

N° 71-526-X au catalogue
ISBN 978-0-660-24069-5

Méthodologie de l'Enquête sur la population active du Canada

Date de diffusion : le 21 décembre 2017



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 STATCAN.infostats-infostats.STATCAN@canada.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 |

Programme des services de dépôt

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

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « 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

© Ministre de l'Industrie, 2017

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.

Remerciements

Nous remercions sincèrement les nombreuses personnes qui ont contribué de différentes façons à ce document.

Nous sommes particulièrement reconnaissants envers les principaux auteurs des divers chapitres : Lihua An, Justin Francis, Guy Laflamme, Yves Lafortune, Scott Meyer, Elisabeth Neusy, Steven Thomas et Sylvia White. Ces personnes ayant travaillé à différents aspects du remaniement de 2015 de la méthodologie de l'Enquête sur la population active étaient les mieux placées pour écrire à ce sujet.

Par ailleurs, de nombreuses personnes à la Division des méthodes d'enquêtes auprès des ménages, à la Division de la statistique du travail et ailleurs à Statistique Canada ont également pris part à la planification, à l'examen, à la vérification, à la révision, à la traduction et à la production de ce document. Nous les remercions toutes sincèrement.

Table des matières

Remerciements	3
Chapitre 1 Introduction et aperçu général	5
Chapitre 2 Plan de sondage	8
Chapitre 3 Création et tenue à jour de la base de logements	27
Chapitre 4 Collecte	34
Chapitre 5 Traitement et imputation	38
Chapitre 6 Pondération et estimation	44
Chapitre 7 Estimation de la variance	53
Chapitre 8 Qualité des données	59
Chapitre 9 Utilisation de la base de sondage ou de l'échantillon de l'EPA pour d'autres enquêtes	69
Références	72
Annexe A.1 Glossaire	74
Annexe A.2 Abréviations	80
Annexe B Caractéristiques de la base de sondage et du plan de sondage	82
Annexe C Plan de sondage de l'Enquête sur la population active	91
Annexe D Cartes des UPE (diagrammes des grappes F01)	93
Annexe E Cartes provinciales	97
Annexe F Définition des variables utilisées pour former les classes d'imputation	113
Annexe G Variables auxiliaires composites	116

Méthodologie de l'Enquête sur la population active du Canada

Chapitre 1 Introduction et aperçu général

1.0 Introduction

La présente publication se veut un ouvrage de référence sur la méthodologie de l'Enquête sur la population active (EPA). Cette publication se concentre sur la méthodologie utilisée pour les 10 provinces du Canada, même si l'EPA vise également les trois territoires. Elle décrit toutes les étapes actuelles de l'enquête et souligne les changements apportés pendant le dernier remaniement du plan de sondage.

Un autre document, appelé *Guide de l'Enquête sur la population active* (n° au catalogue 71-543-G, disponible en ligne), sert de complément au présent rapport et décrit les concepts, les définitions et les données produites dans le cadre de l'EPA.

1.1 Contexte

L'EPA a été créée après la Seconde Guerre mondiale pour répondre à un besoin urgent de données fiables et actuelles sur le marché du travail reflétant le passage d'une économie de guerre à une économie de paix. L'enquête a été conçue pour produire des estimations sur l'emploi et le chômage tant à l'échelle régionale que nationale.

Effectuée sur une base trimestrielle à son origine en 1945, l'EPA est devenue une enquête mensuelle en 1952. En 1960, le Comité interministériel de la statistique du chômage recommandait que l'EPA devienne l'instrument officiel de mesure du chômage au Canada. Après l'adoption de cette recommandation, la demande de données a augmenté, les utilisateurs voulant disposer d'une plus vaste gamme de statistiques sur le marché du travail, notamment des données régionales plus détaillées. L'éventail des estimations produites de cette enquête s'est élargi considérablement au fil des ans et présente aujourd'hui un portrait détaillé du marché du travail canadien.

1.2 Concepts et produits de l'EPA

L'EPA est la source officielle d'estimations mensuelles touchant l'emploi total et le chômage. Parmi les principaux indicateurs mensuels publiés, signalons le taux de chômage, le taux d'emploi et le taux d'activité. Il s'agit également d'une des principales sources d'information sur les caractéristiques sociodémographiques de la population en âge de travailler (notamment l'âge, l'état matrimonial, le niveau de scolarité et la situation familiale).

Les estimations de l'emploi sont produites à divers niveaux, notamment selon le sexe, le groupe d'âge, le secteur d'activité, la profession, le plus haut niveau de scolarité et le statut d'immigrant. Des statistiques sont également produites sur des caractéristiques telles que l'ancienneté, le nombre d'heures habituellement et effectivement travaillées et les salaires des employés. Les questions posées par l'enquête permettent d'étudier une grande variété de sujets d'actualité, comme le travail à temps partiel involontaire, le cumul d'emplois et l'absentéisme au travail.

Les estimations sur le chômage sont produites par secteur d'activité et par profession, ainsi que selon la durée du chômage, le genre de travail recherché et l'activité avant la recherche de travail. Des mesures supplémentaires du chômage sont également produites annuellement pour mettre en lumière le degré de ralentissement du marché du travail et le degré de difficulté associé au chômage. Il existe également des données sur l'activité récente sur le marché du travail des personnes actuellement inactives. Le *Guide de l'Enquête sur la population active* présente une description complète du contenu du questionnaire de l'EPA.

En plus des estimations nationales et provinciales, l'EPA produit également des données pour des régions infra-provinciales, comme les régions économiques de l'assurance-emploi (REAE) et les régions métropolitaines de recensement (RMR). Le gouvernement fédéral et les gouvernements provinciaux se servent des données de l'EPA pour la répartition des ressources financières et autres entre les différents secteurs de compétence politiques et administratifs.

Les estimations normalisées de l'EPA paraissent chaque mois dans la publication *Information sur la population active* (numéro 71-001-X au catalogue et disponible sur Internet). On peut également accéder à une variété

de données propres au marché du travail par l'intermédiaire de CANSIM, la base de données et le système d'extraction électronique de Statistique Canada. On compte plus de 100 tableaux CANSIM représentant plusieurs milliers de séries chronologiques qui sont mises à jour tous les mois ou tous les ans pour tenir compte des nouvelles données de l'EPA.

L'EPA peut produire beaucoup plus d'information que ce qui est publié périodiquement. Des totalisations spéciales peuvent être produites en recouvrement des coûts. Pour plus de détails concernant les produits et services rattachés à l'enquête, veuillez consulter la section 9 du *Guide de l'Enquête sur la population active*.

1.3 Aperçu général de l'enquête et structure du document

Dans les provinces, l'EPA est une enquête mensuelle réalisée auprès des ménages. Son échantillon est représentatif de la population canadienne civile non institutionnalisée de 15 ans et plus. Les personnes suivantes sont exclues de l'enquête : personnes vivant dans les réserves ou autres établissements autochtones, pensionnaires d'un établissement institutionnel, membres à temps plein des Forces canadiennes et résidents de certaines régions extrêmement éloignées ou à très faible densité de population. Collectivement, ces personnes exclues de l'enquête représentent environ 2 % de la population de 15 ans et plus.

Ces groupes sont exclus de la population cible, car la collecte de leurs données soulèverait des problèmes opérationnels. À titre d'exemple, plusieurs membres des Forces canadiennes vivent à des endroits inaccessibles aux intervieweurs de l'EPA (notamment à bord de navires de guerre ou dans des camps militaires et des casernes). Il serait donc difficile d'interviewer ces personnes. Les pensionnaires d'établissements institutionnels sont également exclus, car la plupart d'entre eux sont incapables de participer au marché du travail.

L'enquête utilise un plan de sondage à deux degrés. Au premier degré, on sélectionne un échantillon de régions géographiques. Ces régions sont appelées unités primaires d'échantillonnage (UPE). Dans chacune des UPE sélectionnées, on choisit un échantillon de logements au second degré. Les ménages sont identifiés dans les logements sélectionnés, et toutes les personnes du ménage qui font partie de la population cible sont sélectionnées pour participer à l'enquête. Les logements sélectionnés demeurent dans l'échantillon pour une période de six mois. Les logements sortants sont remplacés par des logements de la même UPE, ou d'une UPE semblable si l'UPE précédente a été éliminée et remplacée. Il en résulte un chevauchement des cinq sixièmes de l'échantillon d'un mois à l'autre, ce qui procure un plan de sondage efficace pour estimer les variations d'un mois à l'autre. Le renouvellement des logements après six mois évite d'imposer un fardeau trop lourd aux répondants des ménages sélectionnés aux fins de l'enquête. La forte proportion d'UPE en commun entre les échantillons à 12 mois d'intervalle procure un plan de sondage efficace pour estimer changements d'une année à l'autre. Les chapitres 2 et 3 fournissent plus de détails au sujet du plan de sondage.

La collecte des données de l'EPA a lieu pendant la semaine suivant la semaine de référence. Normalement, la semaine de référence comprend le quinzième jour du mois. En 2015, environ 88 % des ménages échantillonnés ont répondu au questionnaire de l'EPA chaque mois. L'interview de l'EPA est obligatoire et dure huit minutes en moyenne. La collecte des données est effectuée grâce à un système d'interview assistée par ordinateur. Plusieurs méthodes de collecte sont utilisées, y compris les interviews sur place et les interviews téléphoniques, ainsi qu'un questionnaire électronique. Les détails concernant la stratégie de collecte sont présentés au chapitre 4.

Dans les jours suivant la collecte, on effectue le traitement, l'imputation et la pondération et on dérive des indicateurs de la qualité. Ces étapes sont décrites aux chapitres 5, 6, 7 et 8. Malgré le gros volume de données à traiter chaque mois, Statistique Canada publie les estimations de l'EPA 10 jours seulement après la fin de la collecte.

L'échantillon et la base de sondage de l'EPA sont également utilisés pour un grand nombre des autres enquêtes sociales de Statistique Canada. Ce point est abordé au chapitre 9. Plusieurs annexes, dont chacune traite d'un sujet particulier, ainsi que d'autres documents d'enquête, sont inclus à la fin de la publication.

1.4 Changements apportés en 2015

Tous les 10 ans, après le recensement décennal de la population, l'EPA subit un remaniement de son échantillon pour tenir compte de l'évolution des caractéristiques de la population et du marché du travail, pour s'adapter aux besoins actuels et prévus des utilisateurs de données et pour mettre à jour l'information géographique requise pour mener l'enquête.

Le dernier plan de sondage a été introduit progressivement en janvier 2015 et entièrement mis en place en juin 2015.

Le remaniement de 2015 a apporté un certain nombre de changements importants à la méthodologie de l'enquête. Ces changements visaient à réduire les coûts d'enquête, à utiliser des méthodes de collecte à jour et à permettre aux utilisateurs de données d'effectuer des analyses à la fine pointe du progrès.

Dans cette refonte de l'enquête, les unités primaires d'échantillonnage ont été construites à partir des aires de diffusion définies pour le Recensement de 2011. En plus de rationaliser les travaux en cause pour le remaniement de l'échantillon, ce changement normalise la géographie de l'EPA, ce qui facilite la comparaison des estimations entre enquêtes et l'analyse faisant intervenir une modélisation multiniveau. La stratégie de répartition de l'échantillon a été modifiée pour utiliser des objectifs de qualité qui empêchent les algorithmes d'allocation d'augmenter automatiquement les tailles des échantillons dans les régions affichant un faible taux de chômage. Les modifications apportées aux UPE, à l'allocation et à la stratification sont décrites au chapitre 2.

Une innovation introduite au moment de la refonte de 2005, l'utilisation de listes d'adresses existantes, a été considérablement élargie dans le plan de sondage de 2015. Le Registre des adresses (RA) résidentielles de Statistique Canada a été intégré à un nouveau service de base de sondage pour les enquêtes auprès des ménages. Le Fichier de l'univers des logements (FUL) est un extrait d'adresses du RA qui est maintenant utilisé pour produire la liste d'adresses pour plus de 90 % des UPE dans l'échantillon de l'EPA, ce qui réduit le travail des intervieweurs sur place, qui auraient autrement dû créer la liste d'adresses en observant directement les quartiers/ UPE dans l'échantillon de l'EPA. Le service de base de sondage fournit également les numéros de téléphone qui aideront les intervieweurs à joindre les ménages échantillonnés. Pour plus d'information au sujet du Registre des adresses, voir les chapitres 2 et 3.

L'EPA a ajouté une troisième méthode de collecte en 2015 : les répondants admissibles peuvent maintenant remplir le questionnaire par Internet. Cette nouvelle stratégie est décrite au chapitre 4.

La stratégie d'imputation générale n'a pas changé, mais la liste des variables utilisées pour créer les groupes d'imputation pour l'imputation par donneur a été revue et mise à jour pour inclure le secteur d'activité. Ce changement, ainsi que d'autres changements apportés au processus de vérification et d'imputation, sont décrits au chapitre 5.

Enfin, un important changement a été apporté en 2015 à l'estimation de la variance. Depuis janvier 2015, la méthode bootstrap remplace la méthode du jackknife comme technique d'estimation de la variance pour l'EPA. Ainsi, les utilisateurs peuvent calculer et déclarer les estimations de la variance fondées sur le plan de sondage pour effectuer des analyses avancées par eux-mêmes. Cet important changement est présenté au chapitre 7.

1.5 Changements à venir au-delà de 2015

Un certain nombre de changements supplémentaires sont déjà au programme au cours des années à venir.

En janvier 2016, les systèmes de classification utilisés pour classer l'industrie et la profession en fonction des données de l'EPA seront mis à jour pour tenir compte des normes les plus récentes. Plus précisément, le Système de classification des industries de l'Amérique du Nord de 2007 (SCIAN 2007) utilisé à l'heure actuelle sera mis à jour conformément à la norme du SCIAN 2012, et la Classification nationale des professions pour statistiques de 2006 (CNP-S 2006) sera mise à jour pour tenir compte de la norme de la CNP 2011.

Une deuxième étape de la refonte de 2015 nécessitera également la mise à jour de tous les systèmes d'enquête (collecte, traitement, estimations et calculs) conformément aux normes actuelles. Ce changement sera apporté progressivement sur plusieurs années et devrait se terminer à la fin de l'exercice 2017-2018.

Chapitre 2 Plan de sondage

2.0 Introduction

Comme mentionné au chapitre 1, l'objectif de l'Enquête sur la population active (EPA) est de produire des données fiables et à jour sur l'emploi, le chômage et les caractéristiques de la population d'âge actif à divers niveaux géographiques. En principe, ces données pourraient être obtenues d'une source administrative, d'un recensement de la population ou d'une enquête-échantillon. En l'absence d'une source administrative pouvant produire les estimations requises à l'heure actuelle, et comme il est impossible de réaliser un recensement et de contacter toute la population en âge d'activité tous les mois afin de déterminer le statut d'emploi de chacun, un échantillon de la population est contacté, et les réponses de chaque personne sont utilisées pour produire des estimations mensuelles sur la population active.

Le plan de sondage se compose de toutes les étapes à suivre au moment de sélectionner un échantillon. Il influe sur la qualité des estimations produites et les coûts de l'enquête. Étant donné qu'une bonne part du budget d'une enquête est consacrée à la collecte des données, le plan de sondage s'efforce de réduire les frais de collecte tout en optimisant la qualité des données.

Le présent chapitre décrit les diverses stratégies employées par l'EPA pour atteindre cet objectif dans les 10 provinces. D'abord, la section 2.1 présente certains concepts de base de la théorie des sondages qui seront utilisés tout au long du chapitre. La section 2.2 décrit le plan de sondage général de l'EPA. La répartition de l'échantillon est décrite à la section 2.3. La section 2.4 décrit comment les grappes sont formées, et la section 2.5 décrit comment elles sont stratifiées. Enfin, la section 2.6 décrit le processus de sélection de l'échantillon et la méthodologie de rotation.

2.1 Quelques notions de base de la théorie des sondages

Cette section présente quelques notions requises pour comprendre la description du plan de sondage fournie aux prochaines sections. Pour plus de renseignements, un survol conceptuel de la théorie des sondages est disponible dans *Méthodes et pratiques d'enquête* (Statistique Canada (2003)). Des détails plus techniques se trouvent dans les nombreux ouvrages sur la théorie de l'échantillonnage (p. ex., Cochran 1977 ou Särndal, Swensson et Wretman 1992).

Les données recueillies dans le cadre d'une enquête-échantillon servent à produire des estimations pour la *population cible* – le groupe ou la population d'intérêt. La sélection d'un échantillon nécessite une *base de sondage*, qui devrait correspondre le plus possible à la population cible, bien que des contraintes pratiques puissent l'en empêcher. L'EPA sélectionne un échantillon *probabiliste*, c.-à-d. un sous-ensemble de la population, où les unités sondées sont sélectionnées au hasard. Les *estimations* pour la population sont calculées selon l'information fournie à partir de cet échantillon.

Les estimations peuvent varier selon les personnes sélectionnées dans l'échantillon. De plus, l'estimation produite à partir d'un échantillon diffère de l'estimation produite si la population au complet était interviewée. Ces types de différences sont appelées *erreurs d'échantillonnage*. Les résultats de l'enquête comportent également d'autres erreurs non liées au plan de sondage, appelées *erreurs non dues à l'échantillonnage*.

Deux importantes mesures de l'erreur d'échantillonnage sont le *biais* et la *variance d'échantillonnage*. Supposons qu'il soit possible de sélectionner plusieurs échantillons selon un plan de sondage donné. Pour chacun de ces échantillons, on peut produire une estimation de la caractéristique d'intérêt (par exemple, le nombre de chômeurs, la moyenne du nombre d'heures travaillées) à partir des données observées. Le biais est la différence entre la moyenne des estimations produites à partir de tous les échantillons possibles et la valeur réelle correspondante pour la population totale. La variabilité entre les estimations de l'échantillon, ou le degré de différence entre elles, s'appelle la variance d'échantillonnage.

Le biais peut être causé par différentes sources, comme une base de sondage imparfaite, la méthode utilisée pour produire l'estimation ou la non-réponse à l'enquête. Le plan de sondage peut ajouter un biais lorsque certaines régions sont exclues du champ de l'enquête (p. ex., en raison des frais de collecte élevés). Cette composante de l'erreur peut être difficile à mesurer en pratique, parce que la valeur réelle pour la population est généralement inconnue.

La variance d'échantillonnage mesure la dispersion entre les estimations produites à partir de tous les échantillons possibles. Elle indique le degré de précision d'une estimation: plus la variance d'échantillonnage est petite, meilleure est la précision associée à l'estimation. La variance d'échantillonnage peut être estimée à partir d'un seul échantillon observé, même si elle indique la variabilité entre de nombreux échantillons théoriques.

D'autres mesures de la variabilité sont obtenues à partir de la variance d'échantillonnage. On obtient l'*erreur type* en prenant la racine carrée de la variance d'échantillonnage et on l'utilise souvent pour déterminer un *intervalle de confiance* ou pour effectuer un *test statistique*. L'erreur type est une *mesure absolue* de la variation, puisqu'elle est mesurée dans les mêmes unités que l'estimation. Une autre mesure est le *coefficient de variation (CV)*, qui se définit par l'erreur type divisée par l'estimation. Le CV est une *mesure relative*, puisqu'il ne comporte pas d'unités et est calculé par rapport à l'estimation. Une troisième mesure est l'*effet de plan*, une mesure relative que l'on calcule en divisant la variance d'échantillonnage d'une estimation obtenue en vertu du plan d'enquête par la variance d'échantillonnage d'un échantillon aléatoire simple (EAS) de la même taille. Cette mesure peut être utilisée pour comparer l'efficacité d'un plan de sondage par rapport à un autre. Plus l'erreur type, la longueur de l'intervalle de confiance, le CV ou l'effet de plan sont petits, plus l'estimation est précise.

Le principal objectif d'un plan de sondage efficace consiste à réduire la variance d'échantillonnage compte tenu du budget limité et des contraintes opérationnelles. Un plan de sondage plus efficace peut obtenir la même précision des estimations (mesurée par la variance d'échantillonnage) au moyen d'un échantillon plus petit qu'un autre plan de sondage moins efficace. De même, en supposant une taille totale de l'échantillon fixe, un plan de sondage plus efficace a une variance d'échantillonnage plus petite qu'un plan de sondage moins efficace.

Plusieurs facteurs influencent la variance d'échantillonnage d'une estimation. Les plus importants sont le nombre de personnes dans la population, le nombre de personnes dans l'échantillon, la méthode d'échantillonnage utilisée pour tirer l'échantillon, le taux de réponse et l'homogénéité de la caractéristique d'intérêt dans la population. La taille de la population ne peut pas être contrôlée. Les taux de réponse peuvent parfois être modifiés par le choix de la méthode de collecte des données, mais habituellement pas par le plan de sondage. Toutefois, en contrôlant le nombre de personnes dans l'échantillon, la méthode d'échantillonnage utilisée pour tirer l'échantillon et l'homogénéité dans les groupes échantillonnés, on peut obtenir un plan de sondage plus efficace.

2.2 Vue d'ensemble du plan de sondage

L'EPA utilise un plan de sondage complexe. Une méthode d'échantillonnage plus efficace serait l'échantillonnage aléatoire simple (EAS), où les unités sont sélectionnées au hasard à partir d'une liste aux probabilités égales. Cependant, un échantillon aléatoire simple de personnes nécessite une liste de tous les membres de la population cible, ce qui peut être difficile à obtenir en pratique. De plus, les contraintes opérationnelles pourraient empêcher la faisabilité d'un plan de sondage en fonction de l'EAS, en exigeant l'utilisation d'une méthode d'échantillonnage plus complexe.

Pour la plupart des estimations de l'EPA, la population cible est constituée de toutes les personnes au Canada de 15 ans et plus. Il est impossible de sélectionner directement un échantillon de ces personnes pour une interview, puisqu'il n'existe pas de liste complète et à jour des personnes résidant dans les 10 provinces. Au lieu de sélectionner les personnes directement, il est plus facile de sélectionner des logements, puis de déterminer et d'interviewer les personnes vivant dans les logements sélectionnés. Bien qu'une liste relativement complète et régulièrement mise à jour des adresses des logements soit maintenant disponible (voir la section 2.2.1), la sélection de logements par échantillonnage aléatoire simple donnerait lieu à un échantillon qui serait trop dispersé géographiquement. Par conséquent, les frais de déplacement engagés pour la collecte en personne pourraient être excessifs.

Pour réduire les dépenses de déplacements, l'échantillon des logements est tiré en deux étapes de sélection consécutives. Cette méthode porte le nom d'échantillonnage à deux degrés. Au premier degré, les provinces sont divisées en régions géographiques appelées grappes ou unités primaires d'échantillonnage (UPE). Une sélection aléatoire de ces UPE compose l'échantillon du premier degré. Au deuxième degré, pour chaque UPE sélectionnée, une liste des logements dans la région est établie à partir d'une extraction du Registre des adresses (le Fichier de l'univers des logements (section 2.2.1)) ou au moyen d'un listage sur le terrain. Un échantillon de logements du deuxième degré est sélectionné à partir de ces listes. Les logements sont les unités secondaires d'échantillonnage (USE). Tous les résidents qui font partie de la population cible occupant les logements sélectionnés dans les

grappes sélectionnées constituent l'échantillon de personnes de l'EPA. Cette méthode de sélection à deux degrés est plus complexe mais réduit la dispersion géographique des personnes échantillonnées en les regroupant en grappes, ce qui réduit les coûts.

En janvier 2015, le plan de sondage à l'Île-du-Prince-Édouard (Î.-P.-É.) a été modifié et on a adopté l'échantillonnage à un degré. Autrement dit, les logements sont sélectionnés directement à partir d'une liste, sans aucune grappe. La section 2.5.6 contient de plus amples renseignements au sujet du plan de sondage à un degré à l'Î.-P.-É.

