêt pour Y .

L'intervalle de confiance de Wilson modifié pour un compte Y est basé sur le pivot suivant :

$$\frac{\hat{Y} - Y}{\sqrt{Y(N_C - Y) / n_e}}$$

où

- \hat{Y} est l'estimation de Y ;
- N_C est la taille totale de la population des groupes de calage d'intérêt;
- $n_e = \min(n_c / \text{deff}(\hat{Y}), n_c)$ est la taille totale d'échantillon effective dans les groupes de calage d'intérêt;
- n_c est la taille totale d'échantillon dans les groupes de calage d'intérêt;
- $\text{deff}(\hat{Y}) = \frac{\hat{V}(\hat{Y})}{\hat{Y}(N_C - \hat{Y}) / n_c}$ est l'effet de plan estimé associé à \hat{Y} par rapport à un plan de sondage d'EASAR, avec les tailles de population et d'échantillon déterminées en fonction des groupes de calage d'intérêt;
- $\hat{V}(\hat{Y})$ est la variance estimée de \hat{Y} .

Les tailles de population et d'échantillon sont définies par rapport aux groupes de calage d'intérêt, car cela donne un intervalle de confiance avec de bonnes propriétés. Plus précisément, les simulations montrent que l'intervalle de confiance résultant possède de meilleures propriétés de couverture que les autres façons de définir les tailles de population et d'échantillon (Neusy et coll., 2021).

Tout comme l'intervalle de confiance de Wilson modifié pour les proportions, l'intervalle de confiance de Wilson modifié pour les comptes est basé sur l'hypothèse que le pivot suit approximativement une distribution t de Student. Comme tous les intervalles de confiance du Recensement de 2021, le nombre de degrés de liberté de la distribution t de Student est approximé par R , le nombre de répliques utilisées pour l'estimation de la variance.

À partir du pivot ci-dessus, la version théorique de l'intervalle de confiance de Wilson modifié de 95 % pour un compte Y peut être établie. Cette version de l'intervalle a les bornes inférieure et supérieure suivantes :

$$LB = \frac{\hat{Y} + N_C t^2 / 2n_e}{1 + t^2 / n_e} - \frac{t \sqrt{\hat{Y}(N_C - \hat{Y}) + N_C^2 t^2 / 4n_e}}{\sqrt{n_e}(1 + t^2 / n_e)},$$

$$UB = \frac{\hat{Y} + N_C t^2 / 2n_e}{1 + t^2 / n_e} + \frac{t \sqrt{\hat{Y}(N_C - \hat{Y}) + N_C^2 t^2 / 4n_e}}{\sqrt{n_e}(1 + t^2 / n_e)},$$

où les termes sont définis comme ci-dessus. Dans cette version de l'intervalle, la similitude avec l'intervalle de confiance de Wilson modifié pour les proportions est apparente.

7.5.2 Intervalle de confiance de Wilson modifié pour les comptes : forme approximative

La forme approximative de l'intervalle de confiance de Wilson modifié pour les comptes qui a été mis en œuvre pour le Recensement de 2021 est basée sur l'intervalle théorique ci-dessus. Une approximation a été utilisée en raison des limites du système de tabulation du recensement.

Dans la forme approximative, la borne inférieure et la borne supérieure de l'intervalle de confiance de Wilson de 95 % modifié pour un compte Y sont données par :

$$LB = \hat{Y} + t^2 \frac{1}{2} \frac{\hat{V}(\hat{Y})}{\hat{Y}} - \sqrt{t^2 \hat{V}(\hat{Y}) + \left(t^2 \frac{1}{2} \frac{\hat{V}(\hat{Y})}{\hat{Y}} \right)^2},$$

$$UB = \hat{Y} + t^2 \frac{1}{2} \frac{\hat{V}(\hat{Y})}{\hat{Y}} + \sqrt{t^2 \hat{V}(\hat{Y}) + \left(t^2 \frac{1}{2} \frac{\hat{V}(\hat{Y})}{\hat{Y}} \right)^2},$$

où

- \hat{Y} est l'estimation de Y ;
- t est le 97,5^e centile de la distribution t de Student à R degrés de liberté;
- $\hat{V}(\hat{Y})$ est la variance estimée de \hat{Y} .

7.5.3 Hypothèses pour l'approximation de l'intervalle de confiance de Wilson modifié pour les comptes

Il a été démontré empiriquement que la méthode d'approximation mise en œuvre pour le recensement convient aux plans d'échantillonnage semblables à celui du questionnaire détaillé (Neusy et coll., 2021). L'approximation est basée sur les hypothèses suivantes :

1. La taille totale de l'échantillon dans les groupes de calage d'intérêt n_c est suffisamment grande pour que $1/n_e$ soit proche de 0.
2. Le nombre estimé \hat{Y} est beaucoup plus petit que la taille totale de la population des groupes de calage d'intérêt N_C .

La première hypothèse est généralement valide pour le questionnaire détaillé du recensement, car les groupes de calage d'intérêt correspondent aux SADA ou aux ADA, qui ont toujours une très grande taille d'échantillon. La deuxième hypothèse peut ne pas être respectée, mais uniquement pour des caractéristiques communes et pour de très grands domaines d'intérêt, tels que les SADA. Dans cette situation, la version approximative de l'intervalle de confiance de Wilson modifié est presque identique à l'intervalle de confiance de Student, et la performance des deux méthodes est assez bonne dans le cas de grands domaines d'estimation. Par conséquent, le non-respect de la deuxième hypothèse ne pose pas de problème particulier pour le questionnaire détaillé du recensement.

7.5.4 Propriétés de l'intervalle de confiance de Wilson modifié pour les comptes

L'intervalle de confiance de Wilson modifié pour les comptes a un avantage par rapport aux intervalles de confiance de Wald et de Student, soit d'avoir une borne inférieure qui n'est jamais inférieure à 0. Cette propriété, ainsi que les autres propriétés décrites dans cette section, s'applique à la fois à la version théorique et à la version approximative de l'intervalle de confiance. Étant donné que la valeur des comptes ne peut pas être négative, il convient que l'intervalle de confiance ne contienne pas de valeurs négatives. Comme pour l'intervalle de confiance de Wilson modifié pour les proportions, l'intervalle de confiance de Wilson modifié pour les comptes est asymétrique. Cette asymétrie sera faible lorsque la variance estimée d'un compte est petite par rapport au compte estimé lui-même.

Tout comme les intervalles de confiance de Wald et de Student, l'intervalle de confiance de Wilson modifié pour les comptes peut faire l'objet d'une certaine sous-couverture, particulièrement lorsque la taille d'échantillon est très petite, que la valeur du compte est près de 0 ou près de la taille de la population du domaine d'intérêt ou qu'il existe une forte corrélation entre les membres d'un même ménage. Néanmoins, la méthode de Wilson modifié atteint généralement le taux de couverture nominal dans des situations extrêmes comparativement aux méthodes de Wald et de Student. En général, elle maintient une couverture aussi bonne sinon meilleure que les méthodes de Wald et de Student dans ces situations.