En plus des estimations mensuelles décrites à la section 1.2, l'EPA produit également des estimations du changement entre deux périodes de référence données. En vue d'améliorer la qualité de ces estimations, il est préférable d'accroître le chevauchement entre les échantillons de ces deux périodes. On ne peut augmenter ce recouvrement qu'en conservant les mêmes logements dans l'échantillon durant plusieurs mois. Malheureusement, l'accroissement du recouvrement s'accompagne d'une augmentation du fardeau imposé aux répondants, qui doivent participer à l'enquête à plusieurs reprises. Cette augmentation du fardeau pourrait entraîner une réduction du taux de réponse. Par ailleurs, en plus d'améliorer la qualité, un chevauchement accru réduit également les coûts de la collecte des données d'enquête, puisqu'il est moins coûteux d'obtenir une réponse au cours des mois subséquents qu'au premier mois. Par conséquent, le chevauchement entre les échantillons constitue un compromis entre la qualité des estimations du changement et le coût des opérations d'enquête par rapport au fardeau imposé aux répondants.

On a décidé de garder chaque logement dans l'échantillon de l'EPA pendant six mois consécutifs. Sous réserve de cette contrainte, le recouvrement maximum de l'échantillon entre deux mois consécutifs est de cinq sixièmes. Il est donc nécessaire de remplacer le sixième de l'échantillon de logements chaque mois. Pour mettre en œuvre cette stratégie, la population des UPE de l'EPA est scindée en six groupes de renouvellement¹, un échantillon sélectionné dans chaque groupe représentant la population totale. Le premier groupe de renouvellement est initialement contacté en janvier. Ces logements demeurent alors dans l'échantillon jusqu'en juin inclusivement. En juillet, tous les logements du groupe de renouvellement¹ sont remplacés par un nouvel échantillon de logements du même groupe de renouvellement. Le deuxième groupe de renouvellement se compose des logements initialement sondés de février à juillet inclusivement, et ainsi de suite pour les autres groupes de renouvellement. Le modèle de renouvellement est illustré à la figure 2.1 à la fin du chapitre. Pour plus d'information au sujet du renouvellement de l'échantillon de logements, voir la section 2.6.4.

La stratégie des groupes de renouvellement qui se recoupent comporte certains avantages. D'abord, elle permet l'utilisation de méthodes de traitement et d'estimation plus efficaces (décrites aux chapitres 5 et 6). Elle donne également lieu à une méthode simple pour sélectionner un sous-ensemble de l'échantillon de l'EPA pour d'autres enquêtes de Statistique Canada. Étant donné que chaque groupe de renouvellement représente la population au complet, il est facile de construire l'échantillon pour une autre enquête en regroupant les logements d'un nombre approprié de groupes de renouvellement. De l'information sur l'utilisation de la base de sondage et de l'échantillon de l'EPA pour d'autres enquêtes-ménages se trouve au chapitre 9.

2.2.1 Le Fichier de l'univers des logements et son incidence sur le plan de sondage

Pour le plan de sondage, il est important de savoir à peu près combien de logements et combien de logements occupés (c.-à-d. les logements qui correspondent à des ménages de personnes) se trouvent dans la population de l'EPA. Les chiffres sont utilisés pour la création des UPE, pour la répartition de l'échantillon et pour la stratification. Pour les plans de sondage précédents de l'EPA, ces chiffres provenaient du dernier recensement. Cependant, les derniers chiffres du recensement remontent à mai 2011, alors que ce remaniement a commencé graduellement en janvier 2015.

Pour avoir un dénombrement plus à jour du nombre total de logements dans la population, le Fichier de l'univers des logements (FUL) a été utilisé pour ce plan de sondage. Le FUL est une extraction de la base de données du Registre des adresses (RA) de Statistique Canada, qui renferme les adresses résidentielles (logements). Il est mis à jour tous les trimestres, au moyen des derniers fichiers administratifs disponibles et des résultats du listage et de la vérification sur le terrain.

1. Ils peuvent également être appelés panels de renouvellement. On parle habituellement de plan d'enquête par panel de renouvellement.

Afin de planifier le remaniement, l'extrait de juin 2013 du FUL a été utilisé. Étant donné que le FUL n'indiquait pas les logements occupés, le nombre a dû être estimé. Pour ce faire, on a multiplié le nombre total de logements dans le FUL par le taux d'occupation des logements du Recensement de 2011.

2.3 Répartition de l'échantillon

Comme le décrit le chapitre 1, l'EPA est la source officielle des estimations mensuelles totales de l'emploi et du chômage. L'EPA est également l'une des principales sources d'information sur les caractéristiques sociodémographiques de la population d'âge actif telles que l'âge, l'état matrimonial, le niveau de scolarité et la situation familiale.

L'EPA produit des données pour différentes régions géographiques, y compris à l'échelle nationale, provinciale et infraprovinciale, comme les régions économiques (RE), les régions métropolitaines de recensement (RMR) et les régions économiques d'assurance-emploi (REAE). L'étape de la répartition de l'échantillon précise le nombre ciblé de ménages à sélectionner dans chacune de ces régions². Elle vise à faire en sorte que l'échantillon puisse produire des estimations qui répondent à différents objectifs de l'EPA en ce qui concerne la précision. Il s'agit d'une étape importante, puisque les étapes subséquentes en dépendent, et elle fait en sorte que les ressources de l'enquête soient utilisées efficacement. Pour plus d'information sur les régions géographiques liées au recensement utilisées par l'EPA, voir le Dictionnaire du recensement (Statistique Canada (2012)).

Comme on l'a expliqué à la section 2.1, le nombre d'unités échantillonnées a une incidence directe sur la qualité des estimations produites par l'enquête. Étant donné que la taille de l'échantillon totale est fixe, un excédent d'échantillon attribué à une région donnée produira des estimations pour cette région qui sont de meilleure qualité que ne l'exigent les objectifs de l'enquête, au détriment de la qualité des données dans les autres régions. L'EPA produit des estimations à divers niveaux géographiques (Canada, provinces, régions économiques, etc.). Il est donc nécessaire de parvenir à un compromis approprié pour toutes ces estimations au moment de répartir l'échantillon en fonction d'un budget limité.

Afin de réaliser les objectifs de l'enquête et de maintenir l'efficacité générale du plan d'enquête, l'échantillon de l'EPA est réparti en deux étapes. À la première étape, l'échantillon financé par Statistique Canada est attribué. À la deuxième étape, un échantillon supplémentaire financé par Emploi et Développement social Canada (EDSC) est ajouté. Cette approche à deux étapes est basée sur l'hypothèse selon laquelle le budget de l'EPA de Statistique Canada est assuré sur une longue période, mais le financement d'EDSC pourrait fluctuer au fil du temps. Par conséquent, chaque partie de l'échantillon devrait être attribuée séparément pour répondre aux objectifs appropriés pour lesquels l'échantillon est financé. Le tableau B.4 à l'annexe B fournit la répartition de l'échantillon de l'EPA en fonction de différentes unités géographiques.

2.3.1 Répartition de l'échantillon financé par Statistique Canada

La première étape consiste à répartir l'échantillon financé par Statistique Canada (36 000 ménages) parmi les 10 provinces. Statistique Canada a établi les objectifs de qualité de l'EPA pour les provinces, régions économiques (RE) et régions métropolitaines de recensement (RMR). Tous les objectifs sont basés sur les estimations du nombre de chômeurs. En effet, le chômage est un enjeu revêtant un vif intérêt, et qui nécessite plus de ressources que l'emploi, parce qu'il est plus rare que ce dernier, afin de mesurer au même niveau de qualité pour ce qui est du CV que ne le ferait l'emploi. La répartition a pour objet de faire en sorte que l'échantillon soit en mesure d'atteindre ces objectifs.

Pour apporter des rajustements à la répartition de l'échantillon du plan de sondage précédent, il a fallu prédire la précision des estimations du nombre de chômeurs (mensuelles et moyenne mobile sur trois mois) pour chaque province, RE et RMR pour une taille d'échantillon donnée. Ces prédictions étaient basées sur un modèle d'estimation de CV mettant en cause la taille de l'échantillon, l'estimation du nombre de chômeurs et la variance estimative de cette estimation. Ce modèle est basé sur des données de 80 mois précédents de l'EPA. Par conséquent, on présume implicitement que le nouveau plan de sondage aura une efficacité comparable au

2. L'EPA porte sur un échantillon de logements, et non pas de ménages (logements occupés). Toutefois, la répartition est exprimée en fonction d'un nombre ciblé de ménages. Ce nombre est utilisé pour déterminer le taux d'échantillonnage des ménages dans une région. Lorsque ce taux d'échantillonnage est appliqué à une liste de logements, le nombre de logements échantillonnés devrait en moyenne donner environ le nombre ciblé de ménages. La précision dépendra de la précision du taux d'occupation et de la couverture du FUL.

précédent. On suppose également que les taux de réponse, les taux d'inoccupation et le nombre d'adultes par ménage demeureront constants au fil du temps. Les hypothèses du modèle ont été validées par l'analyse des tendances de l'EPA au cours des neuf dernières années. À partir de ce modèle, on a pu prédire l'incidence des modifications de la répartition sur la qualité des estimations du chômage à venir.

La stratégie de répartition de l'échantillon financé par Statistique Canada est fondée sur les critères suivants :

- Pour chaque province, le CV de l'estimation mensuelle du nombre de chômeurs devrait être inférieur à 7 %;³
- Pour chaque RE, le CV de l'estimation de la moyenne mobile sur trois mois du nombre de chômeurs devrait être inférieur à 25 %;
- La taille minimale de l'échantillon pour chaque région économique est fixée à 200;
- Pour cinq RMR et Lethbridge⁴, le CV de l'estimation de la moyenne mobile sur trois mois du nombre de chômeurs devrait être inférieur à 25 %;

Dans les plans précédents, on s'attendait à ce que les CV soient au-dessous de l'objectif pour tous les mois. Cependant, dans les faits, ce n'est pas le cas. Dans un plan sur 10 ans, certains mois peuvent être des valeurs aberrantes pour certains domaines, ce qui est difficile à prédire. La répartition d'un échantillon plus gros pour neutraliser les valeurs aberrantes peut s'avérer difficile, en plus de risquer de gaspiller les ressources limitées, qu'il vaudrait mieux allouer à une autre région. Pour le remaniement, l'échantillon a été attribué de manière à ce que les objectifs de CV soient atteints pour au moins 90 % des estimations mensuelles au fil du temps.

Au moyen du modèle d'estimation du CV et de la programmation non linéaire, les 36 000 ménages ont été attribués aux provinces, aux RE et aux RMR pour répondre aux contraintes qui précèdent, tout en réduisant au minimum la variance de l'estimation mensuelle nationale du nombre de chômeurs. D'autres détails suivent sur la répartition aux provinces, aux RE et aux RMR.

Répartition aux 10 provinces

Pour les provinces, le plan précédent visait à avoir des CV inférieurs à 7 % pour les estimations mensuelles du nombre de chômeurs. Dans les faits, ce ne fut pas le cas pour plusieurs provinces. Les simulations ont démontré qu'il ne serait pas possible d'atteindre un objectif de CV de 7 % pour toutes les provinces et tous les mois en utilisant uniquement les 36 000 ménages échantillonnés financés par Statistique Canada. Il était impératif que l'échantillon financé par Statistique Canada puisse atteindre les objectifs de qualité annoncés. Étant donné que la taille de l'échantillon était fixe, la seule solution était de modifier les objectifs.

Comme susmentionné, le CV est une mesure de la qualité relative. Pour les niveaux de chômage très faibles, les CV ont tendance à être plus élevés. Par ailleurs, l'erreur type est une mesure de la qualité absolue, en ce qu'elle est mesurée dans les mêmes unités que l'estimation. Pour une taille fixe de l'échantillon, les CV augmentent à mesure que le taux de chômage diminue, même si l'erreur type demeure inchangée. À cause d'un faible taux de chômage dans les Prairies au cours des dernières années, les CV étaient plus élevés que l'objectif de 7 %, même si les erreurs types étaient comparables aux mois affichant un taux de chômage plus élevé et des CV plus faibles. On a décidé d'adopter deux procédures de répartition : répartir aux provinces en fonction des CV lorsque le taux de chômage se situe au-dessus de 5 %; répartir en fonction d'une taille d'erreur type comparable lorsque le taux de chômage est inférieur à 5 %.

L'Î.-P.-É. faisait exception à cette règle. En effet, la taille de l'échantillon à l'Î.-P.-É. est restée au même niveau que pour le plan précédent. Une augmentation de la taille de l'échantillon a été évitée parce qu'une augmentation supplémentaire aurait entraîné la résélection des logements au cours de la durée prévue de 10 ans du nouveau plan, ce qui imposerait un fardeau considérable aux répondants.

Répartition aux régions économiques

Le CV ciblé pour les RE est demeuré inchangé par rapport aux plans précédents de l'EPA. Cependant, certaines modifications ont été apportées aux régions géographiques ciblées, puisqu'il est difficile de veiller à atteindre ces

3. On observe une légère modification de l'objectif de qualité lorsque le taux de chômage est faible, un phénomène expliqué plus loin.

4. Lethbridge devrait devenir une RMR suite au Recensement de 2016.

objectifs pour toutes les RE. Les RE qui ont un petit nombre de ménages sont combinées et l'objectif de précision est appliqué à la RE combinée. Quatre groupes de RE ont été combinés au dernier remaniement. Ils se situaient dans les régions nordiques du Québec, du Manitoba, de la Saskatchewan et de la Colombie-Britannique. Pour ce remaniement, trois autres paires de petites RE rurales ont été combinées à Terre-Neuve-et-Labrador, au Manitoba et en Alberta.

Répartition aux régions métropolitaines de recensement

Il a été envisagé que l'échantillon de Statistique Canada soutienne les objectifs de CV pour toutes les régions métropolitaines de recensement (RMR). Cependant, les simulations ont démontré que l'imposition de tels objectifs mine la précision d'estimations provinciales et nationales plus importantes. Dans la plupart des cas, les régions métropolitaines de recensement correspondent à des régions économiques de l'assurance-emploi (REAE). L'échantillon supplémentaire d'EDSC alloué pour les estimations des REAE (voir la section 2.3.2) fournit assez d'échantillon pour avoir une qualité suffisante pour la plupart des estimations des RMR. Deux RMR (Moncton et Saint John) collectivement forment une seule REAE, et chacune a une qualité suffisante en raison de l'échantillon d'EDSC.

Il y a six régions qui ne correspondent pas bien aux REAE et où les exigences relatives au CV ne sont pas respectées même avec l'échantillon supplémentaire d'EDSC. Cinq de ces régions sont les RMR de Peterborough, Barrie, Brantford, Guelph et Kelowna. La sixième région est Lethbridge. Après chaque recensement, Statistique Canada passe en revue la liste des RMR. Comme Lethbridge devrait devenir une RMR après le Recensement de 2016, il est important de veiller à ce que l'échantillon tiré à Lethbridge soit suffisant pour produire de bonnes estimations après 2016. Une taille d'échantillon plus grande a été attribuée à ces six régions, afin qu'elles aient un échantillon suffisant pour que les CV de l'estimation de la moyenne mobile sur trois mois du nombre de chômeurs soient inférieurs à 25 %.

2.3.2 Répartition de l'échantillon financé par EDSC

Cette étape de la répartition de l'échantillon de l'EPA consiste à ajouter l'échantillon financé par EDSC à l'échantillon de base financé par Statistique Canada. Les estimations du taux de chômage selon l'EPA pour les REAE sont utilisées par EDSC pour établir les critères d'admissibilité à l'assurance-emploi et la durée des prestations dans chaque région. Pour aider à améliorer la précision des estimations du taux de chômage pour les REAE, EDSC paye pour obtenir un échantillon supplémentaire de 16 600 ménages.

Les critères d'admissibilité et les prestations d'EDSC sont déterminés en fonction de fourchettes de taux de chômage. EDSC a besoin d'une précision comparable des estimations dans toutes les REAE, sauf lorsque les taux de chômage d'une région se situent dans les fourchettes supérieure et inférieure. La fourchette inférieure est de 6 % et moins; par conséquent, dans les régions affichant un faible taux de chômage, EDSC n'a qu'à être en mesure de déterminer que le taux de chômage est inférieur à 6 %. De même, la fourchette supérieure dépasse 13 %; par conséquent, dans les régions affichant un taux de chômage très élevé, EDSC n'a qu'à être en mesure de déterminer que le taux de chômage dépasse 13 %.

Avant de répartir l'échantillon financé par EDSC, le nombre d'unités attribuées à chaque REAE par l'échantillon financé par Statistique Canada a dû être déterminé. D'abord, l'échantillon dans une RE ou une RMR a été réparti proportionnellement aux intersections RE-REAE-RMR en fonction de la taille de chaque intersection. En additionnant les intersections, l'échantillon de Statistique Canada attribué à chaque REAE a été déterminé. Pour l'échantillon financé par EDSC, la taille de l'échantillon ciblé dans chaque REAE était fondée sur les critères suivants :

- Pour chaque REAE⁵, le CV de l'estimation du taux de chômage par moyenne mobile sur trois mois doit être inférieur à 15 %. Toutefois, dans les régions affichant un faible taux de chômage, EDSC a seulement besoin d'une qualité suffisante pour conclure que le taux de chômage est inférieur à 6 %. Par conséquent, pour les taux inférieurs à 4,8 %, la taille de l'échantillon nécessaire est telle que l'erreur type doit être suffisamment petite pour conclure que le taux de chômage est inférieur à 6 %. La valeur de 4,8 % a été retenue parce

5. La REAE du Nord du Manitoba est une exception. Malgré le fait que certains de ses CV sont au-dessus de 15 %, cette région a toujours eu un taux de chômage bien au-dessus de 13 %, ce qui fait qu'il était inutile d'améliorer la qualité pour déterminer les prestations d'assurance-emploi. Par conséquent, pour éviter des coûts de collecte très élevés pour le Nord du Manitoba, la taille de l'échantillon a été maintenue au même niveau que dans le plan précédent.

qu'elle représente la valeur maximale pour laquelle la limite supérieure de l'intervalle de confiance du taux de chômage sera de 6 % lorsque le CV est de 15 %.

- La taille minimale de l'échantillon pour chaque REAE est fixée à 500;
- La qualité des estimations produites pour chaque REAE doit être similaire d'une REAE à l'autre.

Encore une fois, la programmation non linéaire a été utilisée pour résoudre ce problème. Après avoir réparti l'échantillon financé par EDSC, la taille de l'échantillon total (Statistique Canada et EDSC) attribuée à chaque REAE a été attribuée aux intersections RE-REAE-RMR, proportionnellement à la taille de chaque intersection. Cette nouvelle répartition a alors été comparée à celle utilisée avant le remaniement pour détecter les erreurs potentielles dans le modèle utilisé et pour prédire l'efficacité du nouveau plan. Les changements de taille des échantillons régionaux et les CV prévus ont été utilisés pour évaluer le nouveau plan.

Après cette dernière étape de la répartition de l'échantillon, deux paramètres ont été produits : la fraction de sondage inverse (FSI) et le nombre de ménages échantillonnés requis pour chaque intersection. La fraction de sondage inverse correspond au nombre de ménages dans l'intersection, divisé par le nombre de ménages attribués à l'échantillon pour l'intersection. Elle sert à déterminer la taille et le nombre de strates du plan (voir la section 2.5) et pendant le processus de sélection de l'échantillon (voir la section 2.6).

2.4 Création des UPE

À la section 2.2, le plan de base était présenté, où l'EPA dispose d'une sélection de grappes comme unités primaires d'échantillonnage (UPE), suivies d'une sélection de logements comme unités secondaires d'échantillonnage (USE)⁶. La première étape du plan de sondage à deux degrés consiste à déterminer les limites géographiques des UPE utilisées pour la première étape de la sélection de l'échantillon en fonction de la taille, de la forme et d'autres facteurs.

La détermination de la taille et de la forme des UPE est un compromis entre les coûts de collecte et l'efficacité du plan de sondage. De la perspective des coûts, la collecte est plus abordable si la forme des UPE est géographiquement compacte et contigüe, ce qui réduit les délais de déplacement entre les logements échantillonnés dans l'UPE. Si les UPE sont trop vastes, il est trop coûteux pour un intervieweur de visiter tous les logements échantillonnés assez souvent pour obtenir des réponses. Par ailleurs, les petites UPE demeurent dans l'échantillon moins longtemps, ce qui augmente les coûts associés au renouvellement des UPE. De plus, lorsque les UPE sont petites, plusieurs UPE devront être sélectionnées. Les UPE sélectionnées seront généralement plus éloignées les unes des autres, ce qui augmentera encore une fois les frais de déplacement.

De la perspective du plan, l'EPA pourrait sélectionner soit un petit nombre d'UPE comportant chacune de nombreux logements, soit de nombreuses UPE comportant chacune un petit nombre de logements. Le dernier cas donne lieu à un plan d'enquête plus efficace⁷. Cependant, l'adoption d'un tel plan annule les avantages d'un plan avec grappes.

Pour déterminer la taille idéale d'une UPE compte tenu des facteurs qui précèdent, deux éléments sont nécessaires. Le premier est un outil pour évaluer la variance d'échantillonnage découlant de divers scénarios. Cet outil peut être construit au moyen des données du recensement. Le deuxième est un modèle relativement précis pour estimer les coûts de collecte pour différents scénarios de taille d'UPE et de nombre de logements sélectionnés par UPE. Pour construire ce modèle, des renseignements détaillés sur les coûts sont nécessaires. Compte tenu de la complexité et des gros changements apportés récemment à la stratégie de collecte (voir le chapitre 4), il s'est avéré pratiquement impossible de construire un modèle valide et à jour des coûts. Par conséquent, il n'a pas été possible de réévaluer la taille idéale des UPE. La taille idéale des UPE qui avait été utilisée lors des deux remaniements précédents (200 ménages) a été préservée, et cette taille a également servi d'objectif pour ce plan.

Une fois la taille cible des UPE établie, les UPE ont été construites à partir des unités géographiques types des aires de diffusion (AD) du Recensement de 2011. Cette approche comporte plusieurs avantages. L'utilisation des AD comme point de départ pour les UPE élimine la nécessité de créer de nouvelles définitions géographiques

6. L'Î.-P.-É., qui a un plan d'échantillonnage à un degré, n'a pas de grappes.

7. Si nous poussons le raisonnement jusqu'au bout, la solution idéale serait de créer des UPE renfermant chacune un logement. C'est l'équivalent d'un échantillonnage à un degré. Toutefois, comme mentionné à la section 2.2, l'échantillonnage à un degré est actuellement trop coûteux pour être mis en œuvre à l'extérieur de l'Î.-P.-É.

pour l'EPA, comme c'était le cas des plans précédents. On rationalise ainsi le processus de création des UPE et on réduit les frais associés au remaniement. Pour l'analyse, l'utilisation d'unités géographiques types facilite la comparaison d'estimations entre enquêtes, le couplage avec les données auxiliaires du recensement et d'autres sources et la modélisation multiniveau. De plus, étant donné que bien des enquêtes-ménages échantillonnent normalement pour des régions définies par des unités géographiques types, l'utilisation des AD comme UPE simplifie l'utilisation de la base de sondage de l'EPA par d'autres enquêtes-ménages, ce qui inclut la capacité de mettre à jour à mesure que de nouvelles AD sont définies, comme ce sera le cas pour le Recensement de 2016.

Malheureusement, certaines AD sont trop vastes, tandis que d'autres sont trop petites pour la contrainte idéale de 200 logements par UPE. Des simulations ont démontré qu'une grande variabilité de la taille de l'UPE augmenterait la variance de l'échantillonnage dans le cadre de la stratégie de sélection d'échantillon de l'EPA (voir la section 2.6). Une certaine variabilité était inévitable pour demeurer proche des unités géographiques types. Une fourchette acceptable de 100 à 600 ménages par UPE a été déterminée. Les AD au-dessous de cette fourchette ont été regroupées avec d'autres AD contiguës pour former de plus grandes UPE. Les AD au-dessus de cette fourchette ont été divisées en UPE contiguës plus petites et compactes au niveau des îlots de diffusion ou des côtés d'îlot du Recensement de 2011.