8. Conclusion

Le questionnaire détaillé du Recensement de 2021 a été marqué par l'ajout d'une quantité considérable de nouveau contenu et par le fait que ses activités de collecte ont été menées pendant une pandémie. Ces deux facteurs ont entraîné des difficultés sur les plans méthodologique et opérationnel. En particulier, des efforts importants ont été déployés pour obtenir des taux de réponse semblables à ceux obtenus en 2016. De plus, des données administratives ont été introduites dans le processus d'imputation pour essayer de réduire au minimum le biais de non-réponse dans les rares endroits où des taux de réponse élevés n'ont pas pu être atteints. Malgré ces défis, les méthodes d'échantillonnage et de pondération constituaient dans l'ensemble une continuation des méthodes du questionnaire détaillé du Recensement de 2016.

L'introduction d'un nouveau système de diffusion des données a permis à Statistique Canada de remanier et d'améliorer la diffusion des indicateurs de qualité des données. Dans le but de permettre aux utilisateurs de données de faire des inférences statistiques valides, de nouvelles méthodes d'intervalles de confiance ont été développées et mises en place. Les intervalles de confiance sont devenus le principal indicateur de la qualité des données fondé sur la variance et sont diffusés avec la majorité des tableaux de données du questionnaire détaillé. De plus, des statistiques détaillées sur les taux de non-réponse et d'imputation par question ont été offertes aux utilisateurs de données pour rendre la qualité des données plus transparente. Pour obtenir de plus amples renseignements, veuillez consulter les [Lignes directrices sur la qualité des données du Recensement de 2021](#), produit n° 98-26-0006 au catalogue de Statistique Canada.

Annexe A – Glossaire

Voici les définitions des termes, variables et concepts principaux du recensement mentionnés dans le présent document. Les utilisateurs peuvent également consulter le [Dictionnaire, Recensement de la population, 2021](#), produit n° 98-301-X au catalogue pour obtenir plus de renseignements.

Aire de diffusion agrégée : Une aire de diffusion agrégée (ADA) est une région géographique de diffusion créée pour le Recensement. Les ADA couvrent l'ensemble du pays et, dans la mesure du possible, ont une population de 5 000 à 15 000 habitants selon les chiffres de population du recensement précédent. On crée les ADA en regroupant des régions géographiques de diffusion existantes, y compris les secteurs de recensement (SR), les subdivisions de recensement (SDR) et les aires de diffusion (AD). Les limites de l'ADA respectent les limites provinciales et territoriales ainsi que les limites des divisions de recensement (DR), des régions métropolitaines de recensement (RMR) et des agglomérations de recensement (AR).

L'objectif de l'ADA consiste à garantir la disponibilité de données du recensement, dans la mesure du possible, pour toutes les régions du Canada.

Division de recensement : Groupe de municipalités voisines les unes des autres qui sont réunies pour des besoins de planification régionale et de gestion de services communs (comme les services de police et d'ambulance). Ces groupes sont créés selon les lois en vigueur dans certaines provinces du Canada. Division de recensement (DR) est le terme général de régions créées en vertu des lois provinciales (comme comté, municipalité régionale de comté [MRC] et *regional district*) ou des régions équivalentes. Dans d'autres provinces et dans les territoires, où les lois ne prévoient pas de telles régions, Statistique Canada définit des régions équivalentes à des fins statistiques en collaboration avec ces provinces et territoires. Les divisions de recensement sont des régions géographiques intermédiaires entre la municipalité (subdivision de recensement) et la province/territoire.

Famille de recensement : Famille de recensement est définie comme suit : un couple marié et les enfants, le cas échéant, du couple et/ou de l'un ou l'autre des conjoints; un couple en union libre et les enfants, le cas échéant, du couple et/ou de l'un ou l'autre des conjoints; ou un parent dans une famille monoparentale, peu importe son état matrimonial, habitant avec au moins un enfant dans le même logement et cet enfant ou ces enfants. Tous les membres d'une famille de recensement particulière habitent le même logement. Les enfants peuvent être des enfants biologiques ou adoptés, peu importe leur âge ou leur état matrimonial, du moment qu'ils habitent dans le logement sans leur propre conjoint marié, conjoint de fait ou enfant. Les petits-enfants habitant avec leurs grands-parents, alors qu'aucun des parents n'est présent, constituent également une famille de recensement.

Subdivision de recensement : Subdivision de recensement (SDR) est le terme générique qui désigne les municipalités (selon les lois provinciales/territoriales) ou les territoires considérés comme des équivalents municipaux à des fins statistiques (p. ex. les réserves indiennes, les établissements indiens et les territoires non organisés). Le statut de municipalité est défini par les lois en vigueur dans les provinces et les territoires au Canada.

Secteur de recensement : Les secteurs de recensement (SR) sont de petites régions géographiques relativement stables qui comptent habituellement une population de moins de 7 500 habitants d'après les données du Programme du recensement précédent sur la population. Ils sont créés au sein de régions métropolitaines de recensement (RMR) et d'agglomérations de recensement (AR) dont le noyau comptait 50 000 habitants ou plus d'après le recensement précédent.

Un comité de spécialistes locaux (par exemple, des planificateurs municipaux) délimite initialement les SR de concert avec Statistique Canada. Une fois qu'une RMR ou qu'une AR a été subdivisée en SR, les SR sont maintenus même si, ultérieurement, la population du noyau de la RMR ou de l'AR devient inférieure à 50 000 habitants.

Unité de collecte¹⁰ : Les unités de collecte (UC) sont de petites unités géographiques utilisées pour la collecte des données du recensement. Les UC couvrent tout le territoire du Canada.

10. Cette définition ne figure pas dans le [Dictionnaire, Recensement de la population, 2021](#), car le dictionnaire est composé principalement de termes liés à la diffusion, tandis qu'« unité de collecte » est un terme relatif à la collecte.

Logement collectif : Logement collectif désigne un logement de nature commerciale, institutionnelle ou communautaire dans lequel une personne ou un groupe de personnes réside ou pourrait résider. Il doit fournir des soins ou des services ou disposer de certaines installations communes, telles qu'une cuisine ou une salle de bain, qui sont partagées par les occupants. Les logements collectifs comprennent, par exemple, les pensions et maisons de chambres, les hôtels, les motels, les établissements pour touristes, les établissements de soins infirmiers, les résidences pour personnes âgées, les hôpitaux, les résidences de personnel, les bases militaires, les campements de travailleurs, les établissements correctionnels et les foyers collectifs.

Aire de diffusion : Une aire de diffusion (AD) est une petite unité géographique relativement stable formée d'un ou de plusieurs îlots de diffusion avoisinants dont la population moyenne est de 400 à 700 habitants d'après les données du Recensement de la population précédent. 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.

Logement : Un logement est défini comme un ensemble de pièces d'habitation. Dans le cadre du recensement, on reconnaît deux genres de logements, soit les logements collectifs et les logements privés. Les premiers sont des logements de nature institutionnelle, communautaire ou commerciale. Les autres désignent un ensemble distinct de pièces d'habitation, auquel on a accès par une entrée privée soit à partir de l'extérieur de l'immeuble, soit à partir d'un hall, d'un foyer, d'un vestibule ou d'un escalier commun situé à l'intérieur de l'immeuble. Il faut qu'on puisse emprunter l'entrée menant au logement sans passer par les pièces d'habitation d'une autre personne ou d'un autre groupe de personnes.

Famille économique : « Famille économique » désigne un groupe de deux personnes ou plus habitant dans le même logement et apparentées par le sang, le mariage, l'union libre, l'adoption ou une relation de famille d'accueil.

Par définition, toutes les personnes qui sont membres d'une famille de recensement sont également membres d'une famille économique. Voici des exemples du vaste concept de famille économique : deux familles de recensement apparentées et vivant ensemble sont considérées comme une famille économique; les frères et sœurs habitant ensemble qui ne sont pas membres d'une famille de recensement sont considérés comme une famille économique; et les nièces ou les neveux habitant avec leur tante ou leur oncle sont considérés comme une famille économique.

Ménage : Ménage réfère à une personne ou à un groupe de personnes qui occupent le même logement et n'ont pas de domicile habituel ailleurs au Canada ou à l'étranger. Le logement peut être un logement collectif ou un logement privé. Le ménage peut se composer d'un groupe familial (famille de recensement), de deux familles ou plus partageant le même logement, d'un groupe de personnes non apparentées ou d'une personne vivant seule. Les membres du ménage qui sont temporairement absents le jour de référence sont considérés comme faisant partie de leur ménage habituel.

Logement privé : « Logement privé » s'entend d'un ensemble séparé de pièces d'habitation possédant une entrée privée soit à partir de l'extérieur de l'immeuble, soit à partir d'un hall, d'un foyer, d'un vestibule ou d'un escalier commun situé à l'intérieur de l'immeuble. Il faut qu'on puisse emprunter l'entrée menant au logement sans passer par les pièces d'habitation d'une autre personne ou d'un autre groupe de personnes.

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 (ce dont témoignent les cheminées, les lignes électriques, les conduites ou les compteurs d'huile ou de gaz, les génératrices, les cordes de bois, les ampoules électriques, les thermopompes, ou les panneaux solaires);
2. fournir un espace clos permettant de s'abriter des intempéries comme en témoigne la présence de murs d'enceinte et d'un toit ainsi que de portes et fenêtres qui protègent 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 parmi les logements privés ordinaires et les 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 uniquement par des résidents étrangers ou par des personnes présentes temporairement, et les logements inoccupés. Les logements

marginiaux sont classés parmi les logements occupés par des résidents habituels ou parmi les logements occupés par des résidents étrangers ou par des personnes présentes temporairement. Les logements marginaux inoccupés le 11 mai 2021 ne font pas partie du parc immobilier.

Logement privé occupé par des résidents habituels : Un logement privé occupé par des résidents habituels désigne un 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 11 mai 2021. Sauf indication contraire, toutes les données présentées dans les produits sur le logement ont trait aux logements privés occupés par des résidents habituels, et non aux logements privés inoccupés ou aux logements occupés uniquement par des résidents étrangers ou par des personnes présentes temporairement.

Ménage privé : Ménage privé s'applique à une personne ou un groupe de personnes qui occupent le même logement et qui n'ont pas de domicile habituel ailleurs au Canada ou à l'étranger. Les ménages sont divisés en deux sous-univers selon qu'ils occupent un logement collectif (ménage collectif) ou un logement privé (ménage privé).

Dans le cadre du recensement, les ménages sont classés en trois catégories : les ménages privés, les ménages collectifs et les ménages à l'extérieur du Canada.

À moins d'indications contraires, toutes les données contenues dans les produits du recensement se rapportent aux ménages privés seulement.

Super aire de diffusion agrégée¹¹ : Les super aires de diffusion agrégées (SADA) ont été créées spécifiquement pour la pondération des données du recensement en respectant des règles préétablies, dont certaines sont obligatoires et d'autres optionnelles. Les SADA sont créées en regroupant des aires de diffusion agrégées (ADA) et elles respectent les limites provinciales et territoriales. Toutes les personnes vivant dans des unités de collecte (UC) de recensement dans les communautés des Premières Nations, les établissements métis, les régions inuites et d'autres régions éloignées sont exclues de la population des SADA. Dans la mesure du possible, les SADA sont créées afin d'avoir une population de 50 000 à 150 000 personnes. Les divisions de recensement (DR) qui ont une population de 40 000 à 50 000 personnes forment leur propre SADA. De plus, les SADA respectent, si possible, en ordre de priorité, les limites des DR, des régions métropolitaines de recensement (RMR), des agglomérations de recensement (AR) et des subdivisions de recensement (SDR). Enfin, les SADA devaient être formées par le regroupement d'ADA adjacentes lorsque cela était faisable, et être aussi compactes que possible.

11. Cette définition ne figure pas dans le [Dictionnaire, Recensement de la population, 2021](#), car le dictionnaire est composé principalement de termes liés à la diffusion, tandis que « super aire de diffusion agrégée » est un terme relatif à la pondération.

Annexe B – Historique de l'échantillonnage dans le cadre du recensement canadien

L'échantillonnage a été utilisé pour la première fois lors du recensement canadien de 1941. Un questionnaire sur le logement a été rempli à tous les 10 logements. Les réponses aux 27 questions de ce questionnaire ont été intégrées aux données de la section sur les personnes et le ménage du questionnaire sur la population correspondant à ces logements. Cela a permis de faire le recoupement des caractéristiques de l'échantillon et des caractéristiques de base. Lors de ce même recensement, on a utilisé l'échantillonnage à l'étape du traitement des données pour obtenir des estimations provisoires des gains 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 des logements en 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. Dans le cadre du Recensement de 1961, on a prélevé un échantillon de 20 % des ménages privés et on a demandé aux personnes âgées de 15 ans et plus de ces ménages de remplir un questionnaire-échantillon sur la population qui comportait 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 coll., 1959), selon lesquels la présence d'un recenseur avait un effet important sur la variance des chiffres du recensement lors d'un recensement par interview. Ainsi, on s'attendait à ce que l'autodénombrement réduise la variance des chiffres du recensement attribuable à la présence du recenseur. En même temps, cela donnerait au répondant plus de temps 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 antérieurs. Une étude des coûts, du fardeau du répondant et de la rapidité d'exécution par rapport à la qualité des données dans le cadre d'un recensement par autodénombrement et par échantillonnage a motivé la décision de recueillir toutes les données, sauf celles portant sur certaines caractéristiques de base, auprès d'un échantillon du tiers de la population au Recensement de 1971. Sauf dans les régions les plus éloignées du Canada, le tiers des ménages privés a reçu le « questionnaire complet » qui comportait toutes les questions du recensement. Le reste des ménages privés a reçu le « questionnaire abrégé » qui ne contenait que les questions de base sur 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 installations sanitaires ainsi que certaines questions en rapport avec la couverture du recensement. Tous les ménages vivant dans des régions éloignées prédéterminées 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 plus détaillée des projets d'application de la méthode d'échantillonnage au Recensement de 1971.