Une exception à cette règle a été faite à Toronto. Toronto renfermait de nombreuses AD comportant plus de 600 ménages, mais bien des AD étaient difficiles à diviser en UPE plus petites en raison de la présence d'immeubles d'appartements en hauteur dotés de plus de 600 logements. Il n'est pas pratique de diviser un seul immeuble en plus d'une UPE. D'abord, l'occupation des logements change fréquemment, ce qui rend difficile la correction en fonction du nombre de ménages dans chaque segment au moment de la division de l'immeuble. De plus, une fois qu'un intervieweur a établi un accès régulier à un immeuble d'appartements, il est tout à fait efficace de continuer la collecte dans les autres logements de l'immeuble. Autrement dit, cette grande UPE n'aurait pas les mêmes coûts élevés qu'une grande UPE composée de maisons isolées. Par conséquent, à Toronto, bien des UPE de 600 à 1000 ménages ont été créées. Pour éviter toute incidence négative sur l'efficacité du plan, ces UPE ont été regroupées en strates spéciales (voir la section 2.5.4).

Une fois toutes les UPE créées, une analyse détaillée a été effectuée pour relever celles qui étaient éloignées d'un centre urbain et aurait probablement un coût de collecte très élevé. Selon la situation, ces UPE ont été soit stratifiées séparément (voir la section 2.5), soit exclues de la base de sondage. Selon le plan précédent de l'EPA, moins de 1 % des ménages au Canada ont été exclus. Pour ce remaniement, environ 100 000 ménages supplémentaires ont été exclus dans les régions du Nord des 10 provinces, élevant le taux des exclusions à 1,5 %. Voir l'annexe B.1 pour plus de détails sur les zones exclues. L'exclusion de personnes appartenant à la population cible d'une enquête introduit intrinsèquement un biais dans les estimations de l'enquête; toutefois, le coût requis pour inclure ces régions était considéré trop élevé par rapport à l'incidence potentielle sur les estimations.

2.5 Stratification

La stratification est le processus par lequel la population est divisée en groupes homogènes absolument exclusifs appelés strates, afin d'améliorer l'efficacité du plan de sondage. Dans bien des enquêtes, les strates sont définies en fonction des domaines d'intérêt géographiques. Dans le cas de l'EPA, des strates sont formées à l'intérieur de chaque domaine d'intérêt : intersections RE-REAE-RMR. Cette stratification supplémentaire vise à s'assurer que l'enquête peut s'adapter aux contraintes de renouvellement, de répartition et de sélection qui sont décrites dans le présent chapitre.

La première étape consiste à déterminer combien de strates s'imposent dans un domaine. Une fois le nombre de strates déterminé, les strates peuvent être définies en fonction des contraintes géographiques, socio-économiques et d'efficacité, et les UPE peuvent être regroupées pour former ces strates. La stratification améliorera l'efficacité du plan de sondage si les UPE regroupées sont homogènes, ce qui veut dire que les ménages qui s'y trouvent ont des caractéristiques semblables. Une fois ce processus terminé, une base de sondage renfermant toutes les UPE et leurs strates correspondantes peut être créée.

2.5.1 Modifications apportées au cours de ce remaniement

Deux stratégies passées, qui avaient été auparavant intégrées à la méthodologie de stratification de l'EPA afin de réduire les coûts associés à la collecte et au listage en personne, ont été abolies pour ce plan de sondage.

D'abord, dans les régions urbaines isolées, un plan de sondage à trois degrés avait été utilisé, de façon à ce qu'à la première étape, une seule agglomération d'un groupe d'agglomérations n'aurait à être visitée à la fois. Ensuite, les UPE rurales qui étaient les plus onéreuses pour la collecte en personne (en raison du taux d'inoccupation élevé, de la distance des centres urbains ou de l'absence d'accès routier) avaient été stratifiées séparément et s'étaient vu attribuer un taux d'échantillonnage réduit pour diminuer le nombre de visites en personne nécessaires. Ces innovations ont réduit les coûts de collecte et de listage, tout en diminuant néanmoins l'efficacité du plan de sondage. De plus, ces innovations pouvaient transformer les chiffres de l'emploi dans certaines industries locales, en particulier lorsqu'une UPE donnée où de nombreux travailleurs étaient concentrés dans une industrie donnée était remplacée par une autre UPE dans une autre région dont l'industrie dominante était différente. Maintenant, compte tenu du plus grand nombre d'interviews réalisées par téléphone plutôt qu'en personne (voir le chapitre 4), de l'exclusion élargie des régions géographiques éloignées (mentionnées à la section 2.4) et du recours au FUL pour fournir des listes de logements, les économies éventuelles offertes par ces deux innovations semblaient moins intéressantes comparativement à la réduction de l'efficacité du plan et à l'incidence sur les estimations; par conséquent, elles ont été abolies pour ce remaniement.

Pour le remaniement, l'échantillonnage à deux degrés a été utilisé dans toutes les provinces sauf l'Î.-P.-É. D'autres modifications ont été apportées à la méthodologie de stratification : l'Î.-P.-É. stratifiée différemment pour faciliter le nouveau plan de sondage à un degré (voir la section 2.5.6), nouveau type de strate spéciale introduit pour régler le problème des grandes UPE à Toronto (voir la section 2.5.4) et prise en compte des besoins particuliers de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) (voir la section 2.5.3).

2.5.2 Taille des strates

La taille et le nombre de strates dans chaque intersection RE-REAE-RMR sont déterminés en fonction de la répartition de l'échantillon, du nombre d'UPE à sélectionner dans chaque strate et du nombre de ménages à sélectionner au sein de chaque UPE (on parle également de rendement de l'échantillon ou de facteur de densité). La répartition à chaque intersection a déjà été expliquée. Le nombre d'UPE est basé sur la stratégie de renouvellement, où le sixième de l'échantillon est renouvelé tous les mois. Pour mettre en œuvre cette approche, il est préférable de sélectionner six UPE (ou parfois 12) dans chaque strate. Enfin, des études antérieures ont déterminé que, pour améliorer l'efficacité du plan de sondage, le rendement de l'échantillon pour une UPE devrait être de 10 dans les strates rurales, de huit dans les strates urbaines et de six dans les strates englobant les RMR de Montréal, de Toronto et de Vancouver. En sélectionnant un plus grand nombre de ménages par UPE dans les strates rurales, on réduit les frais de déplacement par unité pour la collecte en personne. À l'autre extrême, en sélectionnant six ménages par UPE dans les plus grandes RMR, on contribue à augmenter le nombre d'UPE requises dans l'échantillon, améliorant ainsi la précision des estimations. Cette réduction du facteur de rendement augmente également le nombre de strates requises. Des strates plus nombreuses et plus petites devraient donner lieu à une augmentation de l'homogénéité des UPE qui s'y trouvent, ce qui devrait également améliorer l'efficacité du plan de sondage.

En combinant ces contraintes (exigences de répartition, six UPE sélectionnées par strate et nombre fixe de ménages sélectionnés dans chaque UPE), l'exigence relative à la taille de chaque strate dans une intersection, pour ce qui est des ménages, peut être calculée comme suit :

$$M_h = FSI \times 6 \times m_h^* \quad (2.1)$$

où

M_h est le nombre de ménages à regrouper dans chaque strate d'une intersection.

FSI est la fraction de sondage inverse telle qu'établie pendant les deux premières étapes de la répartition de l'échantillon.

m_h^* est le nombre de ménages à sélectionner par UPE. Comme expliqué au paragraphe précédent, ce nombre varie en fonction de la densité de la population de la région (rurale, urbaine, trois plus grandes RMR).

On peut déterminer le nombre de strates requises dans chaque région en divisant le nombre de ménages dans une région par ce résultat et en arrondissant le résultat au nombre entier le plus approprié. Habituellement, les strates dans une intersection RE-REAE-RMR que crée ce processus sont à peu près de la même taille.

2.5.3 Rajustements apportés aux limites géographiques

En utilisant l'expression de la taille de strate décrite précédemment, il a été impossible de créer des strates dans certaines petites intersections RE-REAE-RMR. Par conséquent, ces petites intersections ont été combinées avec une intersection voisine. Cette combinaison visait à ce que le groupe combiné respecte les limites de la RMR ou de la REAE autant que possible. Cette approche accorde implicitement une plus grande importance aux estimations par RMR et par REAE que par RE; par conséquent, l'efficacité du plan de sondage diminuait au niveau de la RE, mais demeurait inchangée pour les REAE et les RMR. Lorsque les limites du Recensement de 2011 pour une RMR ne correspondaient plus aux limites pour la REAE représentant cette ville⁸, les petites intersections ainsi obtenues étaient traitées comme si les limites des REAE correspondaient aux limites des RMR. Après avoir combiné les petites composantes, on a dénombré 120 intersections englobant les 10 provinces où la stratification est survenue.

À l'extérieur des RMR, il était avantageux de créer des strates séparées pour les régions urbaines et rurales pour trois raisons : les strates rurales ont plus de ménages que les strates urbaines (voir l'équation 2.1 à la section 2.5.2); les résidents des régions rurales ont des caractéristiques différentes que ceux des régions urbaines; la stratification correspondant à ces régions permet la mise en œuvre de stratégies de collecte plus appropriées. Dans certains cas, une région urbaine ou rurale est trop petite pour créer une strate au moyen de la taille déterminée à la section 2.5.2. En pareils cas, il faut combiner la région avec une région urbaine ou une région rurale voisine. Chaque cas a été évalué séparément.

L'ESCC utilise régulièrement la base aréolaire et peut compter sur la méthodologie de stratification de l'EPA pour repérer les UPE pour son échantillon. Pour le remaniement de l'EPA, il est avantageux d'apporter certains rajustements pour répondre aux besoins de l'ESCC et pour réduire l'incidence sur l'EPA. L'ESCC tire un échantillon à un taux beaucoup plus élevé dans certaines régions rurales que l'EPA ne le fait, et ses domaines géographiques ne correspondent pas toujours bien aux régions de l'EPA. Par le passé, l'ESCC se contentait de sélectionner plus de logements dans les UPE limitées qui étaient disponibles, ce qui entraînait un renouvellement beaucoup plus rapide que prévu des UPE (voir la section 2.6.4), ou de sélectionner plus d'UPE que l'EPA l'exigeait. Pour le remaniement, la solution était de créer des strates spécifiques supplémentaires de l'ESCC, qui veilleraient à ce qu'il y ait suffisamment d'UPE sélectionnées dans ces régions pour répondre aux exigences de l'ESCC. L'incidence est moindre sur les opérations de l'EPA.

Après les rajustements apportés aux limites géographiques, les UPE peuvent être regroupées en strates. Certaines UPE ont été attribuées à des strates spéciales (voir la section 2.5.4), et les UPE restantes dans chaque intersection ont été stratifiées géographiquement, puis de façon optimale (voir la section 2.5.5).

2.5.4 Strates spéciales

Les strates spéciales peuvent être divisées en deux catégories : celles créées pour améliorer l'efficacité, et celles créées pour cibler des populations précises. La première catégorie est utilisée pour regrouper des UPE éloignées, ainsi que des UPE comportant un grand nombre de logements à Toronto. La deuxième catégorie de strates spéciales aide à cibler les sous-populations d'intérêt pour les analystes qui utilisent les données de l'EPA.

Strates utilisées pour regrouper des UPE difficiles à gérer

On créait deux strates spéciales pour regrouper les UPE difficiles à gérer : les strates éloignées renfermant des UPE géographiquement isolées et où la collecte en personne est difficile, et les strates à Toronto renfermant des UPE comportant un grand nombre de logements. En regroupant ces UPE, le rythme auquel ces UPE sont sélectionnées peut être contrôlé.

8. Les définitions géographiques des REAE remontent à 2000. En raison de l'étalement urbain, bien des RMR qui étaient auparavant des REAE ont maintenant dépassé les limites d'une REAE.

Une importante proportion du Canada est habitée par une faible proportion de la population. Les coûts de collecte sont élevés dans les régions ayant une petite population, tandis que l'incidence de ces régions sur les principales estimations de l'EPA est relativement faible. Ces UPE ont été relevées au moyen des données du Recensement 2011 sur la densité de la population, les distances des centres urbains et l'accessibilité par la route. Si elles étaient en nombre suffisant dans une province, ces UPE étaient regroupées en une strate éloignée. En affectant ces régions à des strates précises, le nombre de ces UPE sélectionnées dans un échantillon donné peut être mieux contrôlé, ce qui améliore le contrôle de l'affectation des ressources de l'EPA.

Comme mentionné à la section 2.4, Toronto contient de nombreuses UPE comportant de 600 à 1000 logements, contrairement aux autres UPE de l'EPA dans les provinces. Si les UPE de plus de 1000 logements étaient stratifiées avec les UPE de 100 logements, il y aurait une augmentation considérable de la variance d'échantillonnage compte tenu de la stratégie de sélection d'échantillon utilisée. Cependant, le plan demeure efficace si la taille des UPE est relativement homogène dans une strate. Étant donné que ces UPE ne pouvaient pas être scindées davantage, elles ont plutôt été regroupées en tant que strates spéciales. Ainsi, la variabilité de la taille des UPE dans les strates de Toronto est réduite, ce qui donne lieu à un plan plus efficace.

Le tableau B.2 à l'annexe B présente le nombre de ménages dans les strates spéciales de la première catégorie.

Strates pour cibler certaines sous-populations

Trois types de sous-populations ont été ciblés par les strates spéciales : les ménages ayant un revenu élevé, les peuples autochtones et les immigrants récents. Pour simplifier, les termes strates à revenu élevé, strates autochtones et strates d'immigrants seront employés dans le reste du présent document, bien qu'ils soient techniquement incorrects puisque ces strates ne contiennent pas uniquement des ménages à revenu élevé, des personnes autochtones ou des immigrants. Des strates à revenu élevé ont été créées dans la plupart des grandes RMR. Des strates autochtones ont été créées en Colombie-Britannique, en Alberta et en Saskatchewan. Des strates d'immigrants ont été créées au Manitoba seulement.

Étant donné que l'EPA tire un échantillon de grappes de logements, au lieu des personnes directement, il est difficile de cibler ces sous-populations rares, en particulier lorsqu'elles ne vivent pas toutes dans le même quartier. Même dans un quartier à plus forte prévalence de la sous-population, bien des ménages n'ont toujours pas de membres de la sous-population, ce qui fait qu'un échantillon de ces logements ne donnera pas nécessairement beaucoup plus de membres qu'un échantillon d'une autre région. Puisqu'il n'existe pas de meilleur outil respectant les contraintes du plan de l'EPA pour cibler des sous-populations, des strates spéciales peuvent aider en veillant au moins à ce qu'une UPE à plus forte prévalence soit sélectionnée.

Pour que les strates spéciales englobent efficacement une population cible, elles doivent avoir une plus forte prévalence de la population cible et représenter un important pourcentage de l'ensemble de la population cible. Cependant, même si elles produisent de bonnes estimations pour leur population cible, les strates spéciales ne peuvent pas être justifiées si leur introduction donne lieu à une importante détérioration de la qualité des principales estimations de l'EPA. Afin de parvenir à un compromis acceptable, deux lignes directrices d'une étude réalisée pour le dernier remaniement ont été utilisées. La première ligne directrice énonce que les strates doivent être basées sur la prévalence de caractéristiques particulières. Par exemple, il serait futile de créer une strate des immigrants dans le Nord du Manitoba, où la proportion d'immigrants est très faible. La deuxième ligne directrice énonce qu'on ne peut pas utiliser plus de 8 % d'un domaine pour créer chaque type de strate spéciale. Cette limite garantit que la création de ces strates n'aura pas d'effet néfaste majeur sur les principales estimations de l'EPA, d'après une étude menée au moyen des données du Recensement de 1996 et de 2001.

Au moyen de ces deux lignes directrices, on créait les strates spéciales en ordre. Pour chaque catégorie, on les créait en relevant les UPE ayant la plus forte prévalence de la sous-population d'intérêt. Les UPE étaient alors continuellement ajoutées aux strates en ordre décroissant de prévalence, jusqu'à ce que la limite de 8 % pour le domaine soit atteinte. Au moyen de cette approche, ces strates ne sont pas contiguës et peuvent être très dispersées géographiquement parlant.

Les strates à revenu élevé étaient créées en premier. Les UPE dans une RMR donnée étaient d'abord classées en ordre décroissant selon la proportion des ménages ayant un revenu de plus de 150 000 \$, en fonction du fichier sur la famille T1 (FFT1) de 2012 produit à partir des déclarations de revenus de 2012 reçues par l'Agence du revenu

du Canada.⁹ Les UPE dans le haut de cette liste étaient affectées à une strate à revenu élevé jusqu'à ce que la taille prédéterminée de la strate ait été atteinte (voir la section 2.5.2). Si la limite de 8 % n'était pas atteinte, une autre strate à revenu élevé était créée pour la même RMR. Ainsi, des strates à revenu élevé ont été créées pour la plupart des RMR.

Pour créer les strates autochtones, la stratégie de base a dû être légèrement modifiée. Les strates à revenu élevé respectent les limites des RMR, mais un nombre considérable de personnes autochtones résident à l'extérieur de ces limites. Par conséquent, on a créé deux strates spéciales séparément dans les RMR et à l'extérieur des RMR. De plus, certaines intersections RE-REAE à l'extérieur des RMR étaient trop petites pour former une strate autochtone, même si plusieurs UPE dans ces intersections avaient une forte concentration de ménages autochtones. Pour remédier à ce problème, les strates autochtones créées à l'extérieur des RMR respectent les limites des REAE seulement, au lieu de celles des intersections RE-REAE. Enfin, les UPE déjà affectées à une strate éloignée ne pouvaient pas être attribuées également à une strate autochtone. Parmi les UPE restantes aux intersections combinées, les UPE ont été classées dans les strates en ordre décroissant selon la proportion des ménages dont au moins un membre a déclaré avoir une identité autochtone dans l'Enquête nationale auprès des ménages de 2011, jusqu'à ce que la limite de 8 % soit atteinte. Comme pour les strates à revenu élevé, plusieurs strates ont été créées au besoin pour atteindre la limite de 8 %.

Pour créer les strates d'immigrants au Manitoba, des strates ont également dû être créées à l'intérieur et à l'extérieur des RMR. Étant donné que la majeure partie de la population des immigrants récents du Manitoba habitent à Winnipeg et que la prévalence des immigrants était faible ailleurs, seulement deux strates d'immigrants ont été créées à l'extérieur de Winnipeg. Pour ces deux strates, les UPE à l'extérieur de Winnipeg ont été classées en ordre décroissant selon la proportion des ménages dont au moins un membre avait immigré au Canada au cours des 10 dernières années d'après l'Enquête nationale auprès des ménages de 2011.

En utilisant seulement 8 % de Winnipeg et des strates à l'extérieur de Winnipeg, on n'aurait pas eu une représentation adéquate de la population cible, puisque les immigrants récents sont disproportionnellement concentrés à Winnipeg. Cependant, l'utilisation de plus de 8 % de la RMR pour créer des strates en triant les UPE en ordre décroissant de prévalence aurait eu de graves conséquences sur les principales estimations de l'EPA pour Winnipeg. Au final, on a décidé que la proportion supérieure de 25 % des UPE¹⁰ en ce qui concerne la prévalence devait être isolée et stratifiée en 12 strates au moyen du même algorithme d'optimisation que celui qui a été utilisé pour stratifier les UPE à l'extérieur des strates spéciales (voir la section 2.5.5). Ainsi, on a pu créer plus de strates pour englober la population cible sans grande incidence sur les estimations de l'EPA. Des simulations ont été effectuées pour évaluer des stratégies de rechange, et cette solution était la plus avantageuse.

Les tableaux B.3 à l'annexe B donnent le nombre de ménages dans les strates spéciales, la prévalence de la population cible et la proportion de la sous-population incluse dans les strates spéciales.

2.5.5 Stratification des UPE restantes

Après avoir formé des strates spéciales, qui englobent seulement une petite portion du territoire canadien, les UPE restantes dans les neuf provinces autres que l'Î.-P.-É.¹¹ sont stratifiées dans les régions géographiques dont il est question à la section 2.5.3. Pour déterminer le nombre de strates devant être créées dans une région, le nombre de ménages n'appartenant pas à des strates spéciales a été divisé par la taille de la strate ciblée de l'équation (2.1) à la section 2.5.2. Étant donné que ce quotient n'est pas un nombre entier, le résultat a été arrondi¹². Si plus d'une strate était nécessaire pour une région, les UPE étaient d'abord stratifiées géographiquement, puis de façon optimale (les deux méthodes sont décrites ci-après).

Stratification géographique

Dans le cas des RMR, chacune était divisée en plusieurs composantes devant servir de base pour la stratification. Les régions suivantes étaient considérées : la plus grande subdivision de recensement (SDR), la deuxième plus

9. Étant donné qu'il n'y a pas de FFT1 pour chaque ménage dans une UPE, la proportion de ménages à revenu élevé a dû être estimée à partir de la proportion parmi les ménages ayant des FFT1 dans l'UPE. Dans la majorité des UPE, les FFT1 étaient disponibles pour 90 % des ménages ou plus. Par conséquent, cette estimation devrait être fiable. Les UPE où seulement quelques FFT1 étaient disponibles ont été exclues.

10. Les UPE déjà affectées à des strates à revenu élevé ont été exclues. Aucune strate autochtone n'a été créée à Winnipeg.

11. La stratification à l'Î.-P.-É. est expliquée à la section 2.5.6

12. Dans certains cas, ce nombre a été arrondi à la baisse, habituellement lorsqu'il y avait trop peu d'UPE pour créer la strate supplémentaire.

grande SDR, la troisième plus grande SDR, les UPE urbaines restantes et les UPE rurales. Ces régions ont été créées uniquement si la RMR nécessitait plusieurs strates et si la région avait la taille de strate ciblée. Autrement, elles étaient combinées avec d'autres régions.

Dans une région, si une seule strate était nécessaire, alors la stratification est complète. Si de deux à neuf strates étaient nécessaires, les UPE étaient stratifiées de façon optimale (voir la description ci-après) dans la composante. Si plus de 10 strates étaient nécessaires, la composante était d'abord divisée en super-strates – des régions compactes comportant un nombre semblable de ménages – pour assurer une meilleure répartition géographique des UPE sélectionnées dans un échantillon. Les UPE étaient alors stratifiées de façon optimale dans les super-strates.

À l'extérieur des RMR, les régions étaient définies au moyen de la plus grande agglomération de recensement (AR) dans une REAE, des UPE urbaines restantes et des UPE rurales restantes. Dans chaque composante, les UPE étaient attribuées au bon nombre de strates au moyen d'une stratification optimale.

Stratification optimale

Après la stratification géographique, les UPE dans les régions nécessitant deux strates ou plus étaient stratifiées de façon optimale. La stratification optimale a pour objet de réduire la variance de l'échantillonnage de plusieurs variables d'intérêt en regroupant les UPE ayant des caractéristiques semblables, créant ainsi des strates le plus homogènes possible tout en respectant les contraintes relatives à la taille des strates déterminées à la section 2.5.2. Pour ce faire, on a utilisé le processus itératif décrit ci-après.

L'algorithme utilisé pour la stratification optimale est basé sur une méthode itérative mise au point par Friedman et Rubin (1967) et modifiée par Drew, Bélanger et Foy (1985). En commençant avec une stratification initiale aléatoire comportant des strates de tailles égales, l'algorithme échange une UPE entre deux strates et vérifie si cette nouvelle stratification diminue une somme pondérée de carrés de données auxiliaires. Si la somme des carrés a diminué, la nouvelle stratification remplace la précédente; autrement, la stratification précédente est conservée. Les échanges d'UPE se poursuivent de façon itérative jusqu'à ce qu'aucun échange ne se traduise par une diminution. Le processus est alors répété au moyen de différentes stratifications initiales. La stratification associée à la plus petite variance est conservée¹³.

La somme pondérée des carrés est calculée en fonction de plusieurs caractéristiques auxiliaires. La liste de ces caractéristiques d'intérêt (29 au total) est identique à la liste du dernier remaniement et est disponible à la fin de l'annexe B. Le revenu du ménage s'est vu attribuer trois fois la valeur du facteur de pondération comparativement au reste des caractéristiques dans la somme pondérée des carrés, parce que le revenu est corrélé à plusieurs variables de l'EPA. Toutes les autres variables ont reçu un facteur de pondération égal pendant le processus.