Le Recensement de 1976 comportait nettement moins de contenu que celui 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 les plus élevés (p. ex. les questions sur la profession, l'industrie et le lieu de travail). Par conséquent, les avantages de l'échantillonnage en matière d'économies et d'allègement du fardeau du répondant étaient moins évidents qu'au Recensement de 1971. Néanmoins, après avoir évalué les économies éventuelles liées à diverses fractions d'échantillonnage et étudié les problèmes sur le plan des relations publiques du retour au dénombrement intégral alors que l'échantillonnage avait fait ses preuves en 1971, Statistique Canada a 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 privés occupés. Les études réalisées à ce moment-là ont révélé que la diminution de la qualité des données (mesurée au moyen de la variance) qui en résultait serait acceptable et qu'elle n'aurait pas une incidence suffisante pour annuler les avantages en matière de réduction des coûts, du fardeau du répondant et des délais d'exécution (Royce, 1983). Le taux d'échantillonnage du cinquième des ménages a été la méthode retenue pour tous les recensements de 1981 à 2006.

En 2011, les renseignements qui étaient auparavant recueillis au moyen du questionnaire détaillé du recensement (à participation obligatoire) ont été recueillis dans le cadre de l'Enquête nationale auprès des ménages (ENM), à participation volontaire. En raison de ce changement, chaque ménage a dû répondre aux 10 questions contenues dans le questionnaire du Recensement de 2011, tandis que 30 % des ménages ont été sélectionnés pour répondre à l'ENM. Par ailleurs, 1 ménage sur 3 n'ayant pas répondu à l'ENM a été sous-échantillonné aux fins de suivi. La fraction d'échantillonnage accrue de 30 % a été appliquée en prévision d'un taux de réponse plus faible à l'ENM. Pour le Recensement de 2016, le gouvernement a de nouveau rendu obligatoire le questionnaire détaillé du recensement pour remplacer l'ENM. La fraction d'échantillonnage a été modifiée à 1 sur 4 en 2016, par rapport à 1 sur 5 pour le questionnaire détaillé du Recensement de 2006, en vue de réduire le risque que le taux de réponse ne revienne pas à ses niveaux élevés du passé.

En 2021, le plan d'échantillonnage et le taux d'échantillonnage de 1 sur 4 sont restés presque identiques à ceux du Recensement de 2016. Seule une petite amélioration opérationnelle a été apportée afin de soutenir le départ aléatoire de la procédure d'échantillonnage systématique. Cela a été fait dans le but de s'assurer que les échantillons des deux cycles étaient statistiquement indépendants, car des échantillons dépendants auraient rendu les comparaisons historiques plus difficiles.

Annexe C – Contraintes utilisées ou exclues dans le processus de pondération

Voici la liste des contraintes potentielles définies au niveau des aires de diffusion agrégées (ADA) et au niveau des super aires de diffusion agrégées (SADA). Il y a 271 contraintes potentielles au niveau des ADA et 203 contraintes potentielles au niveau des SADA. Le tableau présente le nombre de fois qu'une contrainte a été utilisée pour le calage et le nombre de fois que les contraintes ont été exclues. Au total, 408 SADA et 4 207 ADA ont été soumises aux processus de pondération. Même si les contraintes HHADACSD et PPADACSD portent la mention de « contraintes ADA », elles pourraient éventuellement être calées pour plusieurs subdivisions de recensement au sein d'une ADA.

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage, selon la contrainte

Nom de la contrainte	Description	Ajustement pour la couverture et la non-réponse			Calage final		
		Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
ADULTCF	Adultes faisant partie d'une famille de recensement	SADA	155	253	SADA	8	400
AGE00_14	Personnes âgées de 0 à 14 ans	SADA	350	58	Les deux	4 360	255
AGE14	Personnes âgées de 10 à 14 ans	SADA	264	144	Les deux	1 022	3 593
AGE15_24	Personnes âgées de 15 à 24 ans	SADA	317	91	SADA	312	96
AGE15_29	Personnes âgées de 15 à 29 ans	ADA	4 059	148
AGE19	Personnes âgées de 15 à 19 ans	SADA	82	326	Les deux	710	3 905
AGE24	Personnes âgées de 20 à 24 ans	SADA	79	329	Les deux	726	3 889
AGE25_34	Personnes âgées de 25 à 34 ans	SADA	384	24	SADA	385	23
AGE29	Personnes âgées de 25 à 29 ans	SADA	365	43	Les deux	1 426	3 189
AGE30_49	Personnes âgées de 30 à 49 ans	ADA	4 143	64
AGE34	Personnes âgées de 30 à 34 ans	SADA	380	28	Les deux	1 736	2 879
AGE35_44	Personnes âgées de 35 à 44 ans	SADA	405	3	SADA	402	6
AGE39	Personnes âgées de 35 à 39 ans	SADA	389	19	Les deux	1 706	2 909
AGE4	Personnes âgées de 0 à 4 ans	SADA	277	131	Les deux	796	3 819
AGE44	Personnes âgées de 40 à 44 ans	SADA	388	20	Les deux	1 547	3 068
AGE45_54	Personnes âgées de 45 à 54 ans	SADA	383	25	SADA	381	27
AGE49	Personnes âgées de 45 à 49 ans	SADA	373	35	Les deux	1 520	3 095
AGE50_64	Personnes âgées de 50 à 64 ans	ADA	4 022	185
AGE54	Personnes âgées de 50 à 54 ans	SADA	378	30	Les deux	1 665	2 950
AGE55_64	Personnes âgées de 55 à 64 ans	SADA	393	15	SADA	365	43

Rapport technique sur l'échantillonnage et la pondération

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage, selon la contrainte

Nom de la contrainte	Description	Région	Ajustement pour la couverture et la non-réponse		Région	Calage final	
			Nombre de contraintes calées	Nombre de contraintes exclues		Nombre de contraintes calées	Nombre de contraintes exclues
AGE59	Personnes âgées de 55 à 59 ans	SADA	382	26	Les deux	1 685	2 930
AGE64	Personnes âgées de 60 à 64 ans	SADA	379	29	Les deux	1 743	2 872
AGE65PL	Personnes âgées de 65 ans et plus	SADA	406	2	Les deux	4 295	320
AGE74	Personnes âgées de 65 à 74 ans	SADA	406	2	Les deux	2 428	2 187
AGE75PL	Personnes âgées de 75 ans et plus	SADA	404	4	Les deux	2 320	2 295
AGE9	Personnes âgées de 5 à 9 ans	SADA	288	120	Les deux	776	3 839
APT5PLUS	Ménages vivant en appartement dans un immeuble de cinq étages ou plus	SADA	261	147	Les deux	1 210	3 405
APTLT5	Ménages vivant en appartement dans un immeuble de moins de cinq étages	SADA	401	7	Les deux	2 585	2 030
CHILD	Enfants faisant partie d'une famille de recensement	SADA	93	315	SADA	13	395
CHILDFAM	Familles de recensement avec enfants	SADA	238	170	Les deux	3 056	1 559
COMLAWNO_DIV	Personnes divorcées ne vivant pas en couple en union libre	SADA	369	39	Les deux	1 190	3 425
COMLAWNO_OTHERS	Personnes divorcées, séparées ou veuves ne vivant pas en couple en union libre	SADA	369	39	Les deux	3 097	1 518
COMLAWNO_SEP	Personnes séparées ne vivant pas en couple en union libre	SADA	292	116	Les deux	575	4 040
COMLAWNO_SINGLE	Personnes jamais mariées ne vivant pas en couple en union libre	SADA	261	147	Les deux	3 648	967
COMLAWNO_SINGLE_GE15	Personnes de 15 ans et plus jamais mariées ne vivant pas en couple en union libre	SADA	259	149	Les deux	3 656	959
COMLAWNO_SINGLE_LT15	Personnes de moins de 15 ans jamais mariées ne vivant pas en couple en union libre	SADA	350	58	Les deux	4 360	255
COMLAWNO_WID	Personnes veuves ne vivant pas en couple en union libre	SADA	290	118	Les deux	679	3 936
COMLAWYE_MARRIED	Personnes mariées ou vivant en couple en union libre	SADA	245	163	Les deux	3 370	1 245

Rapport technique sur l'échantillonnage et la pondération

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage, selon la contrainte

Nom de la contrainte	Description	Ajustement pour la couverture et la non-réponse			Calage final		
		Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
COMLAW_YE	Personnes vivant en couple en union libre	SADA	260	148	Les deux	2 739	1 876
COUPLE	Personnes vivant en couple (mariées ou en union libre)	SADA	156	252	SADA	33	375
EMPIN_GT50	Personnes dont le revenu d'emploi annuel est au-dessus du 50 ^e percentile de l'ADA	ADA	4 127	80
EMPIN_LE50	Personnes dont le revenu d'emploi annuel est égal ou inférieur au 50 ^e percentile de l'ADA	ADA	4 135	72
EMPIN_P0	Personnes dont le revenu d'emploi annuel est nul au niveau de l'ADA	ADA	4 143	64
EMPIN_P0_GE15	Personnes âgées de 15 ans et plus dont le revenu d'emploi annuel est nul au niveau de l'ADA	ADA	4 057	150
EMPIN_P0_LT15	Personnes âgées de moins de 15 ans dont le revenu d'emploi annuel est nul au niveau de l'ADA	ADA	4 033	174
EMPIN_P100	Personnes dont le revenu d'emploi annuel est au-dessus du 75 ^e percentile de l'ADA	ADA	3 707	500
EMPIN_P25	Personnes dont le revenu d'emploi annuel est égal ou inférieur au 25 ^e percentile de l'ADA	ADA	2 922	1 285
EMPIN_P50	Personnes dont le revenu d'emploi annuel est au-dessus du 25 ^e percentile et égal ou inférieur au 50 ^e percentile de l'ADA	ADA	2 918	1 289
EMPIN_P75	Personnes dont le revenu d'emploi annuel est au-dessus du 50 ^e percentile et égal ou inférieur au 75 ^e percentile de l'ADA	ADA	3 708	499
EMPIN_SADA_GT50	Personnes dont le revenu d'emploi annuel est au-dessus du 50 ^e percentile de la SADA	SADA	408	0	SADA	324	84

Rapport technique sur l'échantillonnage et la pondération

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage, selon la contrainte

Nom de la contrainte	Description	Ajustement pour la couverture et la non-réponse			Calage final		
		Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
EMPIN_SADA_LE50	Personnes dont le revenu d'emploi annuel est égal ou inférieur au 50e percentile de la SADA	SADA	407	1	SADA	292	116
EMPIN_SADA_P0	Personnes dont le revenu d'emploi annuel est nul au niveau de la SADA	SADA	407	1	SADA	360	48
EMPIN_SADA_P0_GE15	Personnes âgées de 15 ans et plus dont le revenu d'emploi annuel est nul au niveau de la SADA	SADA	351	57	SADA	328	80
EMPIN_SADA_P0_LT15	Personnes âgées de moins de 15 ans dont le revenu d'emploi annuel est nul au niveau de la SADA	SADA	350	58	SADA	327	81
EMPIN_SADA_P100	Personnes dont le revenu d'emploi annuel est au-dessus du 75e percentile de la SADA	SADA	407	1	SADA	295	113
EMPIN_SADA_P25	Personnes dont le revenu d'emploi annuel est égal ou inférieur au 25e percentile de la SADA	SADA	272	136	SADA	126	282
EMPIN_SADA_P50	Personnes dont le revenu d'emploi annuel est au-dessus du 25e percentile et égal ou inférieur au 50e percentile de la SADA	SADA	271	137	SADA	145	263
EMPIN_SADA_P75	Personnes dont le revenu d'emploi annuel est au-dessus du 50e percentile et égal ou inférieur au 75e percentile de la SADA	SADA	407	1	SADA	288	120
FEMALE	Femmes+	SADA	403	5	Les deux	4 365	250
FEMALEGE15	Femmes de 15 ans et plus	SADA	371	37	Les deux	3 744	871
FEMALELT15	Femmes de moins de 15 ans	SADA	368	40	Les deux	3 636	979
HHADA	Ménages dans l'ADA	ADA	3 592	615
HHADACSD	Ménages dans la SDR et l'ADA	ADA	4 866	2 066
HHINC_GT50	Ménages dont le revenu annuel est supérieur au 50e percentile de l'ADA	ADA	4 178	29

Rapport technique sur l'échantillonnage et la pondération

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage, selon la contrainte

Nom de la contrainte	Description	Ajustement pour la couverture et la non-réponse			Calage final		
		Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
HHINC_LE50	Ménages dont le revenu annuel est égal ou inférieur au 50 ^e percentile de l'ADA	ADA	4 178	29
HHINC_P100	Ménages dont le revenu annuel est supérieur au 75 ^e percentile de l'ADA	ADA	4 033	174
HHINC_P25	Ménages dont le revenu annuel est égal ou inférieur au 25 ^e percentile de l'ADA	ADA	4 052	155
HHINC_P50	Ménages dont le revenu annuel est supérieur au 25 ^e percentile et égal ou inférieur au 50 ^e percentile de l'ADA	ADA	4 052	155
HHINC_P75	Ménages dont le revenu annuel est supérieur au 50 ^e percentile et égal ou inférieur au 75 ^e percentile de l'ADA	ADA	4 033	174
HHINC_SADA_GT50	Ménages dont le revenu annuel est supérieur au 50 ^e percentile de la SADA	SADA	408	0	SADA	364	44
HHINC_SADA_LE50	Ménages dont le revenu annuel est égal ou inférieur au 50 ^e percentile de la SADA	SADA	408	0	SADA	364	44
HHINC_SADA_P100	Ménages dont le revenu annuel est supérieur au 75 ^e percentile de la SADA	SADA	406	2	SADA	320	88
HHINC_SADA_P25	Ménages dont le revenu annuel est égal ou inférieur au 25 ^e percentile de la SADA	SADA	408	0	SADA	372	36
HHINC_SADA_P50	Ménages dont le revenu annuel est supérieur au 25 ^e percentile et égal ou inférieur au 50 ^e percentile de la SADA	SADA	408	0	SADA	352	56
HHINC_SADA_P75	Ménages dont le revenu annuel est supérieur au 50 ^e percentile et égal ou inférieur au 75 ^e percentile de la SADA	SADA	406	2	SADA	333	75
HHSIZE1	Ménage d'une seule personne	SADA	259	149	Les deux	888	3 727
HHSIZE2	Ménages de deux personnes	SADA	237	171	Les deux	3 904	711
HHSIZE3	Ménages de trois personnes	SADA	288	120	Les deux	3 391	1 224
HHSIZE4	Ménages de quatre personnes	SADA	389	19	Les deux	3 527	1 088