Une fois les stratifications géographique et optimale terminées, on définit la variable géographique de l'EPA en attribuant des identificateurs uniques à chaque strate créée et à chaque UPE attribuée à cette strate. Le résultat est une base aréolaire terminée de l'EPA dans toutes les provinces, sauf l'Î.-P.-É. Cette base est utilisée pour le premier degré de sélection de l'échantillon (section 2.6).

2.5.6 Création de strates à l'Île-du-Prince-Édouard

Pour le remaniement, on a décidé d'utiliser l'échantillonnage à un degré à l'Î.-P.-É. Comme on l'a expliqué aux sections 2.1 et 2.2, le principal avantage de l'échantillonnage à deux degrés est de réduire l'éparpillement géographique des échantillons sélectionnés, ce qui réduit les frais de déplacement pour la collecte des données d'enquête, bien qu'au détriment de l'efficacité du plan. Toutefois, étant donné que l'Î.-P.-É. est une petite province et que bien des cas sont maintenant pris en charge par téléphone et non pas en personne, les frais de déplacement à l'Î.-P.-É. sont minimes. Par le passé, l'autre motif de l'échantillonnage à deux degrés était l'absence d'une liste complète de logements pour une province au complet, ce qui limitait les options de plan d'enquête. Cependant, compte tenu des améliorations apportées récemment au Fichier de l'univers des logements (section 2.2.1), une liste raisonnablement à jour de tous les logements à l'Î.-P.-É. est maintenant disponible. Bien que les CV des estimations mensuelles du chômage pour l'Î.-P.-É. soient souvent au-dessus de l'objectif de 7 %, il n'était pas prudent d'augmenter la taille de l'échantillon, puisque l'Î.-P.-É. a déjà le plus haut taux d'échantillonnage au pays.

13. Cette méthode d'optimisation est appelée escalade à redémarrage aléatoire.

Par conséquent, on a décidé de tout simplement sélectionner les logements à l'Î.-P.-É. dans une liste au hasard (échantillonnage à un degré aléatoire systématique) pour améliorer l'efficacité du plan. Étant donné que l'ESCC doit échantillonner des logements à l'Î.-P.-É. à un taux légèrement plus élevé que pour l'EPA, la FSI (section 2.3.2) a été rajustée de façon à ce que chaque échantillon contienne suffisamment de logements pour l'EPA ou l'ESCC.

Malgré l'absence d'UPE à stratifier, il demeurerait avantageux de stratifier la province pour plusieurs raisons. Sans strates, la sélection aléatoire de logements pourrait donner lieu à des échantillons où un seul logement est sélectionné dans une partie de l'île. En ce qui concerne la stratégie de collecte, il serait difficile de créer une charge de travail à temps plein pour un intervieweur responsable de cette région, ce qui pourrait être très coûteux si cet intervieweur s'occupait également d'autres parties de l'île. Pour ce qui est du plan de sondage, cette partie de l'île serait mal représentée dans l'échantillon, et la variance de l'échantillonnage augmenterait probablement si cette partie de la province avait des caractéristiques différentes. La création de strates permet de faire en sorte que chaque échantillon de l'EPA représente mieux la province dans son ensemble. Par souci d'uniformité avec les unités géographiques normalisées, des strates géographiquement contiguës ont été créées au moyen des AD du recensement.

Comme on l'a expliqué à la section 2.2, un sixième de l'échantillon est remplacé chaque mois, ce qui fait que l'échantillon de l'Î.-P.-É. devait être divisé en six groupes de renouvellement, ce qui est plus difficile en l'absence de grappes dans une strate¹⁴. Pour surmonter cette difficulté, les strates géographiques de l'Î.-P.-É. ont été fusionnées en groupes de six pour former des super-strates. Le processus tenait compte du nombre de ménages dans les super-strates et réduisait au minimum la distance entre les strates inhérentes. Une version modifiée de l'algorithme Friedman-Rubin susmentionné a été utilisée. Ces super-strates respectaient également les limites de Charlottetown et Summerside, les deux grandes villes de la province. Ensuite, chaque strate dans une super-strate était affectée au hasard à un des six groupes de renouvellement, de manière à équilibrer le nombre total de logements occupés dans chaque groupe de renouvellement. Par conséquent, le sixième de l'échantillon de l'Î.-P.-É. peut être remplacé tous les mois.

2.6 Stratégie de sélection de l'échantillon

Une fois la répartition et la stratification terminées, toutes les composantes sont en place pour sélectionner l'échantillon. La présente section fournit une description conceptuelle de la méthode de sélection et de renouvellement utilisée par l'EPA dans les autres provinces que l'Î.-P.-É. Pour l'Î.-P.-É., l'échantillonnage aléatoire systématique des logements dans les strates est utilisé. De l'information supplémentaire sur le traitement de la croissance et de la tenue à jour de la base de sondage se trouve au chapitre 3. Pour une description plus détaillée des probabilités d'échantillonnage et des poids d'échantillonnage, voir le chapitre 6.

2.6.1 Affectation de l'échantillon des UPE aux strates

Lorsqu'un plan à deux degrés est utilisé, la théorie des sondages stipule qu'il est préférable de sélectionner les UPE avec une probabilité proportionnelle à leur taille lorsque la mesure de la taille est également corrélée aux estimations d'intérêt. C'est le cas de l'EPA. Par exemple, le nombre de personnes qui travaillent dans une UPE est fortement corrélé au nombre de personnes qui vivent dans l'UPE. Par conséquent, les UPE pour l'EPA sont idéalement sélectionnées avec une probabilité proportionnelle à leur taille. La mesure de la taille utilisée pour l'EPA est en fonction du nombre de ménages dans l'UPE, selon l'estimation à partir du FUL (des explications se trouvent à la section 2.2.1)¹⁵.

La première étape consiste à déterminer le nombre d'UPE à sélectionner dans chaque strate. Par définition, comme on le décrit en déterminant la taille des strates, ce nombre devrait être de six. Cependant, en raison de l'arrondissement, de la création de strates spéciales et d'autres facteurs, le nombre de strates définies et la taille de l'échantillon requise ne correspondent pas nécessairement à la sélection de six UPE. De plus, pour simplifier le processus de renouvellement de l'échantillon – où un sixième de l'échantillon est renouvelé tous les mois – il est préférable de sélectionner un multiple de six UPE dans chaque strate. La méthode de renouvellement sera décrite plus en détails plus loin dans le chapitre.

14. Voir la section 2.6.1 pour découvrir comment les groupes de renouvellement sont formés dans les autres provinces.

15. En pratique, la mesure de la taille utilisée est la fraction de sondage inverse, qui est obtenue à partir du nombre de ménages. Pour plus d'information sur ce calcul, voir la section 2.7.1.

Pour déterminer le nombre d'UPE à sélectionner dans une strate, le nombre de ménages à sonder dans la strate doit être précisé. Jusqu'ici, la répartition est seulement connue comme le niveau de l'intersection RE-REAE-RMR. Les strates sont associées au même taux d'échantillonnage que l'intersection où elles se trouvent. Par conséquent, dans une intersection RE-REAE-RMR, le taux d'échantillonnage constant suppose que l'échantillon est affecté à des strates proportionnellement au nombre de ménages. La répartition des ménages aux strates est déterminée par le nombre de ménages dans la strate, divisé par la fraction de sondage inverse (FSI) pour l'intersection RE-REAE-RMR.

Le nombre d'UPE à sélectionner dans la strate est déterminé par la répartition des ménages aux strates divisée par le nombre cible de ménages à sonder par UPE sélectionnée. Comme mentionné à la section 2.5.2, ce nombre cible est de six ménages par UPE dans les strates des RMR de Montréal, de Toronto et de Vancouver, de huit dans les strates urbaines à l'extérieur de ces RMR et de 10 dans les strates rurales. Si le résultat de la deuxième division est plus proche de six UPE que 12, six UPE seront sélectionnées dans la strate. Autrement, 12 UPE seront sélectionnées, ce qui peut donner lieu à 18 UPE dans certains cas rares.

L'approche décrite ci-après est basée sur la sélection de six UPE. La même approche s'applique lorsque 12 ou 18 UPE sont sélectionnées.

2.6.2 Aperçu de la méthode RHC

L'EPA sélectionne l'échantillon des UPE au moyen de la méthode Rao-Hartley-Cochran (RHC). La méthode RHC est utilisée parce qu'elle permet de mettre à jour les probabilités de sélection lorsqu'une croissance importante est observée dans certaines UPE. La méthode décrite dans Keyfitz (1951) peut être combinée à la méthode RHC pour mettre à jour les probabilités, tout en optimisant le recoupement des UPE sélectionnées avant et après la mise à jour. Pour plus d'information sur la méthode RHC, voir Rao, Hartley et Cochran (1962).

Lorsque vous utilisez la méthode RHC pour sélectionner plusieurs UPE dans une strate, toutes les UPE doivent d'abord être réparties en groupes, chacune renfermant à peu près le même nombre d'UPE – plus ou moins une. Dans le cas de l'EPA, les groupes utilisés sont les six groupes de renouvellement. Une fois que les UPE ont été distribuées aux groupes de renouvellement, une UPE est sélectionnée par groupe selon la probabilité proportionnelle à la taille dans le groupe. On peut résumer la démarche par l'équation suivante :

$$\pi_{hij} = \frac{M_{hij}}{\sum_{j \in hi} M_{hij}} \quad (2.2)$$

où

M_{hij} est le nombre de ménages dans l'UPE j dans le groupe de renouvellement i de la strate h

$\sum_{j \in hi} M_{hij}$ est le nombre total de ménages dans toutes les UPE du groupe de renouvellement i de la strate h .

π_{hij} est la probabilité de sélection de l'UPE j dans le groupe de renouvellement i de la strate h .

Au lieu d'utiliser le nombre de ménages, l'EPA utilise la fraction de sondage inverse arrondie de l'UPE (FSI_{hij}^*) comme mesure de la taille de l'UPE – voir la description ci-après. Ces valeurs sont utilisées principalement à cause de la façon dont l'échantillon est sélectionné à la deuxième étape. On démontrera plus tard que ce n'est pas très loin de l'utilisation du nombre de ménages pour ce qui est de la probabilité d'échantillonnage pour l'UPE.

Probabilité de sélection au deuxième degré

Les logements sont sélectionnés à l'intérieur des UPE sélectionnées, en fonction de probabilités qui veillent à ce que tous les ménages dans la strate aient la même probabilité générale de sélection. On parle souvent d'un plan auto-pondéré.

Au deuxième degré, les logements sont sélectionnés à partir de la ligne de listage des UPE produite par le Registre des adresses et/ou le listage sur le terrain des UPE¹⁶ au moyen de l'échantillonnage systématique, où les ménages sont prélevés à intervalles réguliers. Cette méthode est recommandée parce qu'elle est simple à utiliser, qu'elle assure une bonne répartition des logements sélectionnés dans l'UPE, qu'elle contrôle le recoupement des échantillons et qu'elle facilite l'ajout de nouveaux logements à l'UPE. Pour sélectionner l'échantillon systématique de logements, la FSI de l'UPE, FSI_{hij} , ainsi qu'un point de départ dans la liste, doivent être déterminés. La FSI_{hij} peut être obtenue à partir du nombre de ménages dans l'UPE et de la FSI de la strate, FSI_h , au moyen de l'équation suivante :

$$FSI_{hij} = \left(\frac{M_{hij}}{\sum_{j \in hi} M_{hij}} \right) FSI_h \quad (2.3)$$

où

FSI_{hij} est la fraction de sondage inverse (FSI) dans l'UPE j dans le groupe de renouvellement i de la strate h .

FSI_h est la fraction de sondage inverse (FSI) de la strate h établie pendant la répartition de l'échantillon.

Étant donné que la FSI_h est constante pour toutes les UPE d'un groupe, la FSI_{hij} est proportionnelle au nombre de ménages dans l'UPE. Pour le remaniement, une transition à un échantillonnage aléatoire simple a été envisagée. Cependant, une étude a démontré que les logements adjacents ont des réponses corrélées, ce qui implique que pour les estimations mensuelles, l'échantillonnage systématique devrait entraîner une meilleure représentation de l'UPE, puisque l'échantillon sera certainement dispersé dans l'ensemble de la région. De plus, l'échantillonnage systématique peut donner une variance réduite des estimations de la variation d'un mois à l'autre puisque des ménages voisins sont utilisés pour remplacer ceux qui doivent être renouvelés.

Le système de sélection de l'EPA ne peut pas utiliser ces FSI directement et est plutôt configuré pour utiliser des fractions de sondage inverses arrondies FSI_{hij}^* . Le résultat de l'équation 2.3 est donc arrondi à la hausse ou à la baisse comme suit

$$\sum_{j \in hi} FSI_{hij}^* = FSI_{hi}^* = FSI_h^*, \forall i \in h \quad (2.4)$$

On parle ici d'un arrondissement contrôlé. La probabilité de sélection au deuxième degré d'un ménage lorsque l'UPE j dans le groupe de renouvellement i de la strate h est sélectionné est $1/FSI_{hij}^*$.

La valeur arrondie FSI_{hij}^* a certaines interprétations utiles. Premièrement, il s'agit de l'intervalle d'échantillonnage à utiliser dans l'échantillonnage systématique si l'UPE correspondante est sélectionnée à la première étape. En appliquant cet intervalle d'échantillonnage, le nombre approprié de ménages sera sélectionné dans l'UPE¹⁷. Deuxièmement, FSI_{hij}^* est le nombre d'échantillons distincts disponibles dans l'UPE. Dans la terminologie de l'EPA, ce concept s'appelle le nombre de *départs aléatoires*.

Probabilités de sélection au premier degré

Comme susmentionné, un plan autopondéré est réalisable lorsque les probabilités du premier et du deuxième degré correspondent. Les probabilités de sélection au deuxième degré ont été définies précédemment comme $1/FSI_{hij}^*$. Afin de préserver l'aspect d'autopondération, les valeurs FSI_{hij}^* doivent être utilisées comme des

16. Le chapitre 3 décrit comment le Registre des adresses et le listage sur le terrain sont utilisés pour créer la base de sondage.

17. Cet objectif correspond au nombre de ménages dans le groupe divisé par la fraction de sondage inverse de la strate.

valeurs de taille dans l'échantillon de la probabilité proportionnelle à la taille. La probabilité de sélection au premier degré associée à chaque UPE est donc la suivante :

$$\pi_{hij}^* = \frac{FSI_{hij}^*}{\sum_{j \in hi} FSI_{hij}^*} = \frac{FSI_{hij}^*}{FSI_h^*}. \quad (2.5)$$

Ce n'est pas très différent de l'utilisation du nombre de ménages comme mesure de la taille. Comme susmentionné, la FSI_{hij}^* est proportionnelle au nombre de ménages dans l'UPE. Dans ce cas-ci,

$$\pi_{hij}^* = \frac{FSI_{hij}^*}{\sum_{j \in hi} FSI_{hij}^*} \approx \frac{M_{hij}}{\sum_{j \in hi} M_{hij}} = \pi_{hij}. \quad (2.6)$$

La seule différence entre les deux probabilités est attribuable à l'arrondissement contrôlé de la FSI_{hij}^* . Par conséquent, la probabilité globale de sélection du ménage k dans l'UPE j dans le groupe de renouvellement i de la strate h est :

$$\pi_{hijk}^* = \frac{FSI_{hij}^*}{\sum_{j \in hi} FSI_{hij}^*} \times \frac{1}{FSI_{hij}^*} = \frac{1}{FSI_h^*}. \quad (2.7)$$

Tel qu'exigé, l'équation 2.7 suggère que la probabilité de sélection est la même pour tous les ménages dans la même strate. Le plan de sondage de l'EPA est donc auto-pondéré dans la strate.

2.6.3 UPE et sélection d'un départ

En pratique, pour sélectionner une UPE dans un groupe de renouvellement, les UPE d'un groupe de renouvellement sont placées en ordre aléatoire. Un nombre entier aléatoire U est alors tiré d'une répartition uniforme dans l'intervalle $[1, FSI_h^*]$. Ce nombre aléatoire U a deux fonctions. D'abord, il sert à identifier la première UPE sélectionnée. Cette UPE est la première pour laquelle le total cumulatif des FSI_{hij}^* est supérieur ou égal à U (ou $\sum_{j \leq j^*} FSI_{hij}^* \geq U$ où l'indicateur j suit l'ordre aléatoire).

Il détermine également le nombre de départs aléatoires à utiliser dans cette première UPE avant de passer à la prochaine UPE. Le nombre de départs à utiliser dans la première UPE est $D_{j^*} = \left(\sum_{j \leq j^*} FSI_{hij}^* \right) - U + 1$. Enfin, un deuxième nombre entier aléatoire $U_{j^*} \in [1, FSI_{hij^*}^*]$ ¹⁸ est sélectionné. Ce nombre indique le premier départ aléatoire à utiliser pour sélectionner l'échantillon de logements pour l'UPE j^* . L'échantillon systématique pour une UPE sélectionnée hij est composé du logement dont le numéro de ligne dans la base de logements est équivalent au point de départ , ainsi que d'autres logements dont les lignes supplémentaires sont dans les intervalles

18. Ce deuxième nombre aléatoire comporte deux fonctions. Il tient plutôt compte du fait que la taille de l'échantillon associée au dernier départ aléatoire est parfois plus petite que celle des premiers départs. Nous espérons donc stabiliser la taille globale de l'échantillon au fil du temps. De plus, nous préparons ainsi le terrain pour appliquer la règle du nombre minimum de départs à utiliser.

de la FSI_{hij}^* . Par conséquent, les logements sont sélectionnés avec les numéros de lignes, de façon à ce que

$d = U_{j^*} + t \times FSI_{hij}^*$, $t = 0, 1, 2, \dots$ jusqu'à ce que d dépasse le nombre de lignes disponibles.

Ces logements demeureront dans l'échantillon pendant une période de six mois.

Gray (1973) et Alexander, Ernst et Haas (1982) utilisent deux approches différentes pour illustrer que cette méthode produit un échantillon qui respecte les probabilités de sélection précisées. Laflamme (2003) démontre le processus de sélection de l'échantillon au moyen d'un diagramme.

2.6.4 Renouvellement de l'échantillon

La section 2.6.3 décrit comment le premier échantillon de logements a été sélectionné dans chaque groupe au moyen de la méthode RHC. Au bout d'une période de six mois, il faut remplacer cet échantillon par de nouveaux logements. En continuant avec l'échantillon donné à la fin de la section précédente, le premier échantillon correspondait au départ aléatoire U_{j^*} de l'UPE j^* .

Si le nombre de départs aléatoires à utiliser pour l'UPE j^* est $D_{j^*} = 1$, le deuxième échantillon de logements correspondra au départ U_{j^*+1} de l'UPE suivante, j^*+1 , où $U_{j^*+1} \in [1, FSI_{hi(j^*+1)}^*]$. Autrement, si $D_{j^*} > 1$, le deuxième échantillon correspondra au départ $U_{j^*} + 1$ de l'UPE j^* (c.-à-d. les voisins de l'échantillon précédent).

Si $U_{j^*} + 1 > FSI_{hij^*}^*$, la sélection retourne en arrière pour utiliser le départ 1 de l'UPE j^* . En règle générale, au moyen de cette méthode, l'UPE j demeure dans l'échantillon pendant D_j périodes de six mois. Lorsqu'il s'avère nécessaire de remplacer les logements sondés, on utilise le départ aléatoire suivant. Après D_j périodes, l'échantillon passe à la valeur U_{j+1} de l'UPE $j+1$. Cette UPE demeurera dans l'échantillon jusqu'à ce que tous les départs aléatoires aient été utilisés. Il en va de même pour les UPE qui sont ajoutées à l'échantillon plus tard.

Cette méthode produit les résultats attendus : les probabilités de sélection sont toujours respectées au fil du temps. Malheureusement, elle comporte un important inconvénient. Comme mentionné, la première UPE demeure dans l'échantillon pendant un nombre de périodes aléatoire, et il arrive que ce nombre soit petit. Ce renouvellement rapide de la première UPE sélectionnée entraînerait une utilisation inefficace des ressources limitées de l'enquête. En fait, l'ajout d'une UPE à un échantillon nécessite beaucoup d'efforts, y compris la préparation du matériel, le listage éventuel de l'UPE et parfois, l'embauche et la formation d'un intervieweur. Par souci d'efficacité, il serait préférable d'amortir cet investissement en évitant un renouvellement trop rapide de la première UPE dans la mesure du possible.

Pour surmonter ce problème, l'EPA a élaboré une correction qui accroît le nombre de départs aléatoires de la première UPE sans introduire de biais dans les probabilités de sélection. Lorsque D_{j^*} est trop petit, en fonction d'un critère prédéterminé, il est augmenté afin de garder cette UPE dans l'échantillon plus longtemps. En pareil cas, le nombre de départs aléatoires à utiliser pour l'UPE j^*+1 doit être réduit proportionnellement afin d'éviter d'introduire un biais dans les probabilités de sélection. Certaines contraintes sont requises pour faire en sorte que l'augmentation du nombre de départs associés à la première UPE ne réduira pas trop le nombre de départs pour l'enquête à partir de la deuxième UPE. Gray (1973) démontre que cette approche n'introduit pas de biais dans les probabilités de sélection, tandis que Laflamme (2003) fournit des explications sur ces contraintes.

Cette méthode est appliquée séparément à chaque groupe de renouvellement. Cependant, les échantillons ne sont pas tous renouvelés en même temps. Un groupe RHC dans le groupe de renouvellement 1 est renouvelé en janvier et en juillet de chaque année. Les groupes RHC dans le groupe de renouvellement 2 sont renouvelés en février et en août, et ainsi de suite. En utilisant cette méthode, au début du remaniement, une liste peut être produite renfermant tous les départs qui seront dans l'échantillon de l'EPA pour chaque mois au cours des 10 prochaines années.

Figure 2.1
Renouvellement de l'échantillon de l'EPA

		Mois de l'enquête											
		JANV.	FÉVR.	MARS	AVR.	MAI	JUIN	JUILL.	AOÛT	SEPT.	OCT.	NOV.	DÉC.
Groupe de renouvellement	2	6 ^e											
	3	5 ^e	6 ^e										
	4	4 ^e	5 ^e	6 ^e									
	5	3 rd	4 ^e	5 ^e	6 ^e								
	6	2 ^e	3 rd	4 ^e	5 ^e	6 ^e							
	1	1 ^{er}	2 ^e	3 rd	4 ^e	5 ^e	6 ^e						
	2		1 ^{er}	2 ^e	3 rd	4 ^e	5 ^e	6 ^e					
	3			1 ^{er}	2 ^e	3 rd	4 ^e	5 ^e	6 ^e				
	4				1 ^{er}	2 ^e	3 rd	4 ^e	5 ^e	6 ^e			
	5					1 ^{er}	2 ^e	3 rd	4 ^e	5 ^e	6 ^e		
	6						1 ^{er}	2 ^e	3 rd	4 ^e	5 ^e	6 ^e	
	1							1 ^{er}	2 ^e	3 rd	4 ^e	5 ^e	6 ^e
	2								1 ^{er}	2 ^e	3 rd	4 ^e	5 ^e
	3									1 ^{er}	2 ^e	3 rd	4 ^e
	4										1 ^{er}	2 ^e	3 rd
	5											1 ^{er}	2 ^e
	6												1 ^{er}

Ce diagramme illustre le plan de renouvellement de l'échantillon de l'EPA. Les couleurs indiquent à quel groupe de renouvellement appartiennent les logements (l'orange correspond au groupe de renouvellement 1, le rose, au groupe de renouvellement 2, le bleu, au groupe de renouvellement 3, le vert, au groupe de renouvellement 4, le gris, au groupe de renouvellement 5 et le jaune, au groupe de renouvellement 6). Les chiffres dans les cases indiquent depuis combien de mois les logements associés à un groupe de renouvellement en particulier font partie de l'enquête. Comme l'indique le diagramme, un sixième de l'échantillon est renouvelé tous les mois. Par exemple, en avril, les logements du groupe de renouvellement bleu sont à leur deuxième mois de l'enquête, tandis que les logements du groupe de renouvellement gris sont à leur sixième et dernier mois de participation. Les logements du groupe de renouvellement gris seront remplacés par de nouveaux logements en mai, comme l'indique le diagramme. Les logements qui sont supprimés sont généralement remplacés par des logements des mêmes unités d'échantillonnage primaires respectives.