Rapport technique sur l'échantillonnage et la pondération

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage, selon la contrainte

Nom de la contrainte	Description	Région	Ajustement pour la couverture et la non-réponse		Région	Calage final	
			Nombre de contraintes calées	Nombre de contraintes exclues		Nombre de contraintes calées	Nombre de contraintes exclues
HHSIZE5	Ménages de cinq personnes	SADA	193	215	Les deux	848	3 767
HHSIZEGE5	Ménages de cinq personnes ou plus	SADA	7	401	Les deux	135	4 480
HHSIZEGE6	Ménages de six personnes ou plus	SADA	15	393	Les deux	42	4 573
INEFAM	Personnes dans une famille économique	SADA	282	126	SADA	177	231
IR_LINK_NO	Personnes qui n'ont pas pu être appariées au Registre des Indiens	SADA	152	256	Les deux	403	4 212
IR_LINK_YE	Personnes qui ont pu être appariées au Registre des Indiens	SADA	152	256	Les deux	334	4 281
LIM_NO	Personnes ne faisant pas partie d'un ménage à faible revenu (après impôt)	SADA	403	5	Les deux	2 820	1 795
LIM_YE	Personnes faisant partie d'un ménage à faible revenu (après impôt)	SADA	403	5	Les deux	2 820	1 795
LONEPAR	Parents dans les familles monoparentales	SADA	159	249	SADA	14	394
MALE	Hommes+	SADA	403	5	Les deux	4 365	250
MALEGE15	Hommes de 15 ans et plus	SADA	362	46	Les deux	3 708	907
MALELT15	Hommes de moins de 15 ans	SADA	360	48	Les deux	3 606	1 009
MARRIED	Personnes mariées	SADA	373	35	Les deux	2 831	1 784
NB_NOTINCF	Personnes ne faisant pas partie d'une famille de recensement	SADA	220	188	Les deux	3 732	883
NOCLDFAM	Familles de recensement sans enfants	SADA	223	185	Les deux	1 630	2 985
NOINEFAM	Personnes ne faisant pas partie d'une famille économique	SADA	282	126	SADA	177	231
NOINEFAMHHSIZEEQ1	Personnes ne faisant pas partie d'une famille économique — Dans un ménage d'une seule personne	SADA	259	149	SADA	50	358
NOINEFAMHHSIZEGT1	Personnes ne faisant pas partie d'une famille économique — Dans un ménage de deux personnes ou plus	SADA	205	203	SADA	59	349

Rapport technique sur l'échantillonnage et la pondération

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage, selon la contrainte

Nom de la contrainte	Description	Région	Ajustement pour la couverture et la non-réponse		Région	Calage final	
			Nombre de contraintes calées	Nombre de contraintes exclues		Nombre de contraintes calées	Nombre de contraintes exclues
NOTINFAM	Personnes ne faisant pas partie d'une famille de recensement	SADA	220	188	SADA	174	234
NOTINFAMHHSIZEEQ1	Personnes ne faisant pas partie d'une famille de recensement — Dans un ménage d'une seule personne	SADA	259	149	SADA	50	358
NOTINFAMHHSIZEGT1	Personnes ne faisant pas partie d'une famille de recensement — Dans un ménage de deux personnes ou plus	SADA	163	245	SADA	32	376
OLN_BI	Langues officielles — Français et anglais	SADA	301	107	Les deux	1 859	2 756
OLN_EN	Langue officielle — Anglais	SADA	69	339	Les deux	495	4 120
OLN_FR	Langue officielle — Français	SADA	80	328	Les deux	552	4 063
OLN_NO	Langue officielle — Aucune	SADA	157	251	Les deux	248	4 367
POBG2_1	Lieu de naissance — Albanie	SADA	0	408	Les deux	5	4 610
POBG2_10	Lieu de naissance — Brésil	SADA	16	392	Les deux	58	4 557
POBG2_11	Lieu de naissance — Bulgarie et Roumanie	SADA	26	382	Les deux	52	4 563
POBG2_16	Lieu de naissance — République démocratique du Congo et République du Congo	SADA	1	407	Les deux	9	4 606
POBG2_17	Lieu de naissance — Cameroun, République centrafricaine, Tchad et Gabon	SADA	2	406	Les deux	12	4 603
POBG2_18	Lieu de naissance — Angola et Sao Tomé-et-Principe	SADA	0	408	Les deux	1	4 614
POBG2_19	Lieu de naissance — Kazakhstan, Kirghizistan, Tadjikistan, Turkménistan et Ouzbékistan	SADA	0	408	Les deux	2	4 613
POBG2_20	Lieu de naissance — Chili	SADA	1	407	Les deux	9	4 606
POBG2_21	Lieu de naissance — Chine, Hong Kong, Macao et Taïwan	SADA	83	325	Les deux	162	4 453
POBG2_22	Lieu de naissance — Colombie, Équateur et Pérou	SADA	47	361	Les deux	107	4 508

Rapport technique sur l'échantillonnage et la pondération

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage, selon la contrainte

Nom de la contrainte	Description	Région	Ajustement pour la couverture et la non-réponse		Région	Calage final	
			Nombre de contraintes calées	Nombre de contraintes exclues		Nombre de contraintes calées	Nombre de contraintes exclues
POBG2_24	Lieu de naissance — République tchèque, Hongrie et Slovaquie	SADA	9	399	Les deux	16	4 599
POBG2_25	Lieu de naissance — Burundi et Rwanda	SADA	2	406	Les deux	4	4 611
POBG2_26	Lieu de naissance — Érythrée, Kenya, Ouganda, Tanzanie et Zambie	SADA	19	389	Les deux	51	4 564
POBG2_27	Lieu de naissance — Comores, Djibouti, Madagascar, Malawi, Maurice, Seychelles, Somalie et Zimbabwe	SADA	6	402	Les deux	16	4 599
POBG2_28	Lieu de naissance — Bélarus, Moldavie, Fédération de Russie et Ukraine	SADA	37	371	Les deux	67	4 548
POBG2_29	Lieu de naissance — Égypte, Soudan du Sud et Soudan	SADA	13	395	Les deux	10	4 605
POBG2_3	Lieu de naissance — Australie et Nouvelle-Zélande	SADA	0	408	Les deux	3	4 612
POBG2_30	Lieu de naissance — Éthiopie	SADA	10	398	Les deux	12	4 603
POBG2_31	Lieu de naissance — France, Luxembourg et Monaco	SADA	18	390	Les deux	32	4 583
POBG2_32	Lieu de naissance — Cambodge, Laos et Viet Nam	SADA	16	392	Les deux	24	4 591
POBG2_33	Lieu de naissance — Cuba, République dominicaine et Haïti	SADA	8	400	Les deux	13	4 602
POBG2_34	Lieu de naissance — Grèce	SADA	2	406	Les deux	10	4 605
POBG2_35	Lieu de naissance — Guyana et Suriname	SADA	18	390	Les deux	39	4 576
POBG2_36	Lieu de naissance — Saint-Siège (État de la Cité du Vatican), Italie et Saint-Marin	SADA	29	379	Les deux	47	4 568
POBG2_37	Lieu de naissance — Bahamas, Jamaïque et Porto Rico	SADA	43	365	Les deux	71	4 544

Rapport technique sur l'échantillonnage et la pondération

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage, selon la contrainte

Nom de la contrainte	Description	Ajustement pour la couverture et la non-réponse			Calage final		
		Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
POBG2_38	Lieu de naissance — Antigua-et-Barbuda, Aruba, Barbade, Bonaire, Saint-Eustache et Saba, Curaçao, Dominique, Grenade, Saint-Kitts-et-Nevis, Sainte-Lucie, Saint-Martin (partie néerlandaise), Saint-Vincent-et-les Grenadines, Trinité-et-Tobago et îles Vierges des États-Unis	SADA	8	400	Les deux	15	4 600
POBG2_39	Lieu de naissance — Japon	SADA	12	396	Les deux	18	4 597
POBG2_4	Lieu de naissance — Autriche, Allemagne et Liechtenstein	SADA	19	389	Les deux	47	4 568
POBG2_40	Lieu de naissance — Corée du Nord et Corée du Sud	SADA	10	398	Les deux	19	4 596
POBG2_41	Lieu de naissance — Libéria	SADA	0	408	Les deux	2	4 613
POBG2_42	Lieu de naissance — Algérie, Libye, Maroc et Tunisie	SADA	14	394	Les deux	10	4 605
POBG2_43	Lieu de naissance — Brunéi Darussalam, Indonésie, Malaisie, Philippines, Singapour et Timor-Leste	SADA	96	312	Les deux	118	4 497
POBG2_45	Lieu de naissance — Mexique	SADA	12	396	Les deux	26	4 589
POBG2_46	Lieu de naissance — Arabie saoudite, Bahreïn, Qatar, Émirats arabes unis et Yémen	SADA	16	392	Les deux	27	4 588
POBG2_47	Lieu de naissance — Liban, Syrie	SADA	26	382	Les deux	49	4 566
POBG2_48	Lieu de naissance — Afghanistan, Chypre, Iran, Iraq, Israël, Jordanie, Koweït, Oman, Turquie, Cisjordanie et bande de Gaza (Palestine)	SADA	87	321	Les deux	107	4 508
POBG2_50	Lieu de naissance — Mozambique	SADA	0	408	Les deux	1	4 614
POBG2_51	Lieu de naissance — Népal	SADA	1	407	Les deux	2	4 613

Rapport technique sur l'échantillonnage et la pondération

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage, selon la contrainte

Nom de la contrainte	Description	Région	Ajustement pour la couverture et la non-réponse		Région	Calage final	
			Nombre de contraintes calées	Nombre de contraintes exclues		Nombre de contraintes calées	Nombre de contraintes exclues
POBG2_54	Lieu de naissance — Pologne	SADA	37	371	Les deux	52	4 563
POBG2_55	Lieu de naissance — Région de l'Océanie (en excluant l'Australie et la Nouvelle-Zélande)	SADA	0	408	Les deux	2	4 613
POBG2_56	Lieu de naissance — Portugal	SADA	30	378	Les deux	40	4 575
POBG2_57	Lieu de naissance — Argentine, Bolivie, Paraguay et Uruguay	SADA	2	406	Les deux	2	4 613
POBG2_59	Lieu de naissance — Namibie et République d'Afrique du Sud	SADA	0	408	Les deux	2	4 613
POBG2_6	Lieu de naissance — Belgique et Pays-Bas	SADA	4	404	Les deux	9	4 606
POBG2_60	Lieu de naissance — Sri Lanka	SADA	7	401	Les deux	16	4 599
POBG2_63	Lieu de naissance — Arménie, Azerbaïdjan et Géorgie	SADA	0	408	Les deux	1	4 614
POBG2_64	Lieu de naissance — Bangladesh, Inde et Pakistan	SADA	39	369	Les deux	69	4 546
POBG2_65	Lieu de naissance — Union des républiques socialistes soviétiques, ancienne	SADA	5	403	Les deux	17	4 598
POBG2_66	Lieu de naissance — Guernesey, Irlande, île de Man, Jersey, Sercq et Royaume-Uni	SADA	5	403	Les deux	0	4 615
POBG2_67	Lieu de naissance — États-Unis	SADA	171	237	Les deux	245	4 370
POBG2_68	Lieu de naissance — Venezuela	SADA	4	404	Les deux	6	4 609
POBG2_69	Lieu de naissance — Gambie, Ghana, Nigéria et Sierra Leone	SADA	4	404	Les deux	3	4 612
POBG2_7	Lieu de naissance — Belize, Salvador, Guatemala et Honduras	SADA	8	400	Les deux	15	4 600
POBG2_70	Lieu de naissance — Bénin, Burkina Faso, Côte d'Ivoire, Guinée, Mali, Mauritanie, Niger, Sénégal et Togo	SADA	2	406	Les deux	2	4 613

Rapport technique sur l'échantillonnage et la pondération

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage, selon la contrainte

Nom de la contrainte	Description	Région	Ajustement pour la couverture et la non-réponse		Région	Calage final	
			Nombre de contraintes calées	Nombre de contraintes exclues		Nombre de contraintes calées	Nombre de contraintes exclues
POBG2_71	Lieu de naissance — Bosnie-Herzégovine, Croatie, Kosovo, République de Macédoine, Monténégro, Serbie et Slovénie	SADA	40	368	Les deux	65	4 550
POBG2_9	Lieu de naissance — Botswana, Lesotho et Swaziland	SADA	1	407	Les deux	0	4 615
POBG3_10	Lieu de naissance — Europe du Nord	SADA	220	188	Les deux	275	4 340
POBG3_12	Lieu de naissance — Océanie	SADA	29	379	Les deux	48	4 567
POBG3_14	Lieu de naissance — Amérique du Sud	SADA	151	257	Les deux	214	4 401
POBG3_15	Lieu de naissance — Asie du Sud-Est	SADA	204	204	Les deux	517	4 098
POBG3_16	Lieu de naissance — Afrique australe	SADA	13	395	Les deux	33	4 582
POBG3_17	Lieu de naissance — Asie du Sud	SADA	181	227	Les deux	648	3 967
POBG3_18	Lieu de naissance — Europe du Sud	SADA	160	248	Les deux	173	4 442
POBG3_19	Lieu de naissance — Union des républiques socialistes soviétiques, ancienne	SADA	88	320	Les deux	142	4 473
POBG3_2	Lieu de naissance — Antilles et Bermudes	SADA	130	278	Les deux	162	4 453
POBG3_20	Lieu de naissance — États-Unis	SADA	171	237	Les deux	245	4 370
POBG3_21	Lieu de naissance — Afrique de l'Ouest	SADA	65	343	Les deux	83	4 532
POBG3_22	Lieu de naissance — Europe de l'Ouest	SADA	169	239	Les deux	265	4 350
POBG3_3	Lieu de naissance — Afrique centrale	SADA	11	397	Les deux	18	4 597
POBG3_4	Lieu de naissance — Amérique centrale	SADA	79	329	Les deux	94	4 521
POBG3_5	Lieu de naissance — Afrique de l'Est	SADA	61	347	Les deux	107	4 508
POBG3_6	Lieu de naissance — Asie de l'Est	SADA	164	244	Les deux	570	4 045
POBG3_7	Lieu de naissance — Europe de l'Est (en excluant l'ancienne Union des républiques socialistes soviétiques)	SADA	116	292	Les deux	157	4 458
POBG3_8	Lieu de naissance — Asie centrale de l'Ouest et Moyen-Orient	SADA	141	267	Les deux	330	4 285
POBG3_9	Lieu de naissance — Afrique du Nord	SADA	69	339	Les deux	132	4 483
PPADA	Personnes dans l'ADA	ADA	3 222	985