Chapitre 3 Création et tenue à jour de la base de logements

3.0 Introduction

Comme mentionné au chapitre précédent, l'Enquête sur la population active (EPA) utilise un plan de sondage à deux degrés dans toutes les provinces sauf à l'Île-du-Prince-Édouard. Un des avantages de cette approche est que l'échantillon est concentré dans un nombre limité de régions; par conséquent, il est possible de réaliser des interviews sur place. Au premier degré, des unités primaires d'échantillonnage (UPE) – également appelées grappes – correspondant à des régions géographiques sont sélectionnées. Il s'agit de parcelles de terrain relativement petites, souvent des aires de diffusion du recensement (AD). Dans les UPE sélectionnées, des logements sont sélectionnés au deuxième degré d'échantillonnage.

Aux deux degrés du processus d'échantillonnage, une base de sondage, c.-à-d. une liste de toutes les unités (grappes ou logements) qui font partie de la population cible, s'impose. Une base de sondage de bonne qualité aura des erreurs de couverture limitées et facilitera le contact avec les unités échantillonnées. Étant donné que de nouvelles unités sont continuellement ajoutées à la population cible et supprimées de cette dernière, il est important que la tenue à jour soit effectuée sur la base de sondage. Les détails sur la création de la base de sondage pour les UPE et les aspects du plan pour la sélection des ménages ont été décrits au chapitre 2.

3.1 Création de la base de logements

Dans les UPE sélectionnées, une liste complète de logements (une base de sondage) est requise afin de sélectionner l'échantillon du deuxième degré. La liste est obtenue soit par un exercice de listage effectué sur le terrain, soit à partir d'une liste existante, en particulier le Registre des adresses (RA). Une fois disponible, la liste de logements sera utilisée tant que l'UPE soit dans l'échantillon. Un défi constant consiste à déterminer quelles UPE venant d'être échantillonnées devraient faire l'objet d'un listage, et lesquelles peuvent se fier aux renseignements du RA. Le listage sur le terrain est une solution plus coûteuse qui devrait être évitée dans la mesure du possible. Il se produit habituellement lorsque l'information dans les listes existantes est de faible qualité.

3.1.1 Le Registre des adresses

Le RA est une base de données qui a été créée au départ pour les données du Recensement de la population canadienne de 1991, en vue d'améliorer la couverture du recensement. Il a été créé au moyen de plusieurs fichiers administratifs, notamment les factures de téléphone et les fichiers de permis de construire. Toute de suite après ce recensement, le RA a été mis à jour au moyen de la liste des adresses créée pendant le processus de dénombrement du recensement. Depuis cette première itération, le RA a continué d'être tenu à jour tous les trimestres au moyen de fichiers administratifs et du programme de listage du recensement, ainsi que des renseignements du recensement disponibles tous les cinq ans.

Le RA visait au départ à fournir et à tenir une liste d'adresses pour les collectivités ayant une population de plus de 50 000 personnes. La couverture du RA a été élargie après chaque recensement subséquent pour inclure les petits centres de population et les régions à l'extérieur des centres de population. À l'heure actuelle, le RA a une couverture nationale, mais il est plus précis dans les agglomérations.

En 2015, le RA incluait plus de 15 millions d'adresses. La grande majorité de ces adresses – environ 90 % – ont été associées à des logements résidentiels valides pendant le Recensement de 2011. Parmi les adresses restantes, 7 % ont été obtenues par le biais de mises à jour des fichiers administratifs et de listage sur le terrain en prévision du Recensement de 2016, et 3 % étaient des logements valides pendant un recensement précédent.

Pour apparaître dans le RA, un logement résidentiel doit posséder une adresse de voirie standard valide, ou une forme quelconque d'adresse descriptive. Pour les besoins de l'enquête, les adresses descriptives sont souvent incomplètes et ne fournissent pas nécessairement assez d'information pour repérer le logement. Lorsqu'il y a une proportion considérable d'adresses descriptives, la zone pourrait nécessiter un listage sur le terrain.

Deux fichiers clés sont extraits de la base de données du RA pour le processus de création de la base de logements de l'EPA : le Fichier de l'univers des logements et le Fichier de numéros de téléphone résidentiels.

Fichier de l'univers des logements

Le Fichier de l'univers des logements (FUL) est un extrait d'adresses du RA. Des règles s'appliquent pour faire en sorte que la liste contienne uniquement les logements qui correspondent à la population cible de l'EPA. Ces règles évoluent au fil du temps, à mesure que les méthodes pour détecter les adresses erronées ou en double s'améliorent. Les logements collectifs représentent également une petite proportion de la population cible de l'EPA, et ces logements sont disponibles par le biais du processus d'extraction du RA.

Fichier de numéros de téléphone résidentiels

Le Fichier de numéros de téléphone résidentiels (FTR) est une liste de numéros de téléphone résidentiels valides au Canada. Un grand nombre de ces numéros (88 % en 2015) peuvent être associés à l'adresse d'un logement qui se trouve dans le FUL. Le FTR peut donc être utilisé pour ajouter des numéros de téléphone – des coordonnées essentielles – à une grande partie des logements échantillonnés.

3.1.2 La Base nationale de données géographiques

Pour utiliser le RA dans un contexte de plan de sondage à deux degrés, chaque adresse doit d'abord être affectée à une UPE en particulier. Pour ce faire, on apparie le RA à la Base nationale de données géographiques (BNDG). La BNDG contient des couches cartographiques qui incluent les limites de l'UPE, les réseaux routiers, les voies navigables et d'autres repères géographiques. Cette information peut être utilisée pour associer des adresses à des tronçons de rue. Ces tronçons peuvent être au niveau de l'îlot (un îlot est un polygone composé de côtés de tronçons de rue contenus dans une AD) ou plus précisément au niveau du côté d'îlot (un seul tronçon de rue). Ces tronçons sont alors associés à une AD ou une UPE, ce qui associe véritablement le logement à l'UPE.

La BNDG est gérée en partenariat avec Élections Canada et évolue continuellement en raison de l'ajout régulier de routes et des mises à jour des limites géographiques, comme les limites municipales. Tous les trois mois, une nouvelle version de la BNDG est produite.

3.1.3 Mise en ordre de la liste d'adresses

Les adresses dans la base de logements doivent être organisées dans une liste ayant un ordre particulier pouvant être maintenu au fil du temps. Cet ordre facilite la recherche des logements sélectionnés, en plus d'aider les intervieweurs à reconnaître les omissions. L'ordre des adresses est créé par un processus de mise en ordre qui énumère les côtés d'îlots dans un ordre qui englobe la totalité de l'UPE, tout en réduisant au minimum la distance totale parcourue par l'intervieweur au moment de vérifier la liste d'adresses. Cet algorithme utilise les renseignements géographiques dans l'UPE provenant de la BNDG et est particulièrement utile pour les employés sur le terrain lorsque toutes les adresses peuvent être géocodées en fonction des côtés d'îlot. L'algorithme est exécuté pour la base de sondage au complet des UPE pour chaque version de la BNDG. Autrement dit, dans chaque UPE sélectionnée, la liste d'adresses est organisée dans un ordre particulier afin de faciliter et d'optimiser le listage.

3.2 Chargement et listage sur le terrain

Une fois que les logements ont été attribués à leurs UPE, des indicateurs de qualité pour la liste d'adresses peuvent être développés. La qualité détermine si la région exigera un listage sur le terrain ou si les renseignements du RA-BNDG suffiront comme liste à utiliser comme base de sondage pour les logements dans l'UPE.

Idéalement, les listes dans toutes les UPE seraient vérifiées sur le terrain (listage sur le terrain), mais le budget limite le nombre d'UPE pouvant tomber dans cette catégorie. La qualité de la liste d'adresses pour une UPE donnée dépend de la qualité du RA, de la qualité de la BNDG et de l'efficacité des règles d'admissibilité du FUL. L'objectif de cette stratégie consiste à utiliser le plus possible le RA, tout en tenant compte du fait que sa qualité fluctue d'une région à une autre.

On sait que la qualité du RA est la plus élevée dans les centres de population. Ces centres de population correspondent essentiellement au « secteur d'envoi par la poste », où la méthode de collecte du recensement consiste à joindre les ménages par courrier. Cette région correspond à environ 80 % des logements. D'après ces renseignements, les UPE sont classées dans un des trois groupes suivants :

Le groupe 0 du RA contient les UPE dans le secteur d'envoi par la poste. Aucun listage de départ sur le terrain n'est effectué, et le premier échantillon d'unités est sélectionné à partir de la liste basée sur le RA. Malgré l'absence d'un listage sur le terrain comme « UPE de l'EPA », une grande partie du secteur d'envoi par la poste fait l'objet d'un listage sur le terrain dans le cadre du programme de listage du recensement. Les résultats de ce listage sont traités par l'équipe du RA et apparaissent au bout du compte dans le FUL.

Le groupe 1 du RA correspond aux UPE n'appartenant pas aux secteurs d'envoi par la poste et sans listage de départ. La liste du RA est considérée comme de bonne qualité, d'après un ensemble de statistiques et d'indicateurs. L'échantillon initial de logements est sélectionné directement de la liste basée sur le RA.

Le groupe 2 du RA correspond aux UPE n'appartenant pas aux secteurs d'envoi par la poste avec un listage de départ. L'UPE doit être listée sur le terrain avant la sélection du premier échantillon.

La répartition de 2015 a attribué 72 % des UPE échantillonnées au groupe 0 du RA, 19 % au groupe 1 du RA et 9 % au groupe 2 du RA. Il s'agit d'un changement important par rapport au lancement du plan de 2005, où au moins 61 % des UPE avaient besoin d'un listage de départ. À mesure que les UPE sont ajoutées et retirées de l'échantillon et que la qualité du RA évolue (en particulier après le Recensement de 2016), la répartition des groupes du RA changera probablement.

3.2.1 Chargement initial

Pour les UPE du groupe 0 du RA ou du groupe 1 du RA, la liste de logements utilisée pour la sélection de l'échantillon est remplie à partir de la liste d'adresses de logements disponible dans le FUL associées à ces UPE. Ce processus s'appelle le chargement initial. L'échantillon de logements de l'EPA est sélectionné directement de cette liste.

Les UPE non listées ont tendance à avoir une plus forte proportion d'unités échantillonnées ayant reçu le code « invalide » ou « démoli » au moment de la collecte des données d'enquête. On parlera des erreurs de couverture au chapitre 8.

3.2.2 Listage de départ

Les UPE dans le groupe 2 du RA doivent faire l'objet d'un listage de départ. L'objectif du listage de départ consiste à dresser une liste complète et exacte des logements pour la première sélection d'échantillon dans une UPE. La liste de départ est préremplie avec les logements associés avec cette UPE en fonction du FUL. Chaque logement dans la liste est validé, modifié ou désactivé par les employés sur le terrain. De nouveaux logements peuvent également être ajoutés à la liste.

Mise en correspondance des UPE

Afin d'effectuer efficacement le listage sur le terrain, les limites de l'UPE doivent être affichées sur une carte. Une traduction adéquate du contenu de la carte par rapport aux caractéristiques physiques sur le terrain est primordiale pour déterminer quels logements appartiennent à l'UPE. De plus, les numéros d'îlot et les plages d'adresses sur la carte peuvent aider à repérer des adresses en particulier. Les adresses ou les descriptions des logements sont saisies par l'intervieweur sur le terrain au moyen de l'application de listage de Statistique Canada. Les cartes de l'UPE sont produites au moyen du logiciel Système généralisé de cartographie, en place depuis 2009. L'annexe D contient des exemples de cartes d'UPE et décrit plus de détails au sujet de leur création et de leurs utilisations.

Listage des logements collectifs

Le listage des logements collectifs n'est pas aussi simple que celui des logements privés occupés. Il y a deux grands critères pour le listage de logements collectifs. Premièrement, les détenus ne font pas partie de la population visée par l'EPA. De même, les résidents temporaires ayant un lieu de résidence habituel ailleurs ne sont pas admissibles. Généralement, seuls la résidence du propriétaire, les résidences des employés et les logements pour les résidents hors établissement (p. ex., les unités dans une résidence pour personnes âgées) seraient listés.

3.3 Mise à jour de la base de sondage

Peu importe si l'UPE a fait l'objet d'un listage de départ ou pas, chaque mois, il est possible de mettre à jour ou de corriger la liste de logements. Par conséquent, la plupart des problèmes de base de sondage sont temporaires et peuvent être rectifiés pour les occasions d'échantillonnage subséquentes.

3.3.1 Mise à jour et tenue à jour de la liste

Une fois qu'une UPE a été sélectionnée, les mises à jour régulières de la liste d'adresses peuvent être faites tous les trimestres à partir de chaque nouvelle version du FUL (mise à jour de la liste) ou tous les mois à partir de la vérification sur le terrain (tenue à jour de la liste). Pour les grappes du groupe 0 du RA, une combinaison de mise à jour et de tenue à jour de la liste est utilisée. Pour les grappes des groupes 1 et 2 du RA, la tenue à jour de la liste est la source principale.

Dans le cas de la tenue à jour de la liste, les logements peuvent être ajoutés, modifiés ou désactivés (avec un motif quelconque pour la désactivation). L'ordre des logements peut être changé dans la liste, ce qui a une incidence sur le numéro de séquence d'impression, mais le numéro d'identification permanent dans la grappe, la ligne de listage, demeure inchangé. Cette approche permet à l'intervieweur d'avoir un ordre de liste de préférence, tout en conservant efficacement l'historique de l'échantillon de chaque logement.

La tenue à jour s'effectue normalement « par renouvellement », c'est-à-dire pendant le premier mois d'échantillonnage (p. ex., pendant les mois de 'naissance' de janvier ou de juillet pour les UPE du renouvellement 1). Habituellement, l'intervieweur doit visiter l'UPE, parce qu'au moins une partie des logements sélectionnés doivent être contactés en personne.

Logements sélectionnés par les intervieweurs

La tenue à jour de la liste peut donner lieu à des logements sélectionnés par les intervieweurs (LSI). Il s'agit de nouveaux cas de l'EPA que l'intervieweur IPAQ doit traiter.

Il y a deux formes de LSI créés pendant la tenue à jour de la liste. D'abord, pendant la durée de vie de l'UPE, l'intervieweur peut régulièrement ajouter de nouveaux logements à mesure que la population s'accroît. Étant donné que la liste de logements est non limitative, d'autres logements peuvent être sélectionnés sur le terrain. Les structures ajoutées à la fin de la liste sont échantillonnées au moyen de la fraction de sondage inverse (FSI) au niveau de l'UPE et à partir de la prochaine ligne d'interview provenant de la dernière sélection de l'échantillon dans l'UPE. Une fois qu'on a sélectionné un logement, on détermine la prochaine ligne d'interview en appliquant la FSI de l'UPE.

On désigne la seconde forme de LSI par le terme « multiples ». Au cours de l'interview dans un logement sélectionné, l'intervieweur peut déterminer que la structure contient des logements distincts, généralement des unités de sous-sol ou d'étage pas facilement visibles de la rue. Étant donné que la liste des logements n'indique pas ces unités additionnelles sous forme de lignes distinctes, ces logements n'ont aucune chance d'être sélectionnés pendant la durée de vie de l'UPE. Afin de compenser les logements omis ainsi et dans d'autres cas semblables non résolus, on sélectionne toutes les unités omises avec le logement d'origine. Puis, on les ajoute à la liste comme multiples du logement sélectionné à l'origine et l'application génère un cas pour chaque multiple.

3.3.2 Traitement des zones en croissance

Comme les listes de logements de l'UPE sont ouvertes, une croissance extrême est possible. Les intervieweurs ne sont pas nécessairement en mesure de tenir à jour de longues listes à cause du coût associé à cette tenue à jour,

ainsi que du temps nécessaire à la tenue d'interviews pour l'important afflux de nouvelles unités échantillonnées qui accompagne une aussi longue liste. Bien que cette croissance extrême soit observée dans moins de 1 % des UPE, des options doivent être disponibles pour la gérer et la traiter.

Sous-échantillonnage de l'UPE

En fonction de la rétroaction provenant du terrain, les UPE à forte croissance peuvent entraver la capacité de l'intervieweur de terminer toutes les interviews qui lui ont été attribuées. Le degré de difficulté à cet égard peut être amplifié dans le cas d'une tâche associée à un mois de 'naissance', en particulier si la fraction de ménages nécessitant des interviews sur place est élevée. En pareils cas de croissance isolée, l'UPE est sous-échantillonnée pour réduire le fardeau. L'EPA utilise deux formes de sous-échantillonnage.

La première est une simple modification du taux d'échantillonnage pour l'UPE en question. Cette technique – également appelée sous-échantillonnage de grappe ou mécanique – est utilisée dans la majorité des cas. Souvent, il suffit de diminuer le taux d'échantillonnage par un facteur de deux afin de réduire la charge de travail de l'intervieweur de moitié.

La deuxième forme de sous-échantillonnage est l'insertion d'un degré supplémentaire de sélection de l'échantillon. Dans cette technique, des sous-grappes sont formées comme unités secondaires d'échantillonnage (USE). Par convention, les UPE peuvent être appelées des grappes, et les composantes des UPE peuvent être appelées les sous-grappes. En cas de forte croissance, les employés du bureau central délimitent quatre sous-grappes ou plus de tailles à peu près équivalentes pour ce qui est du nombre de ménages dans l'UPE. Deux des USE sont alors sélectionnées pour l'activité d'enquête, et les facteurs de sous-échantillonnage sont créés.

Les modifications de sous-échantillonnage entravent la probabilité d'échantillonnage des ménages. Des descriptions des rajustements à prendre en compte à cet égard se trouvent dans l'explication des poids au chapitre 6.

Mise à jour de la strate

Rarement, la croissance d'une UPE est tellement extrême qu'elle peut plus que décupler le nombre de ménages. Dans ce scénario, le sous-échantillonnage de l'UPE peut introduire des facteurs d'échantillonnage extrêmes ou être insuffisant pour réduire la charge de travail de l'intervieweur. De plus, les facteurs de sous-échantillonnage peuvent créer une forte variabilité des probabilités d'échantillonnage et peuvent entraver la précision des estimations. En pareils cas, il est préférable de remanier la strate. Normalement, d'autres UPE dans la strate auront également enregistré une importante croissance.

Pour un remaniement au niveau de la strate, les UPE de départ affichant une croissance extrême sont re-délimitées en plusieurs nouvelles UPE comportant environ 230 ménages chacune, ce qui est la taille moyenne d'une UPE. Le nombre de logements estimatif pour toutes les UPE dans la strate est nécessaire, que les UPE soient nouvelles ou qu'elles aient conservé leurs limites originales. Ces chiffres peuvent souvent être obtenus en fonction du dernier FUL. Avec ces intrants révisés, le programme de mise à jour des strates est exécuté de manière à recréer les groupes de renouvellement aléatoires et à rétablir les fractions d'échantillonnage au niveau de l'UPE. Ce programme, basé sur Keyfitz (1951) et modifié par Drew, Choudhry et Gray (1978), conserve autant des UPE sélectionnées que possible au moment de la mise à jour.

Les UPE nouvellement sélectionnés doivent être listées sur le terrain ou chargées d'information provenant du RA. Le nouvel échantillon est introduit graduellement pendant une période de six mois.

3.3.3 Surveillance du rendement de l'UPE

Au fil du temps, le rendement de l'UPE des ménages est surveillé de près. Une UPE dont le rendement des ménages est exceptionnellement faible ou fort pourrait nécessiter une attention spéciale ou un traitement particulier. Un très faible rendement des ménages porte à croire qu'un changement fondamental est survenu depuis que les comptes d'origine ont été établis en juin 2013. Un fort rendement des ménages indiquent habituellement des zones de croissance, mais il peut également indiquer que les logements sont passés à la mauvaise UPE dans le FUL. Un suivi sur place ou des enquêtes menées par le bureau central sont effectués pour justifier ou pour corriger les incohérences.

3.3.4 Stabilisation de la taille de l'échantillon

Au fil du temps, on observe une augmentation générale lente de la taille, ou de l'effectif, de la population. Si l'on ne fait rien pour la contrôler, cette croissance augmenterait la taille de l'échantillon et les frais de collecte des données d'enquête. Pour contenir la taille de l'échantillon, on peut avoir recours à la stabilisation.

Objectifs de nombre d'unités

La première étape de la stabilisation consiste à déterminer dans quelles circonstances la stabilisation est nécessaire. Les objectifs de nombre d'unités – le nombre d'unités échantillonnées nécessaires dans une région pour obtenir l'échantillon de ménages souhaité – sont déterminés. Les objectifs de nombre d'unités tiennent compte du fait que certaines unités dans la base de sondage ne seront pas nécessairement des logements valides et qu'une fraction donnée de logements valides ne sont pas occupés (c.-à-d. que ce ne sont pas des ménages). Chaque zone de stabilisation est un ensemble de strates qui correspond à peu près à une région économique de l'assurance-emploi (REAE) ou à une proportion quelconque d'une REAE. Les objectifs de nombre d'unités devraient fonctionner pour tous les renouvellements, et rarement devoir être mis à jour. Les résultats de la collecte récente indiquent généralement dans quelles circonstances des rajustements aux objectifs de nombre d'unités sont justifiés – que ce soit à cause d'une insuffisance ou d'un surplus de ménages.

Sélection de la stabilisation

Les objectifs de nombre d'unités sont comparés au nombre d'unités obtenues de l'échantillonnage à partir des listes de logements les plus à jour aux taux prescrits. Les régions nécessitant une stabilisation sont celles où l'échantillon obtenu des listes de logements les plus récentes contiennent plus d'unités que nécessaire d'après les objectifs de nombre d'unités. Le nombre d'unités à éliminer est le nombre d'unités dans l'échantillon de départ moins l'objectif de nombre d'unités.

Certaines régions sont définies sans qu'on s'attende à y éliminer des unités du fait de la petite taille de l'échantillon et de la forte variabilité relative. Les UPE à forte croissance comportant un sous-échantillonnage sont également exemptées de la stabilisation pour éviter une inflation supplémentaire des facteurs de sous-échantillonnage déjà présents. À partir des unités restantes, un sous-échantillon systématique d'unités est sélectionné pour être éliminé du processus de collecte. Les probabilités de sélection des unités non éliminées sont rajustées pour assurer une représentation adéquate de la population.

D'autres enquêtes qui sélectionnent des unités à partir de la base de sondage de l'EPA peuvent s'occuper de leur propre stabilisation, en éliminant des unités de leur échantillon de départ. On reparlera de ces enquêtes au chapitre 9.

Ajustement du poids de stabilisation

Le poids de stabilisation, utilisé pour compenser pour les logements éliminés de l'échantillon, est calculé après l'élimination. Les strates dans un secteur de stabilisation n'ont pas toutes la même FSI de strate, et le calcul de l'ajustement des poids en tient compte, en veillant à ce que les unités échantillonnées représentent bien la population.

L'exemple qui suit illustre la façon dont les facteurs de stabilisation n'ont pas d'incidence sur la contribution pondérée du secteur de stabilisation au complet.

Imaginez un secteur de stabilisation de trois strates, A, B et C, les FSI au niveau de la strate se situant à 400, 500 et 600, et le rendement d'unités avant la stabilisation se situant à 10, 10 et 10 respectivement. Supposez en outre que l'objectif de nombre d'unités pour ce secteur de stabilisation est de 28, c'est-à-dire que deux unités devraient être éliminées. Pour cet exemple, on présume qu'une unité a été éliminée de la strate A et qu'une unité a été éliminée de la strate B.

La contribution pondérée de ce secteur de stabilisation devrait se chiffrer à $15\,000 = 10 \times 400 + 10 \times 500 + 10 \times 600$. Les deux unités ayant été éliminées, la contribution pondérée devient $9 \times 400 + 9 \times 500 + 10 \times 600 = 14\,100$. Le facteur de stabilisation est tel que la contribution pondérée de ce secteur de stabilisation est conservé. Dans cet exemple, le facteur de $15\,000/14\,100$ est appliqué aux unités sélectionnées qui restent dans l'échantillon, et la contribution de ces unités à ce secteur de stabilisation est d'exactement 15 000, cet ajustement étant appliqué aux poids.

Considérations spéciales

Les logements sélectionnés sur le terrain en raison de la croissance de l'UPE sont relevés après le processus de stabilisation et n'ont donc aucune chance d'être inclus dans le programme de stabilisation. En théorie, il ne faudrait pas appliquer de poids de stabilisation à ces logements. Cependant, nos systèmes actuels attribuent les facteurs de stabilisation au niveau de la strate, et les LSI sont sujets au même facteur de stabilisation que les autres unités de la strate. L'incidence est minime, puisque le nombre de LSI en croissance est faible et que les facteurs de stabilisation sont proche de 1 ou exactement de 1. Les logements multiples, des logements comportant plusieurs unités répertoriées par erreur comme des résidences individuelles¹⁹, se font attribuer le poids de stabilisation, ce qui leur attribue en fait le poids de la résidence principale.

19. Voir la section 3.3.1 pour plus d'information.

Chapitre 4 Collecte

4.0 Introduction

Étant donné que l'Enquête sur la population active (EPA) est réalisée tous les mois, on obtient les données utilisées pour produire les différentes estimations de l'EPA en communiquant avec les ménages de l'échantillon tous les mois.

Le calendrier complet des activités de l'EPA pour un mois donné comporte quatre étapes différentes (prétraitement, collecte, traitement et diffusion), qui durent environ quatre semaines complètes au total. Par conséquent, les activités de collecte doivent suivre un échéancier serré établi conformément aux exigences des autres processus d'enquête.

La collecte des données aux fins de l'EPA a lieu tous les mois durant la semaine qui suit la semaine de référence de l'EPA, qui est habituellement la semaine où tombe le 15 du mois. Les interviews débutent le dimanche de la semaine de collecte et se poursuivent généralement jusqu'au mardi de la semaine suivante. Décembre fait exception à la règle : pour éviter de réaliser des interviews à une date trop proche de Noël, la collecte est effectuée plus tôt que les autres mois. Dans de rares cas, la collecte est prolongée d'une journée.

Une équipe d'environ 1 300 intervieweurs participent aux activités de collecte tous les mois. Les données recueillies par les intervieweurs sont transmises au bureau central à des fins de traitement.

Le présent chapitre décrit les méthodes de collecte utilisées et certaines règles d'enquête particulières liées à la collecte.

4.1 Méthodes de collecte

Pour les besoins de la collecte, il y a deux types de ménages dans l'EPA : les nouveaux ménages (les 'naissances') et les ménages subséquents (les 'subséquents'). Les naissances sont les ménages qui en sont à leur premier mois de participation à l'enquête. Ils représentent environ le sixième de l'échantillon mensuel. Les ménages subséquents se situent entre leur deuxième et leur sixième mois de participation à l'enquête. Ils représentent environ les cinq sixièmes de l'échantillon mensuel. Il convient de souligner que, pour ce qui est de la collecte, les naissances ne correspondent pas toujours au premier mois de collecte des données d'enquête pour le logement. En cas de changement complet des occupants du ménage (p. ex., en raison d'un déménagement), de non-réponse, et/ou d'un logement inoccupé au cours des mois précédents, un ménage est encore considéré comme une naissance, puisqu'il s'agit du premier mois de réponse pour ces occupants.

Depuis mars 2015, les interviews de l'EPA sont réalisées au moyen de trois méthodes de collecte : interviews sur place assistées par ordinateur (IPAO), interviews téléphoniques assistées par ordinateur (ITAO) et interviews Web assistées par ordinateur (IWAO), également appelées interviews par questionnaire électronique (QE). De mars à octobre 2015, le QE a été offert comme option à la moitié de chaque groupe de renouvellement. L'option du QE est maintenant offerte à tous les répondants admissibles depuis novembre 2015. La méthode de collecte utilisée pour un cas particulier dépend de plusieurs critères liés au genre de ménage (naissance ou subséquent), de l'admissibilité du ménage au QE et des préférences du répondant du ménage. La figure 4.1 à la fin du chapitre illustre le processus d'attribution de la méthode de collecte.

4.1.1 Méthodes de collecte pour les nouveaux ménages

Avant 2004, les IPAO étaient utilisées pour tous les ménages au premier mois de l'enquête, les intervieweurs visitant les ménages en personne pour réaliser l'interview.

Les ITAO pour les naissances ont été introduites en 2004 afin de réduire les coûts de la collecte associés à une interview initiale sur place. Cette approche, appelée premier contact par téléphone (PCT), utilise le fichier de numéros de téléphone résidentiels (FTR)²⁰ pour obtenir un numéro de téléphone pour les ménages sélectionnés. Le PCT est utilisé dans les régions où l'information utilisée pour créer le FTR est mise à jour plus souvent et où les

20. Voir le chapitre 3 pour plus d'information sur le FTR.

adresses des logements ont tendance à avoir une forme standard, qui offre une meilleure correspondance entre les adresses et les numéros de téléphone.

En 2015, la liste des strates de PCT a été allongée et comporte maintenant 977 des 1 153 strates dans les 10 provinces. De tous les ménages dans ces strates pour lesquelles un numéro de téléphone est obtenu, seulement environ 3 000 (28 % des naissances) sont attribués aux centres d'appels des ITAO pour une ITAO. Cette sélection des ménages convenant le mieux à une ITAO est basée sur des indicateurs de qualité obtenus à partir des caractéristiques de l'adresse du ménage et des variables du FTR. L'établissement à 3 000 du nombre de ménages attribués aux ITAO tous les mois aide les bureaux régionaux à planifier et à gérer les ressources.

L'approche du PCT a également été étendue pour permettre aux intervieweurs IPAO d'utiliser les numéros de téléphone filaires appariés aux ménages sélectionnés qui n'ont pas été conservés pour les ITAO. Cet élargissement englobe autant les ménages des strates avec PCT que des strates sans PCT. Les intervieweurs IPAO peuvent opter, depuis septembre 2012, d'établir un premier contact par téléphone. Tous les mois, un numéro de téléphone est fourni pour environ la moitié des ménages attribués aux intervieweurs IPAO. De ces numéros, une première tentative de contact est habituellement effectuée pour environ 60 % des ménages, et à peu près 25 % n'ont pas besoin de visites en personne pour une IPAO.

À l'heure actuelle, le QE n'est pas disponible pour les nouveaux ménages, puisque pour le moment, il n'y a pas de liste administrative d'adresses électroniques pour les ménages sélectionnés.

4.1.2 Méthodes de collecte pour les ménages subséquents

À la fin de la première interview, une série de questions est utilisée pour déterminer l'admissibilité du ménage à un QE et la méthode de collecte privilégiée par le ménage pour les mois subséquents.

Afin d'être admissible à un QE, plusieurs conditions doivent être respectées. D'abord, la première interview doit être considérée comme complète. Ensuite, la personne désignée comme la meilleure personne-ressource pour les mois subséquents doit être la personne à qui s'adresse l'intervieweur et la personne qui a fourni l'information pour le mois de naissance pour tous les membres du ménage. Aucun nom fictif ne doit avoir été fourni pour les membres du ménage. Enfin, un numéro de téléphone valide à 10 chiffres et une adresse de listage complète doivent également avoir été consignés.

Si toutes les conditions sont respectées, on demande au répondant s'il préférerait répondre à l'enquête par Internet le mois suivant, ou être contacté directement par un intervieweur de Statistique Canada. Si Internet est la méthode privilégiée, on demande au répondant de fournir une adresse électronique. Si une adresse électronique est fournie, on attribue ce ménage au QE pour les mois restants.

Si les conditions ne sont pas toutes respectées, ou si le répondant refuse de donner une adresse électronique, on propose au répondant de procéder à une ITAO le mois suivant. Si le répondant n'a pas de numéro de téléphone ou s'il indique une préférence pour une visite en personne, une IPAO est effectuée le mois suivant.

Pour les ménages répondant par IPAO ou par ITAO pour un mois subséquent, l'admissibilité au QE est réévaluée chaque mois. Si les conditions sont respectées, on offre au répondant de répondre par QE le mois suivant. De même, pour les ménages qui répondent par IPAO pour un mois subséquent et qui ne sont pas admissibles au QE, on propose de procéder à une ITAO le mois suivant.

Avant l'introduction du QE, environ 96 % des interviews subséquentes étaient menées par ITAO. Avec la mise en œuvre complète du QE, environ 20 % des interviews subséquentes seront réalisées par QE, environ 76 % par ITAO et la proportion restante de 4 % par IPAO.

4.1.3 Transferts de mode de collecte

L'EPA a une certaine capacité de transférer un cas d'un mode de collecte à un autre. Voici quatre situations qui déclencheront ce genre de transfert.

- a. La composante de contact du questionnaire est utilisée pour permettre la vérification de l'adresse du ménage. Cette vérification est primordiale pour s'assurer que le ménage contacté est celui qui occupe le logement sélectionné. Si on obtient une confirmation que le numéro de téléphone est invalide ou qu'il ne

correspond pas au logement sélectionné, d'autres sources sont consultées pour trouver un numéro de téléphone valide correspondant au ménage sélectionné. Lorsqu'on ne trouve pas de numéro valide, le cas ITAO sera transféré à un intervieweur sur le terrain, qui se rendra au logement sélectionné pour une IPAO.

- b. Lorsqu'un numéro de téléphone reste sans réponse pendant deux mois consécutifs, même si le numéro n'est pas confirmé comme invalide, le cas sera transféré à un intervieweur IPAO pour le troisième mois de l'enquête.
- c. Si le répondant demande une interview sur place, les cas ITAO peuvent être transférés à des intervieweurs sur place pendant la collecte.
- d. Si aucune réponse n'a été reçue d'un ménage attribué au QE dans les quatre premiers jours de la collecte, le cas sera transmis à un intervieweur ITAO, qui utilisera le numéro de téléphone fourni pour essayer de réaliser une interview téléphonique.

4.2 Règles d'enquête au sujet de la collecte

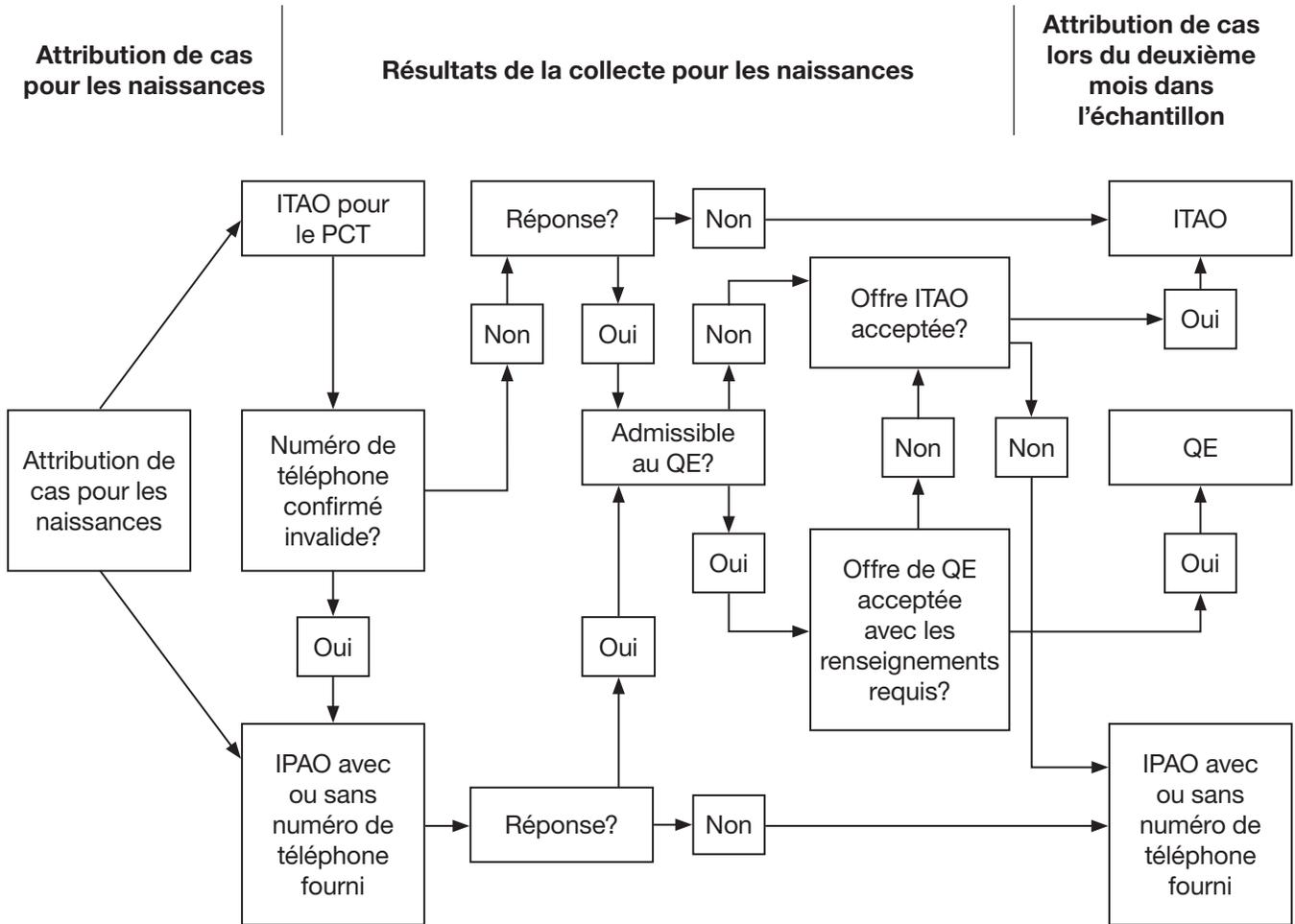
L'EPA a plusieurs règles de collecte pour alléger le fardeau du répondant et les coûts de collecte, tout en atteignant des taux de réponse et une qualité des données élevés. Les règles les plus importantes sont décrites ci-après.

Les réponses pour les membres du ménage sont habituellement fournies par un seul membre du ménage bien renseigné. Les réponses pour les membres du ménage obtenues indirectement auprès du répondant s'appellent des « réponses par personne interposée ». Elles sont permises parce qu'il serait trop coûteux en temps et en frais de faire plusieurs visites ou appels pour obtenir cette information directement auprès de chaque membre du ménage. Environ 60 % des données de l'EPA sont obtenues au moyen de cette méthode.

Pendant la première interview, l'intervieweur recueille des renseignements sociodémographiques et sur la population active concernant tous les membres du ménage sélectionné. Dans les interviews subséquentes, l'intervieweur vérifiera la liste des membres du ménage, puis il recueillera des renseignements sur la population active pour le mois en cours. Pour les personnes de 70 ans et plus, on réduit le fardeau imposé au répondant en réutilisant les réponses au sujet des renseignements sur la population active fournies à la première interview pour les mois subséquents.

Différents types de lettres sont envoyées aux ménages sélectionnés pour contribuer à maintenir un taux de réponse élevé. Par exemple, lorsqu'un ménage est sélectionné pour la première fois, une lettre de présentation et un dépliant d'information sont envoyés par la poste avant la première interview. Des lettres de refus sont également envoyées pour convaincre les ménages qui hésitent à participer.

Figure 4.1 Organigramme des attributions de méthodes de collecte pour les naissances et le deuxième mois dans l'échantillon



Chapitre 5 Traitement et imputation

5.0 Introduction

Après la collecte, les données de l'Enquête sur la population active (EPA) doivent passer par plusieurs étapes de traitement avant que des estimations soient produites. Pour faciliter la production d'estimations à partir d'un fichier de microdonnées complet et cohérent, la vérification, l'imputation et l'ajustement de poids sont utilisés pour repérer et neutraliser les données invalides, incohérences et manquantes. Le traitement des données peut être réparti en quatre étapes, que voici :

1. Réception des données des bureaux régionaux et phase I de la vérification
2. Phase II de la vérification
3. Imputation hot deck
4. Traitement post-imputation

Les données invalides et incohérentes sont repérées et remplacées par des données valides et cohérentes au moyen de vérifications et de diverses méthodes d'imputation, selon le type de non-réponse. La non-réponse partielle, où les réponses à une ou plusieurs questions du questionnaire sont inconnues, est traitée au moyen d'une imputation historique, d'une imputation par déduction ou d'une imputation hot deck, selon l'historique de réponse du répondant et les données d'enquête recueillies pour l'enregistrement. La non-réponse des personnes, où il est impossible d'obtenir des données d'enquête pour une personne, est traitée au moyen de la méthode d'imputation hot deck. La non-réponse des ménages, où il est impossible d'obtenir des données d'enquête pour le ménage au complet, est traitée au moyen de l'imputation hot deck ou d'un ajustement de poids pour la non-réponse (voir le chapitre 6), selon l'historique de réponse du ménage.

Les sections suivantes expliquent les étapes du traitement en détail, en particulier le système d'imputation hot deck.

5.1 Réception des données et phase I de la vérification

Pendant la période de collecte, les cas pour lesquels des interviews ont été complétées sont envoyés des bureaux régionaux au bureau central tous les jours. Les données sont alors traitées au bureau central. L'EPA recueille des renseignements sociodémographiques (p. ex., l'âge, le sexe, les études, le statut d'immigrant et le statut autochtone) et sur la population active (p. ex., situation d'activité, catégorie de travailleur, industrie et gains). La vérification des réponses et la vérification de la cohérence des blocs sont effectuées en plusieurs étapes. La vérification de la phase I comprend quatre étapes : l'acceptation des enregistrements, la vérification des caractéristiques démographiques, l'acceptation des réponses à la composante Information sur la population active (IPA) et la vérification de l'industrie/profession.

À l'étape de l'acceptation des enregistrements, chaque enregistrement est vérifié pour faire en sorte que toutes les composantes nécessaires ont été effectuées pendant l'interview. Il s'agit de vérifier qu'il y a des données démographiques pour chaque membre du ménage et qu'il y a des données sur la population active pour les personnes concernées d'après leur code de réponse finale, leur âge, leur appartenance au ménage, etc. Les valeurs manquantes et incohérentes de l'âge et de l'appartenance au ménage sont imputées à cette étape au moyen d'une imputation historique ou d'une imputation par déduction.

La vérification démographique met en cause une vérification détaillée des renseignements démographiques et il s'agit de la dernière étape de la vérification au niveau du ménage. À cette étape, toutes les données démographiques pour les personnes dans le ménage sont vérifiées aux niveaux individuel et familial. Une série de contrôles vérifie la validité et la cohérence des réponses aux questions pour chaque personne et entre les membres du ménage. Des corrections automatisées et manuelles peuvent être faites à cette étape.

La prochaine étape de la vérification est l'étape de l'acceptation des réponses à la composante IPA. À cette étape, chaque enregistrement fait l'objet d'un processus de prévérification pour vérifier la validité des données sur la population active reçues. Le flux des questions est vérifié pour déterminer si les réponses pour les données sur la population active suivent une voie unique, cohérente et correcte. Ce processus vérifie également la gamme et la validité des réponses dans le parcours.

La dernière étape de la vérification de la phase I est le codage des industries et des professions. Les enregistrements nécessitant un codage sont cernés et codés au moyen d'un système automatisé ou d'un système manuel lorsque le système automatisé ne peut pas attribuer un code complet. Les codes sont validés et vérifiés pour en assurer la cohérence. L'industrie est codée aux normes du Système de classification des industries de l'Amérique du Nord (SCIAN) et la profession est codée selon la norme de la Classification nationale des professions pour statistiques (CNP-S). Le SCIAN de 2012 et la CNP de 2011 sont actuellement utilisés dans le système de traitement de l'EPA.

5.2 Phase II de la vérification

La phase II de la vérification comprend la résolution des réponses « Ne sait pas » et « Refus » et la vérification détaillée de la cohérence. Pendant cette étape de la vérification, chaque enregistrement est vérifié pour déterminer s'il contient des entrées « Ne sait pas » ou « Refus ». Ces réponses sont considérées comme des cas de non-réponse partielle. La totalité de la non-réponse partielle est cernée et traitée au moyen d'une imputation par déduction ou d'une imputation historique dans la mesure du possible. Les vérifications de la cohérence sont appliquées pour veiller à ce que chaque enregistrement soit cohérent à l'interne. Si ce processus ne réussit pas, les réponses manquantes ou incohérentes sont relevées pour la méthode d'imputation hot deck.

5.3 Imputation hot deck

Dans le cas de l'imputation hot deck pour l'EPA, les valeurs manquantes d'un receveur sont remplacées par les valeurs correspondantes d'un donneur sélectionné au hasard dans la même classe d'imputation. Les classes d'imputation sont définies en fonction des variables disponibles pour les destinataires et les donneurs potentiels. Deux séries distinctes de classes d'imputation sont formées, une série de classes pour la non-réponse partielle, et une autre série pour la non-réponse des personnes et des ménages et la non-réponse partielle qui ne peut pas être résolue par une imputation des réponses manquantes.

En janvier 2005, une stratégie longitudinale d'imputation hot deck a été mise en œuvre en fonction de la recherche de Bocci et Beaumont (2004). Cette stratégie est principalement utilisée pour traiter la non-réponse des personnes et des ménages. La stratégie s'appuie sur les valeurs du mois précédent (possiblement imputées) de certaines variables, ainsi que certaines variables sociodémographiques du mois courant comme variables d'appariement pour former les classes d'imputation pour les donneurs et les receveurs. Récemment, l'efficacité de ces variables d'appariement pour le traitement de la non-réponse des personnes et des ménages a été examinée par White et Benhin (2013). L'étude a donné lieu à un ensemble amélioré de variables d'appariement, qui a été mis en œuvre en janvier 2015.

Les sous-sections suivantes décrivent les étapes de l'imputation hot deck de l'EPA plus en détail. La figure 5.1, à la fin du présent chapitre, présente une vue d'ensemble du processus du système d'imputation hot deck (SIHD). Les spécifications plus détaillées de la stratégie du SIHD se trouvent dans le document de Lorenz (1996).

5.3.1 Imputation lors du prétraitement

Avant de pouvoir effectuer l'imputation réelle des valeurs manquantes en utilisant la méthode hot deck, il faut suivre quelques étapes de prétraitement des données.

Premièrement, les réponses sont déterminées, et chaque enregistrement de réponse se voit attribuer un type d'imputation provisoire. Les données extraites du système de traitement du bureau central (STBC) sont réparties en fichiers de réponse et en fichiers de non-réponse. Les enregistrements de non-réponse seront pris en compte au moyen d'un ajustement de poids en fonction de la non-réponse, dont on reparlera au chapitre 6, et les enregistrements de réponse feront l'objet d'une imputation hot deck. Les enregistrements dans le fichier de réponse à ce stade-ci n'ont pas nécessairement répondu pendant le mois en cours. Si la personne avait répondu le mois précédent, on considère le cas comme une réponse à cette étape. Tous les enregistrements de réponse sont initialement répartis en trois groupes : A, B et C. Le groupe A contient les donneurs potentiels. Il s'agit ici de toutes les personnes pour qui les données déclarées ne contiennent aucune valeur manquante et qui sont cohérentes à l'interne. Le groupe B est constitué de toutes les personnes qui n'ont pas de valeurs manquantes et qui sont cohérentes à l'interne après la première étape de la vérification, mais qui n'appartiennent pas au

groupe A parce qu'au moins une de leurs réponses a été imputée pendant la vérification. Les personnes restantes constituent le groupe C et nécessitent une imputation.

La deuxième étape du prétraitement calcule les variables d'appariement de l'imputation. Certaines variables ne se trouvent pas initialement dans une forme qui peut être utilisée directement pour l'imputation. Par exemple, les deux variables de groupes professionnels OCC4 et OCC10 sont obtenues à partir de la variable de la CNP. Toujours à cette étape, on convertit les données sur les gains en salaire horaire en divisant les gains totaux pour la période par les heures travaillées. On s'assure ainsi qu'il y a une mesure uniforme des gains et que la valeur imputée pour les gains correspond à la valeur pour le nombre d'heures déclaré par un bénéficiaire.