Rapport technique sur l'échantillonnage et la pondération

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage, selon la contrainte

Nom de la contrainte	Description	Ajustement pour la couverture et la non-réponse			Calage final		
		Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
PPADACSD	Personnes dans le chevauchement géographique de la SDR et de l'ADA	ADA	4 830	2 102
SNGLDET	Ménages vivant dans une maison individuelle non attenante	SADA	399	9	Les deux	3 639	976
TOTCFAM	Familles de recensement	SADA	367	41	Les deux	1 244	3 371
TOTHHL	Ménages	SADA	408	0	Les deux	4 615	0
TOTPERS	Personnes	SADA	408	0	Les deux	4 615	0
TPERGE15	Personnes âgées de 15 ans et plus	SADA	350	58	Les deux	4 360	255
TPERLT15	Personnes âgées de moins de 15 ans	SADA	350	58	Les deux	4 360	255
YRIMD_1900	Immigrants admis au pays avant 1981	SADA	333	75	SADA	191	217
YRIMD_1980	Immigrants admis au pays de 1980 à 1985	SADA	194	214	SADA	100	308
YRIMD_1986	Immigrants admis au pays de 1986 à 1990	SADA	179	229	SADA	121	287
YRIMD_1991	Immigrants admis au pays de 1991 à 1995	SADA	202	206	SADA	97	311
YRIMD_1996	Immigrants admis au pays de 1996 à 2000	SADA	197	211	SADA	94	314
YRIMD_2001	Immigrants admis au pays de 2001 à 2005	SADA	243	165	SADA	115	293
YRIMD_2006	Immigrants admis au pays de 2006 à 2010	SADA	246	162	SADA	133	275
YRIMD_2011	Immigrants admis au pays de 2011 à 2015	SADA	233	175	SADA	168	240
YRIMD_2016	Immigrants admis au pays de 2016 à 2021	SADA	271	137	SADA	161	247
YRIMD_M3	Personnes dont l'année d'immigration est manquante	SADA	119	289	SADA	65	343
YRIMD_M5	Personnes dont l'année d'immigration est manquante	SADA	192	216	SADA	148	260
YRIMG1_1900	Immigrants admis au pays avant 1981	SADA	333	75	SADA	191	217
YRIMG1_1980	Immigrants admis au pays de 1980 à 1990	SADA	254	154	SADA	143	265
YRIMG1_1991	Immigrants admis au pays de 1991 à 2000	SADA	251	157	SADA	127	281
YRIMG1_2001	Immigrants admis au pays de 2001 à 2010	SADA	300	108	SADA	142	266
YRIMG1_2011	Immigrants admis au pays de 2011 à 2015	SADA	233	175	SADA	168	240
YRIMG1_2016	Immigrants admis au pays de 2016 à 2021	SADA	271	137	SADA	161	247

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage, selon la contrainte

Nom de la contrainte	Description	Ajustement pour la couverture et la non-réponse			Calage final		
		Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
YRIMG1_M3	Personnes dont l'année d'immigration est manquante	SADA	119	289	SADA	65	343
YRIMG1_M5	Personnes dont l'année d'immigration est manquante	SADA	192	216	SADA	148	260

... n'ayant pas lieu de figurer

SDR = Subdivision de recensement

SADA = Super aire de diffusion agrégée

ADA = Aire de diffusion agrégée

Les deux = SADA et ADA

Hommes+ = comprend les hommes (ou les garçons) ainsi que certaines personnes non binaires

Femmes+ = comprend les femmes (ou les filles) ainsi que certaines personnes non binaires

Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2021.

Bibliographie

- Devin, N. et F. Verret. 2016. « The development of a variance estimation methodology for large-scale dissemination of quality indicators for the 2016 Canadian census long form sample ». *JSM Proceedings*. Survey Research Methods Section, American Statistical Association. Chicago, États-Unis.
- 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*, 59(308) : p. 1016 à 1041.
- Folsom, R.E. Jr. et A.C. Singh. 2000. « The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification ». *Proceedings of the Survey Research Methods Section*. American Statistical Association : p. 598 à 603.
- Hansen, M.H., W.N. Hurwitz et M.A. Bershada. 1959. « Measurement errors in censuses and surveys ». *Bulletin of the International Statistical Institute*, 38 : p. 359 à 374.
- Judkins, D.R. 1990. « Fay's method for variance estimation ». *Journal of Official Statistics*. Statistics Sweden, 6(3) : p. 223 à 239.
- Kott, P.S. et D.A. Carr. 1997. « Developing an Estimation Strategy for a Pesticide Data Program ». *Journal of Official Statistics*, 13(4) : p. 367 à 383.
- Neusy, E. et H. Mantel. 2016. « Confidence Intervals for Proportions Estimated from Complex Survey Data ». *Proceedings of the Survey Methods Section*, Assemblée annuelle de la Société statistique du Canada, juin 2016.
- Neusy, E., S.-A. Savard, M. Hidiroglou et V. Martin. 2021. « Modified Wilson Intervals for Estimated Counts with Application to Census 2021 Long Form Estimation ». Présentation au Comité consultatif des méthodes statistiques, mai 2021, document interne, Statistique Canada.
- Rao, J.N.K., et J. Shao. 1999. « Modified balanced repeated replication for complex survey data ». *Biometrika*, Oxford University Press, 86(2) : p. 403 à 415.
- Royce, D. 1983. *The Use of Sampling in the 1981 Canadian Census*. Rapport interne, Statistique Canada. Ottawa (Ontario).
- Savard, S.-A. 2021. « Inference for Census Long-Form Weighted Counts ». *Proceedings of the Survey Methods Section*. Assemblée annuelle de la Société statistique du Canada, juin 2021.
- Shao, J. et Q. Tang. 2011. « Random group variance estimators for survey data with random hot deck imputation ». *Journal of Official Statistics*. Statistics Sweden, 27(3) : p. 507 à 526.
- Statistique Canada. 2020. *Système canadien de vérification et d'imputation du recensement (SCANVIR), version 5.4, guide de base de l'utilisateur*. Ottawa (Ontario).
- Statistique Canada. 2018. [Rapport technique sur l'échantillonnage et la pondération, Recensement de la population, 2016](#), produit n° 98-306-X2016001 au catalogue. Ottawa (Ontario) (site consulté le 6 janvier 2023).
- Statistique Canada. 2022. [Guide du Recensement de la population, 2021](#), produit n° 98-304-X2021001 au catalogue. Ottawa (Ontario). Version mise à jour le 30 novembre 2022 (site consulté le 6 janvier 2023).
- Statistique Canada. 2023. *Evaluating confidence interval methods for the 2021 Census using the Census Monte Carlo simulation environment*. Document de travail interne. Ottawa, Canada.
- Wilson, E.B. 1927. « Probable Inference, the Law of Succession, and Statistical Inference ». *Journal of the American Statistical Association*, 22(158) : p. 209 à 212.