La troisième étape du prétraitement est la détermination des valeurs aberrantes dans les gains et le peaufinage des groupes A, B et C. Les valeurs des gains qui sont extrêmement élevées ou extrêmement faibles sont considérées comme suspectes et sont donc réputées manquantes et imputées. Les personnes qui ont déclaré des gains très élevés ou très faibles sans être extrêmes conservent ces valeurs de gains, mais elles sont exclues des donneurs potentiels et affectées au groupe B. On forme des catégories de détection de valeurs aberrantes en croisant les variables province-sexe-groupe d'âge, et le groupe professionnel. Différentes valeurs limites basées sur la méthode du quartile sont établies dans chaque catégorie. La méthode du quartile pour la détection de valeurs aberrantes est décrite dans la publication de Statistique Canada *Méthodes et pratiques d'enquête* (2003).

Après la détection de valeurs aberrantes, les enregistrements du groupe A constituent le bassin de donneurs potentiels. Les enregistrements du groupe C sont les destinataires et seront imputés au moyen de la méthode d'imputation hot deck. Les enregistrements du groupe B n'ont pas besoin d'être imputés et ne sont pas admissibles non plus en tant que donneurs.

La dernière étape du prétraitement consiste à attribuer un chemin temporaire (TPATH) à chaque enregistrement, dans la mesure du possible. Cette variable TPATH sera utilisée comme importante variable d'appariement dans l'imputation pour la non-réponse partielle. L'utilisation de TPATH sera expliquée en détail à la prochaine section.

5.3.2 Imputation en cas de non-réponse partielle

Une fois toutes les étapes de prétraitement terminées, les valeurs manquantes peuvent être imputées. L'imputation hot deck aléatoire dans les classes est utilisée pour remplir les valeurs manquantes. Cette procédure est appliquée de façon à ce que les données des receveurs respectent les règles de contrôle de la cohérence et de la validité après l'imputation. Par exemple, les variables exigeant des valeurs non vides pour un receveur donné doivent être imputées au moyen de valeurs non vides. Dans une classe d'imputation donnée, on impute chaque receveur en sélectionnant une série de donneurs selon une méthode d'échantillonnage aléatoire simple sans remise jusqu'à ce qu'un donneur satisfaisant à toutes les règles de vérification soit trouvé. Une fois qu'un donneur approprié a été trouvé, tous les éléments manquants d'un receveur sont imputés au moyen des données de ce donneur.

Les classes d'imputation initiales sont formées par le croisement des 18 variables catégoriques suivantes :

1. TPATH (12 catégories)
2. LMLFS3 (3 catégories)
3. COW (3 catégories)
4. OCC4 (4 catégories)
5. PROV (10 catégories)
6. AGE3P3 (3 catégories)
7. ABQ1 (2 catégories)
8. IMM (3 catégories)
9. LMLFS7 (7 catégories)
10. LMINDG (20 catégories)
11. MULTJOB (2 catégories)
12. AGE3P1 (5 catégories)

13. SEX (2 catégories)
14. OCC10 (10 catégories)
15. AGE2 (8 catégories)
16. STUD (2 catégories)
17. EDUC (2 catégories)
18. DWELRENT (2 catégories).

L'ordre de ces variables indique leur importance pour expliquer les variables sur la population active comme déterminé par les études empiriques dans l'étude de White et Benhin (2013). Une description détaillée des valeurs des catégories de chaque variable se trouve à l'annexe F.

Soulignons que les variables LMLFS3, LMLFS7 et LMINDG correspondent aux valeurs du mois dernier.

La variable TPATH joue un rôle important dans le système d'imputation. Les sept premières valeurs possibles de TPATH correspondent aux sept valeurs possibles de la variable sur la situation d'activité, LFSSTAT. Chaque donneur se fait attribuer une valeur pour LFSSTAT en fonction des données déclarées. Pour les receveurs, la valeur de LFSSTAT n'est peut-être pas connue; toutefois, il peut y avoir suffisamment d'information pour exclure une ou plusieurs des sept valeurs possibles. La variable TPATH est utilisée pour veiller à ce que seules des valeurs valides pour LFSSTAT soient imputées aux receveurs en reproduisant chaque donneur par son nombre de valeurs TPATH valides et en attribuant une seule valeur de TPATH à chaque receveur. À la fin de l'étape de l'imputation, les donneurs reproduits sont supprimés. Par exemple, supposons que pour un donneur, LFSSTAT = 2. Ce donneur a donc trois valeurs TPATH valides : 2, 8 et 10 (voir l'annexe F pour la définition de toutes les valeurs TPATH possibles). Le donneur est donc reproduit trois fois, chaque répétition recevant une des trois valeurs TPATH valides. Lorsque les classes d'imputation sont formées, chaque répétition du donneur sera dans une classe d'imputation distincte.

L'imputation est effectuée dans chaque classe qui contient suffisamment de donneurs pour respecter les deux contraintes suivantes :

- i. Le nombre de donneurs doit être plus grand que le nombre de receveurs de cette classe;
- ii. Chaque classe doit comprendre un minimum de trois donneurs.

Si l'une ou l'autre de ces contraintes n'est pas satisfaite, la variable la moins importante (DWELRENT) est retirée de la liste de variables de classe d'imputation, et le processus d'imputation est réessayé pour les receveurs restants. Si après cette deuxième série d'imputation, il reste encore des receveurs n'ayant pas été imputés en raison de classes qui ne répondent pas aux deux conditions précitées, l'imputation est effectuée à une troisième reprise, en retirant la deuxième variable la moins importante (EDUC). Ce processus qui consiste à enlever une variable afin d'effectuer l'imputation continue jusqu'à ce que tous les receveurs soient imputés ou jusqu'à ce qu'il reste seulement les cinq premières variables – TPATH, LMLFS3, COW, OCC4 et PROV. Les receveurs non encore imputés à ce stade-ci sont envoyés pour l'imputation de l'enregistrement au complet, de telle sorte que toutes les variables liées à l'activité du receveur, y compris celles qui ont été déclarées, sont remplacées par celles d'un donneur sélectionné au hasard au moyen d'un ensemble différent de variables d'appariement – voir la section 5.3.3.

Dans une classe d'imputation donnée satisfaisant aux deux contraintes susmentionnées, on impute chaque receveur en sélectionnant d'abord un donneur de façon à ce que les règles de vérification de la validité soient satisfaites. En l'absence d'un tel donneur, l'enregistrement est envoyé à l'imputation de la totalité de l'enregistrement, comme le décrit la section 5.3.3. Si un donneur approprié est trouvé (c.-à-d. un donneur qui satisfait aux règles de vérification de la validité après l'imputation des valeurs manquantes du receveur), les valeurs manquantes du receveur sont remplacées par les valeurs correspondantes du donneur et les règles de contrôle de la cohérence sont vérifiées. Si toutes les règles de vérification sont satisfaites, le processus d'imputation pour ce receveur est terminé; autrement, un deuxième donneur approprié (c.-à-d. qui satisfait aux contrôles de validité) est essayé, et les règles de contrôle de la cohérence sont revérifiées. Si toutes les règles de vérification sont satisfaites après cette deuxième tentative, le processus d'imputation pour ce receveur est terminé; autrement, l'enregistrement au complet sera imputé au moyen des valeurs du dernier donneur essayé. Cette imputation pour

l'enregistrement au complet est légèrement différente du processus d'imputation décrit plus loin, parce qu'elle utilise un ensemble de variables d'appariement différent et plus précis.

5.3.3 Imputation en cas de non-réponse des particuliers

La non-réponse des personnes ou des ménages lorsque les données du mois précédent sont disponibles et la non-réponse partielle qui n'a pas pu être traitée par une imputation des valeurs manquantes sont traitées par l'imputation de l'enregistrement au complet (longitudinale). Selon cette stratégie, les données du mois précédent (possiblement imputées) pour certaines variables et données du mois courant pour d'autres variables sont utilisées pour former des classes d'imputation. Cette stratégie est également utilisée pour la non-réponse des personnes et des ménages en l'absence d'une réponse le mois précédent, mais en présence d'une réponse par le passé. Les donneurs et les receveurs pour l'imputation de l'enregistrement au complet sont tous deux des enregistrements-personnes, même lorsqu'il est question de la non-réponse du ménage au complet. Pour les ménages où cette imputation n'est pas possible, on procède plutôt à un rajustement de poids pour la non-réponse.

Les variables actuellement croisées pour former des classes d'imputation initiales pour l'imputation de l'enregistrement au complet sont indiquées ci-après par ordre d'importance :

1. PROV (10 catégories)
2. LMLFS3 (3 catégories)
3. AGE GP1 (5 catégories)
4. SEX (2 catégories)
5. LMINDG (20 catégories)
6. LMLFS7 (7 catégories)
7. REAE (56 catégories)
8. EDUC (2 catégories)
9. ABQ1 (2 catégories)

Les catégories pour ces variables sont décrites en détail à l'annexe F.

Comme dans le cas de l'imputation de la non-réponse partielle, les deux mêmes contraintes d'imputation s'appliquent aux classes d'imputation :

- i. Le nombre de donneurs doit être plus grand que le nombre de receveurs de cette classe;
- ii. Chaque classe doit comprendre au moins trois donneurs.

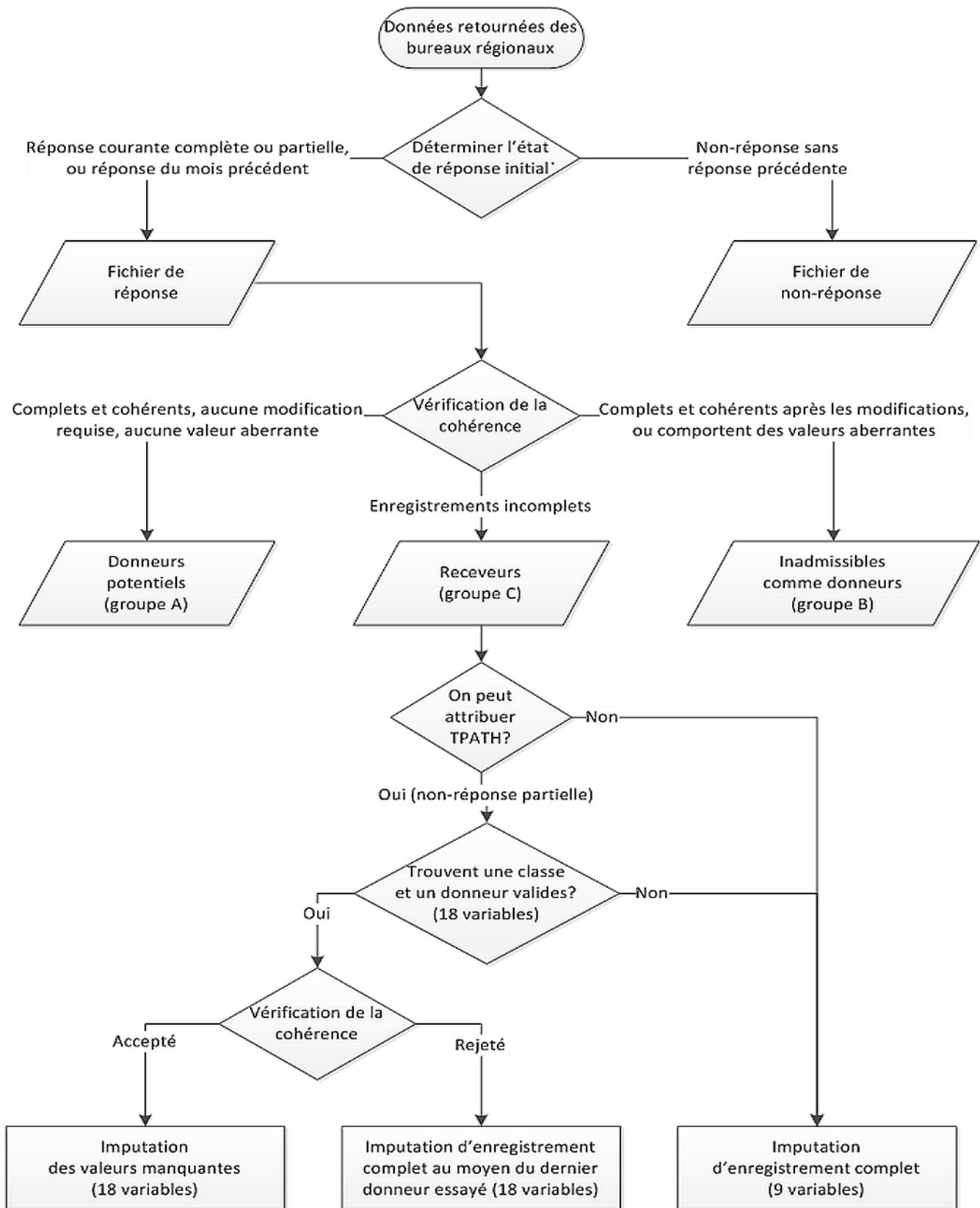
Si une de ces contraintes n'est pas satisfaite, les classes sont réduites par l'élimination de la variable la moins importante. Le processus d'élimination d'une variable et de remaniement des classes d'imputation se poursuit jusqu'à ce que tous les receveurs soient imputés.

Les enregistrements des receveurs sont imputés au moyen des valeurs des donneurs du mois en cours, même si les classes d'imputation sont basées sur les valeurs du mois courant et du mois précédent. De plus, la validité et les vérifications du contrôle de la cohérence ne sont pas nécessaires lorsque l'imputation de l'enregistrement au complet est effectuée parce que les donneurs doivent déjà satisfaire à toutes les règles de validité et de cohérence.

5.4 Processus post-imputation

Le processus post-imputation comprend l'élimination des donneurs en double qui ont été créés pendant le calcul de TPATH, l'établissement de tous les indicateurs de sortie au niveau du groupe pour indiquer que l'imputation a eu lieu, et le post-traitement de données sur les gains en calculant les gains horaires et hebdomadaires pour tous les employés en fonction des heures et des salaires déclarés ou imputés. La situation vis-à-vis de l'activité et certaines autres variables sont également obtenues.

Figure 5.1 – Organigramme simplifié du Système d'imputation hot deck de l'EPA



Chapitre 6 Pondération et estimation

6.0 Introduction

L'estimation est le processus d'enquête par lequel les estimations des paramètres de population inconnus sont produites à partir des données de l'échantillon, en combinaison possible avec des données auxiliaires provenant d'autres sources. Les paramètres de population d'intérêt comprennent notamment les totaux, les moyennes et les ratios de population, de même que leurs moyennes sur plusieurs mois d'enquête.

Les estimations de l'Enquête sur la population active (EPA) sont produites à partir des poids associés à chaque personne pour qui des données de l'EPA sont disponibles. Le présent chapitre décrit les étapes à suivre pour calculer les poids définitifs pour l'estimation. La section 6.1 décrit le calcul des poids déterminés par le plan de sondage décrit au chapitre 2. La section 6.2 décrit comment les poids déterminés par le plan de sondage sont rajustés en fonction des ménages non répondants pour devenir ce que l'on appelle des sous-poids. La section 6.3 décrit le calage composite qui est appliqué aux sous-poids par souci d'uniformité avec les estimations externes de la population, pour tenir compte du sous-dénombrement et pour améliorer l'efficacité des estimations. Cette section décrit également la méthode intégrée de pondération, qui assure un poids définitif commun pour chaque personne dans un ménage. Enfin, la section 6.4 décrit comment les poids sont utilisés pour calculer certains des principaux paramètres de population estimés par l'EPA.

6.1 Poids déterminé par le plan de sondage

Le poids déterminé par le plan de sondage d'une personne l est égal à l'inverse de sa probabilité de sélection dans l'échantillon π_l^D . On peut utiliser la formule suivante : $w_l^D = 1/\pi_l^D$. Le poids déterminé par le plan de sondage est souvent interprété comme le nombre d'unités de la population cible que représente l'unité échantillonnée. Comme chacune des personnes d'un ménage sélectionné est incluse dans l'échantillon, le calcul de la probabilité de sélection d'une personne donnée équivaut au calcul de la probabilité de sélection de son ménage.

6.1.1 Poids de base

Comme décrit à la section 2.6.2, la probabilité générale de sélection du ménage k dans l'UPE j du groupe de renouvellement i de la strate h est $\pi_{hijk}^* = 1/FSI_h^*$, pour tous les ménages de la strate h . Vous vous souviendrez que FSI_h^* est la fraction de sondage inverse arrondie pour la strate h , comme on l'a établi pendant la répartition de l'échantillon.

Dans toutes les provinces, sauf à l'Île-du-Prince-Édouard (Î.-P.-É.), l'EPA utilise un plan de sondage à deux degrés pour sélectionner des ménages. Par conséquent, le calcul des poids de base est différent pour l'Î.-P.-É. que pour le reste des provinces; cependant, étant donné que les ménages sont sélectionnés systématiquement en fonction de la FSI de la strate, la probabilité de sélection à l'Î.-P.-É. est $\pi_{hijk}^* = 1/FSI_h^*$ comme dans les autres provinces.

Étant donné que les données de l'EPA sont recueillies pour toutes les personnes admissibles dans un ménage sélectionné, la probabilité de sélection de base d'une personne l dans la strate h de n'importe quelle province est $\pi_{hijkl}^B = \pi_{hijk}^* = 1/FSI_h^*$ et son poids de base est

$$w_{hl}^B \equiv w_{hijkl}^B = 1/\pi_{hijkl}^B = FSI_h^*$$

Ce plan de sondage est autopondéré dans les strates parce qu'il a un poids de base constant dans chaque strate.

Les poids déterminés par le plan de sondage seraient équivalents aux poids de base si le plan de sondage et la population demeuraient inchangés. Cependant, comme les unités primaires d'échantillonnage (UPE) sont appelées à croître au fil du temps et que le taux d'échantillonnage systématique est fixe, la taille de l'échantillon augmenterait continuellement. Pour éviter cette situation, la taille de l'échantillon est contrôlée au moyen des procédures d'échantillonnage décrites à la section 3.3.2 : les UPE peuvent être sous-échantillonnées au moyen

de la méthode de sous-échantillonnage des UPE ou de la méthode de formation de sous-grappes; la strate peut être remaniée en fonction de l'information à jour. Ces méthodes changent la probabilité de sélection des ménages (et des personnes). Il est donc nécessaire de rajuster les poids de base pour créer des poids propres aux grappes pour tenir compte de ces procédures d'échantillonnage.

6.1.2 Poids de la grappe

Les poids de grappes sont utilisés pour les strates comportant un plan à deux degrés, c.-à-d. les strates pour toutes les provinces sauf l'Î.-P.-É. Une grappe correspond à une UPE dans ces strates. Dans les centres de population, la construction peut donner lieu à une augmentation considérable du nombre de ménages dans certaines grappes au fil du temps. Les intervieweurs sont affectés à des grappes, et si une importante croissance survient dans une ou dans plusieurs de leurs grappes, leur charge de travail augmenterait tout autant, ce qui aurait une incidence sur la qualité du travail de l'intervieweur et sa capacité de terminer la tâche. Lorsque le nombre de logements dans une grappe augmente à plus du double du niveau de départ, sans devenir trop extrême, la grappe peut être sous-échantillonnée au hasard au moyen du sous-échantillonnage de grappe (ou mécanique) ou de la formation de sous-grappes. Ces méthodes de sous-échantillonnage modifient les probabilités de sélection des ménages. Par conséquent, le poids de base w_{hl}^B est modifié par un facteur de correction de grappe a_{hl}^P pour donner le poids de la grappe

$$w_{hl}^P = w_{hl}^B a_{hl}^P.$$

Malheureusement, la propriété d'autopondération est perdue lorsque l'une ou l'autre de ces méthodes est utilisée. Des détails supplémentaires sur ces méthodes se trouvent dans le document de recherche de Kennedy (1998). Lorsque la croissance est extrême, le sous-échantillonnage risque de ne pas être pratique et on procède à la mise à jour de la strate, comme décrit ci-après.

Sous-échantillonnage de grappe

Cette méthode est la plus simple et la plus courante de toutes les méthodes de sous-échantillonnage. On modifie le taux d'échantillonnage afin de réduire le nombre de ménages sélectionnés dans la grappe. Si la grappe a initialement été échantillonnée à un taux $1/FSI_{hij}^*$ et que le sous-échantillonnage donne un taux d'échantillonnage de $1/FSI_{hij}^{**}$, le sous-poids de la grappe est $a_{hl}^P = FSI_{hij}^{**}/FSI_{hij}^*$. On multiplie le poids de base des ménages interviewés par ce facteur. Avant d'utiliser cette méthode, on doit aussi s'assurer que la croissance suffit pour justifier un facteur d'au moins 2. En raison des valeurs aberrantes auxquelles pourraient faire face les enquêtes spéciales effectuées au moyen de la base de sondage de l'EPA, la valeur maximale du facteur de correction de grappe correspond à 3.

Formation de sous-grappes

Lorsqu'une grappe subit une croissance équivalant à plus du triple de sa taille et que ses tracés de rues sont bien définis, la grappe en croissance est divisée en quatre sous-grappes ou plus. Un échantillon de deux de ces petites sous-grappes est tiré, puis un échantillon de ménages est prélevé dans chaque sous-grappe sélectionnée. Cette procédure équivaut à ajouter un autre degré d'échantillonnage dans les grappes en croissance. Elle ne modifie pas la probabilité de sélection des grappes, mais plutôt la probabilité de sélection des ménages dans les grappes en croissance. Le sous-poids d'une grappe représente ce processus de sélection.

Mises à jour de strates

Lorsque la croissance est tellement extrême que les processus de sous-échantillonnage susmentionnés sont insuffisants, une mise à jour de la strate s'impose, comme le décrit la section 3.3.2. On doit avoir des comptes mis à jour des logements dans toutes les grappes de la strate et on forme de nouvelles grappes en subdivisant

des grappes existantes dans la base de sondage à partir de ce nouveau recensement. On effectue une mise à jour de l'échantillon de la strate, selon la méthode Keyfitz (1951), modifiée par Drew, Choudhry et Gray (1978), en conservant le plus grand nombre possible des UPE sélectionnées à l'origine. Le nouvel échantillon est introduit sur une période de six mois. On applique un coefficient de pondération provisoire à toutes les UPE de la strate jusqu'à la mise en place complète. Ce coefficient de pondération permet de tenir compte des nouvelles informations provenant du plus récent compte des logements, informations autrement non reflétées dans l'échantillon actif.

6.1.3 Poids de stabilisation

Au dernier degré d'échantillonnage, on effectue un échantillonnage systématique à un taux fixe. Comme on emploie toujours le même taux de sondage, la croissance de la population et, par conséquent, du nombre de ménages, entraînerait un élargissement progressif de l'échantillon et une augmentation des coûts de sondage si aucune stabilisation de l'échantillon n'était effectuée. La stabilisation de l'échantillon consiste à sélectionner de façon aléatoire des ménages de l'échantillon afin de garder la taille de ce dernier à son niveau prévu. On effectue cette sélection aléatoire en échantillonnant systématiquement dans chaque secteur de stabilisation et indépendamment d'un secteur de stabilisation à un autre. Un secteur de stabilisation se définit comme un secteur renfermant tous les ménages appartenant à la même région économique de l'assurance-emploi (REAE) et au même groupe de renouvellement.

La stabilisation de l'échantillon modifie la probabilité de sélection des ménages. On modifie donc le poids de

la grappe $w_{hl}^P = w_{hl}^B a_{hl}^P$ par un facteur de correction de la stabilisation a_{hl}^S pour donner le poids de stabilisation $w_{hl}^S = w_{hl}^B a_{hl}^P a_{hl}^S$. Par définition, le poids déterminé par le plan de sondage d'une personne l dans la strate h , w_{hl}^D est égal à son poids de stabilisation w_{hl}^S , c.-à-d.

$$w_{hl}^D \equiv w_{hl}^S = w_{hl}^B a_{hl}^P a_{hl}^S.$$

Calcul de la correction par stabilisation

Le facteur de correction par stabilisation a_{hl}^S est calculé séparément dans les sous-secteurs. Par définition, un sous-secteur réunit toutes les strates d'un secteur de stabilisation ayant un taux d'échantillonnage commun. La pondération de stabilisation s'écarte légèrement du principe de pondération par l'inverse de la probabilité de sélection, car on l'applique dans les sous-secteurs plutôt que dans les secteurs de stabilisation. On appelle souvent cette méthode de pondération stratification a posteriori, les strates a posteriori correspondant aux sous-secteurs.

Pour simplifier, supposons un secteur de stabilisation où tous les ménages présentent une probabilité de sélection de base de 1 sur 200 au moment de la conception de l'enquête et un facteur de correction de grappe de 1. Dans cet exemple, le secteur de stabilisation n'est pas réparti en sous-secteurs. Si la taille prévue de l'échantillon du secteur de stabilisation est de 300 ménages au moment de la conception de l'enquête et si les taux d'échantillonnage utilisés produisent en fait 350 ménages, il faut retirer aléatoirement 50 ménages du secteur de stabilisation. La probabilité de sélection des ménages passe donc de 1 sur 200 à 3 sur 700 (c'est-à-dire $1/200$ multiplié par $300/350$). En multipliant le poids de base de 200 par le facteur $350/300$, on obtient le poids de stabilisation $700/3=233,333333$.

Les ménages qui présentent une des deux caractéristiques suivantes sont exclus de la stabilisation de l'échantillon et de la pondération de stabilisation :

- Les ménages qui appartiennent à une grappe sous-échantillonnée avec la méthode I ou II décrite à la section 6.1.2;
- Les ménages habitant un logement de construction récente qui a été ajouté à la liste des grappes et qu'on ne peut retirer de l'échantillon (logement sélectionné par l'intervieweur).

Comme ces ménages n'ont pas la chance d'être retirés de l'échantillon, ils sont aussi exclus de la pondération de stabilisation.

6.2 Sous-poids

Bien que l'on s'efforce d'interviewer tous les ménages dans l'échantillon sélectionné, les refus et d'autres facteurs rendent impossible de contacter certains ménages. On traite d'abord une partie de cette non-réponse des ménages au moyen d'une méthode d'imputation longitudinale (voir la section 5.3.3). On traite ensuite les ménages non répondants qui restent en les retirant du fichier et en corrigeant les poids déterminés par le plan de sondage pour les ménages répondants, y compris ceux qui ont été imputés, au moyen d'un facteur de correction de non-réponse. Le principe de base consiste à choisir un modèle qui convient aux probabilités de réponse inconnues, puis à calculer les facteurs de correction de non-réponse selon l'inverse des probabilités de réponse estimées.

Dans l'EPA, le modèle de non-réponse utilisé est le modèle uniforme de non-réponse à l'intérieur des catégories. Avec ce modèle, on présume que tous les ménages appartenant à une catégorie donnée de non-réponse c présentent la même probabilité de réponse P_c . La probabilité de réponse estimée \hat{P}_c correspond simplement au taux de réponse pondéré des ménages dans la catégorie c . Le facteur de correction de non-réponse pour une personne l appartenant à un ménage répondant dans la catégorie c correspond à $a_{cl}^{NA} = 1/\hat{P}_c$ et le poids corrigé de non-réponse ou le sous-poids correspond à la formule

$$w_{cl}^{NA} = w_{cl}^B a_{cl}^P a_{cl}^S a_{cl}^{NA} = w_{cl}^D a_{cl}^{NA}.$$

Chaque membre d'un ménage répondant donné a droit au même facteur de correction de non-réponse et donc, au même sous-poids.

6.2.1 Catégories de non-réponse

La clé pour réduire le biais de non-réponse consiste à cerner les catégories de non-réponse qui expliquent bien le mécanisme de non-réponse inconnue et dont l'élaboration justifie l'hypothèse d'une probabilité de réponse constante à l'intérieur des catégories. Sur le plan de l'efficacité, il est aussi souhaitable que les catégories de non-réponse soient les plus homogènes possible par rapport aux principales variables d'intérêt. Autrement dit, la formation des catégories devrait faire en sorte que les répondants d'une catégorie donnée s'apparentent aux non-répondants au chapitre des variables d'intérêt principales. Ainsi, les variables servant à élaborer les catégories devraient être liées au mécanisme de non-réponse et aux variables d'intérêt principales.

Dans l'EPA, chaque strate autochtone ou de revenu élevé forme une catégorie distincte de non-réponse. On définit les autres catégories par le croisement des variables PROVINCE, EIER, TYPE et ROTATION (sauf les ménages qui appartiennent à une catégorie autochtone ou de revenu élevé). La variable TYPE comporte quatre catégories et indique le type de strate à laquelle appartient un ménage : région éloignée, région rurale, région urbaine hors région métropolitaine de recensement (RMR) (y compris la strate à un degré de l'Î.-P.-É.) et RMR urbaine. La variable ROTATION correspond aux six groupes de renouvellement. Il convient de souligner que les catégories de non-réponse ne se chevauchent pas et que, réunies, elles englobent l'ensemble de la population. On procède à un regroupement des catégories lorsqu'un facteur de correction de la non-réponse est supérieur à 2 dans une catégorie donnée. Pour se faire, on élimine la dernière variable de catégorie (ROTATION) et on recalcule les facteurs de correction de la non-réponse parmi les catégories redéfinies (PROVINCE par REAE par TYPE). La catégorie problématique obtient alors le nouveau facteur de correction, ainsi que toutes les autres catégories (c.-à-d. les groupes de renouvellement) dans la même PROVINCE, la même REAE et le même TYPE. On regroupe les catégories de non-réponse afin d'éviter les facteurs importants de correction de non-réponse, puisqu'ils ont tendance à accroître la variabilité des estimations.

6.3 Poids finaux

La dernière étape du processus de pondération consiste à calculer les poids finaux, qui sont utilisés pour obtenir les estimations officielles. Le calage composite et la méthode intégrée de pondération sont utilisés pour produire

les poids définitifs. La méthode intégrée de pondération est utilisée pour assurer un poids final commun pour chaque personne du ménage.

6.3.1 Calage composite

Le calage sur marges est utilisé pour les trois raisons suivantes : assurer la cohérence avec les estimations projetées du recensement et avec toutes les enquêtes qui utilisent ces estimations du recensement; tenir compte du sous-dénombrement; améliorer l'efficacité des estimations. Afin de tenir compte de la sous-couverture et d'améliorer l'efficacité des estimations, les variables auxiliaires servant au calage doivent être corrélées avec les variables d'intérêt principales. Une façon d'atteindre cet objectif consiste à choisir les variables auxiliaires en modélisant les variables d'intérêt. Par exemple, un modèle approprié permet de constater que l'emploi ou le chômage est lié à l'âge et au sexe d'une personne.

L'EPA utilise le calage composite (ou l'estimation composite par régression) pour produire les poids finaux. Le calage composite est essentiellement la même chose que le calage sur marges, sauf que certains totaux de contrôle sont les estimations des données d'enquête du mois précédent, et les variables auxiliaires associées à ces totaux de contrôle sont imputés pour certaines unités.

Le calage composite peut donner lieu à une amélioration considérable de l'efficacité des estimations en cas de forte corrélation d'un mois à l'autre de l'information recueillie. Cet accroissement est attribuable au chevauchement de l'échantillon de l'EPA. D'un côté, on y gagne en efficacité, car le calage composite utilise l'information provenant du groupe de renouvellement sortant pour le mois précédent. D'un autre côté, on y perd en efficacité en raison des valeurs manquantes dans le groupe de renouvellement des naissances. Dans l'ensemble, on a constaté de façon empirique que le calage composite est avantageux pour l'EPA.

Comme le calage sur marges, le calage composite est une technique qui permet de déterminer les poids w_l^{CC} pour toutes les personnes dans le sous-ensemble de toutes les personnes dans l'échantillon s qui appartiennent à un ménage répondant ou imputé $l \in s_r$, le plus proche possible des sous-poids w_l^{NA} , sous réserve de certaines contraintes. En termes plus officiels, on obtient les poids de calage composite w_l^{CC} dans l'EPA en réduisant au minimum la fonction de distance

$$\sum_{l \in s_r} \frac{(w_l^{CC} - w_l^{NA})^2}{w_l^{NA}} \quad (6.1)$$

sous réserve de deux ensembles de contraintes : contraintes du calage sur marges et contraintes du calage composite.

Le premier ensemble de contraintes, les contraintes du calage sur marges, exige que les estimations basées sur les poids, w_l^{CC} , pour un vecteur des variables auxiliaires \mathbf{x} , $\hat{\mathbf{X}}^{CC} = \sum_{l \in s_r} w_l^{CC} \mathbf{x}_l$, soient égales au vecteur des chiffres de population connus, $\mathbf{X} = \sum_{l \in P} \mathbf{x}_l$. Autrement dit, les contraintes du calage sur marges peuvent être données par $\sum_{l \in s_r} w_l^{CC} \mathbf{x}_l = \mathbf{X}$. Dans l'EPA, ces chiffres de population connus, souvent appelés totaux de contrôle, sont en fait des estimations du recensement, pour le mois courant, du nombre de personnes de 15 ans et plus dans les régions économiques (RE) et les RMR/agglomérations de recensement (AR), ainsi que du nombre de personnes compris dans les 24 groupes d'âge-sexe par province. On emploie des totaux de contrôle additionnels afin de garantir que le nombre estimatif de personnes de 15 ans et plus est le même pour chaque

groupe de renouvellement. Pour procéder au calage sur marges, il faut connaître le vecteur \mathbf{x} pour chaque personne $l \in s_r$. Dans le cas de l'EPA, cela veut dire que le groupe âge-sexe, la RE et la RMR/AR de chaque personne $l \in s_r$ doivent être connus.

Le deuxième ensemble de contraintes, les contraintes de calage composite, met en cause des totaux de contrôle qui correspondent aux estimations des données d'enquête du mois précédent, ainsi que des variables auxiliaires associées à ces totaux de contrôle estimatifs. Les variables auxiliaires ne sont pas nécessairement connues pour toutes les personnes $l \in s_r$ et sont donc imputées pour certaines personnes. Ces totaux de contrôle et ces variables auxiliaires sont appelés respectivement totaux de contrôle composites et variables auxiliaires composites. Il existe 28 variables auxiliaires composites pour chaque province, et elles sont toutes définies par rapport aux données d'enquête du mois précédent (voir l'annexe G pour une liste complète).

Imputation de variables de contrôle auxiliaires

Si le vecteur des variables auxiliaires composites pour l'unité l , indiqué par $\mathbf{z}_{t-1,l}$, est défini pour le mois précédent (le mois $t-1$), le vecteur correspondant des totaux de contrôle estimatifs, indiqué par $\hat{\mathbf{Z}}$, doit également être calculé au moyen des données du mois précédent. Le vecteur des variables auxiliaires composites \mathbf{z}_{t-1} n'est pas observé pour les personnes appartenant au groupe de renouvellement naissant, puisqu'elles n'ont pas été interviewées au cours du mois précédent. L'imputation est utilisée pour remplir les valeurs manquantes pour ces unités au moyen d'un agencement de deux méthodes d'imputation.

Dans la première méthode, l'imputation par la moyenne est utilisée pour obtenir le vecteur modifié :

$$\mathbf{z}_{\bullet,l}^{(1)} = \begin{cases} \mathbf{z}_{t-1,l} & \text{if } l \in s_r - s_r^b \\ \hat{\mathbf{Z}}/N_{15+} & \text{if } l \in s_r^b \end{cases}$$

où s_r^b est le sous-ensemble de personnes $l \in s_r$ qui appartiennent au groupe de renouvellement naissant et N_{15+} est le nombre provincial de personnes de 15 ans et plus. Dans une étude empirique précédente, on a pu démontrer que cette méthode d'imputation permettait d'évaluer de façon efficace les paramètres de population définis pour le mois courant t .

Dans la deuxième méthode d'imputation, le vecteur modifié $\mathbf{z}_{\bullet,l}^{(2)}$ est défini comme suit :

$$\mathbf{z}_{\bullet,l}^{(2)} = \begin{cases} \mathbf{z}_{t-1,l} + (\delta_l^{-1} - 1)(\mathbf{z}_{t-1,l} - \mathbf{z}_{t,l}) & \text{if } l \in s_r - s_r^b \\ \mathbf{z}_{t,l} & \text{if } l \in s_r^b \end{cases},$$

où $\mathbf{z}_{t,l}$ est le vecteur $\mathbf{z}_{t-1,l}$ défini pendant le mois en cours t et δ_l est la probabilité que $l \in s_r - s_r^b$ étant donné que $l \in s_r$. Dans l'EPA, $\delta_l = 5/6$, pour $l \in s_r$, et cette valeur est remplacée dans l'équation précédente par l'estimation $\hat{\delta}_l = \sum_{l \in s_r - s_r^b} w_l^{NA} / \sum_{l \in s_r} w_l^{NA}$. Essentiellement, il s'agit d'effectuer une imputation rétrospective (imputation à partir des valeurs du mois courant pour produire les valeurs du mois précédent) afin d'imputer \mathbf{z}_{t-1} pour le groupe de renouvellement naissant, car on sait qu'il existe une forte corrélation d'un mois à l'autre pour les variables auxiliaires composites. Toutefois, les valeurs de \mathbf{z}_{t-1} hors du groupe de renouvellement naissant sont modifiées étant donné que l'imputation rétrospective élimine tout changement pour les personnes appartenant au groupe de renouvellement naissant. On détermine la correction hors du groupe de renouvellement naissant de manière à préserver l'absence de biais asymptotique qui caractérise les estimations. Dans une étude empirique

précédente, on a constaté que cette méthode d'imputation (qui permet d'en arriver à $\mathbf{z}_{\bullet l}^{(2)}$) permettait d'évaluer de façon efficace les paramètres de population définis comme des différences entre deux mois successifs.

Comme mentionné, ni $\mathbf{z}_{\bullet l}^{(1)}$ ni $\mathbf{z}_{\bullet l}^{(2)}$ ne sont utilisés dans l'enquête. Au lieu de cela, une combinaison des deux méthodes est utilisée. Les variables auxiliaires composites sont définies comme suit

$$\mathbf{z}_{\bullet l} = (1 - \alpha)\mathbf{z}_{\bullet l}^{(1)} + \alpha\mathbf{z}_{\bullet l}^{(2)},$$

où α est une constante qui équivaut à 2/3. On parvient ainsi à un compromis entre les deux méthodes d'imputation. Chen et Liu (2002) ont effectué une étude sur le choix de α . Bocci et Beaumont (2005) ont également étudié d'autres méthodes d'imputation qui font appel à l'imputation calée.

On obtient donc les poids de calage composite de l'EPA w_l^{CC} en réduisant au minimum la fonction de distance donnée par l'équation (6.1), sous réserve des deux ensembles de contraintes

$$\sum_{l \in s_r} w_l^{CC} \begin{pmatrix} \mathbf{x}_l \\ \mathbf{z}_{\bullet l} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \hat{\mathbf{Z}} \end{pmatrix}$$

La réduction au minimum entraîne les poids de calage composite $w_l^{CC} = w_l^{NA} g_l^{CC}$ où le facteur de correction par calage composite g_l^{CC} est donné par

$$g_l^{CC} = (\mathbf{x}'_l, \mathbf{z}'_{\bullet l}) \left(\sum_{l \in s_r} w_l^{NA} (\mathbf{x}'_l, \mathbf{z}'_{\bullet l}) (\mathbf{x}'_l, \mathbf{z}'_{\bullet l})' \right)^{-1} (\mathbf{X}', \hat{\mathbf{Z}})'$$

Vous pouvez trouver d'autres renseignements sur le calage composite de l'EPA dans les rapports de Singh, Kennedy et Wu (2001), de Fuller et Rao (2001) et de Gambino, Kennedy et Singh (2001). Gambino, Kennedy et Singh (2001) traitent aussi des questions liées aux personnes manquantes ou inadmissibles pour le mois précédent, hors du groupe de renouvellement naissant. On impute les valeurs manquantes au moyen de la méthode d'imputation par hot deck aléatoire et on applique la formule $\mathbf{z}_{\bullet l} = \mathbf{0}$ aux personnes inadmissibles le mois précédent. Il s'agit de déterminer $\mathbf{z}_{\bullet l}$ de sorte que $\sum_{l \in s_r} w_l^{NA} \mathbf{z}_{\bullet l}$ demeure, comme $\hat{\mathbf{Z}}$, une estimation du vecteur inconnu des totaux de contrôle \mathbf{Z} , laquelle est définie pour le mois précédent. On traite de la façon habituelle les valeurs pour les personnes manquantes ou inadmissibles pour le mois courant.

6.3.2 Méthode de pondération intégrée

Étant donné que certaines variables auxiliaires et toutes les variables auxiliaires composites sont définies à l'échelle de la personne, les poids de calage composite w_l^{CC} ne sont pas constants dans un ménage, contrairement aux sous-poids w_l^{NA} . Cela ne pose aucun problème tant qu'on cherche à évaluer les paramètres de population liés aux personnes, comme le nombre total de personnes ayant un emploi au sein de la population. Toutefois, dans l'EPA, on s'intéresse parfois aussi à l'estimation de paramètres de population liés aux ménages. Par exemple, on pourrait chercher à évaluer le nombre total de ménages ayant une certaine caractéristique, notamment la présence d'au moins un membre occupant un emploi. Il existe plus d'une possibilité de pondération pour ces paramètres de population.

Pour éviter de produire deux ensembles de poids finaux, on a adopté dans l'EPA la méthode de pondération intégrée qui permet d'obtenir un seul ensemble de poids finaux pour les paramètres de population liés tant aux

personnes qu'aux ménages; voir Lemaître et Dufour (1987). Avec cette méthode, le poids de calage composite final est constant pour tous les membres du ménage. Il s'agit de remplacer \mathbf{X}_l et $\mathbf{Z}_{\bullet l}$ pour une personne donnée l par la moyenne de \mathbf{x} et \mathbf{Z}_{\bullet} pour tous les membres du même ménage, puis de calculer les poids de calage ou les poids de calage composite conformément à la section 6.3.1. On garantit ainsi un poids final commun pour tous les membres du ménage. Cette contrainte additionnelle sur les poids finaux devrait réduire l'efficacité des estimations. Cependant, Pandey, Alavi et Beaumont (2003) ont constaté de façon empirique que la réduction de l'efficacité est faible dans le contexte de l'EPA.

6.3.3 Traitement des poids négatifs et arrondissement

Il arrive que le calage sur marges donne des poids négatifs. En pareil cas, un calage composite est réeffectué sur les poids post-calage, les poids négatifs étant réinitialisés à leurs sous-poids. S'il subsiste encore des poids finaux négatifs après ce second calage composite, on attribue à ces poids négatifs une valeur égale à 1 et on accepte que la contrainte de calage composite ne soit pas respectée. Cette situation se produit rarement. Après les deux séries de calage composite, le poids est arrondi à l'entier le plus proche, pour produire le poids final.

6.4 Estimation

Une fois calculés, les poids finaux sont utilisés pour estimer plusieurs types de paramètres d'intérêt d'une population, y compris les exemples suivants de totaux, de taux et de moyennes mobiles.

Chaque mois, l'EPA calcule le nombre de personnes occupées dans la population. Si y_l est une variable binaire indiquant si une personne donnée l de la population occupe un emploi ($y_l = 1$) ou pas ($y_l = 0$), le chiffre de population Y représente le nombre de personnes occupées dans la population P . Le chiffre de population est calculé comme suit

$$Y = \sum_{l \in P} y_l$$

Au moyen des poids finaux, ce chiffre de population peut être estimé comme suit

$$\hat{Y}^{CC} = \sum_{l \in s_r} w_l^{CC} y_l$$

où s_r est le sous-ensemble de toutes les personnes de s appartenant à un ménage répondant ou imputé et w_l^{CC} est le poids de calage composite, ou le poids final, associé à la personne l .

L'EPA calcule également le taux de chômage chaque mois. Si y_{1l} est une variable binaire indiquant si une personne donnée l de la population est en chômage ($y_{1l} = 1$) ou pas ($y_{1l} = 0$) et y_{2l} est une variable binaire indiquant si la personne l fait partie de la population active ($y_{2l} = 1$) ou pas ($y_{2l} = 0$), le taux de population r_{y_1, y_2} représente le taux de chômage au sein de la population.

$$r_{y_1, y_2} = \frac{\sum_{l \in P} y_{1l}}{\sum_{l \in P} y_{2l}}$$

Il est possible de l'estimer à partir des poids finaux w_l^{CC} comme suit

$$\hat{r}_{y_1, y_2}^{CC} = \frac{\sum_{l \in s_r} w_l^{CC} y_{1l}}{\sum_{l \in s_r} w_l^{CC} y_{2l}}$$

De plus, tous les mois, l'EPA produit des estimations de la moyenne mobile sur trois mois des taux de chômage pour chaque REAE à partir des données des trois derniers mois. Si la moyenne mobile sur T mois d'un total Y au temps t est

$$\theta_t^Y = \sum_{q=0}^{T-1} \frac{Y_{t-q}}{T}$$

et que l'estimation repose sur les poids finaux au moyen de la formule

$$\hat{\theta}_t^Y = \sum_{q=0}^{T-1} \frac{\hat{Y}_{t-q}^{CC}}{T}$$

alors la moyenne mobile estimative sur trois mois pour le taux de chômage peut être calculée comme suit

$$\begin{aligned} \hat{r}_{\theta_t^{Y_1}, \theta_t^{Y_2}} &= \frac{\hat{\theta}_t^{Y_1}}{\hat{\theta}_t^{Y_2}} = \frac{\sum_{q=0}^2 \frac{\hat{Y}_{1,t-q}^{CC}}{3}}{\sum_{q=0}^2 \frac{\hat{Y}_{2,t-q}^{CC}}{3}} \\ &= \frac{\sum_{q=0}^2 \hat{Y}_{1,t-q}^{CC}}{\sum_{q=0}^2 \hat{Y}_{2,t-q}^{CC}} \end{aligned}$$

Les moyennes mobiles estimatives sont utilisées parce qu'elles sont plus stables que les estimations mensuelles; toutefois, leur interprétation est différente, puisqu'elles estiment un paramètre de population différent.

Chapitre 7 Estimation de la variance

7.0 Introduction

Dans une enquête basée sur un échantillon probabiliste comme l'Enquête sur la population active (EPA), les inférences statistiques doivent tenir compte de l'erreur d'échantillonnage. La variance mesure la précision d'un estimateur. Compte tenu de la complexité de la méthode d'estimation et du plan de sondage, une forme explicite de l'estimateur de la variance n'est pas facile à obtenir pour l'EPA. L'enquête utilise donc une méthode de rééchantillonnage pour l'estimation de la variance.

Pour le remaniement de 2015, on a apporté un important changement à la méthode d'estimation de la variance de l'EPA. Auparavant, l'estimation de la variance était basée sur une méthode de rééchantillonnage appelée la méthode du jackknife. Un système d'estimation de la variance conçu sur mesure pour l'EPA utilisait la méthode du jackknife pour produire les estimations de la variance des totaux, des taux ou des proportions, des changements et des moyennes mobiles. Depuis janvier 2015, l'estimation de la variance est basée sur une méthode de rééchantillonnage appelée la méthode bootstrap. Chaque mois, 1 000 ensembles de poids bootstrap de l'EPA sont produits, et ces poids bootstrap peuvent être utilisés avec différents progiciels standard pour produire les estimations de la variance. Les estimations de la variance obtenues au moyen de la nouvelle méthodologie ont une valeur similaire à celle des estimations obtenues au moyen de l'ancienne méthodologie. Le principal avantage de la nouvelle méthodologie est qu'une fois que les poids bootstrap sont générés, ils peuvent être utilisés pour produire les estimations de la variance pour un bien plus large éventail d'analyses que l'ancien système.

Le présent chapitre décrira comment la variance est estimée pour l'EPA. La section 7.1 présente la méthode bootstrap particulière qui est mise en œuvre, la méthode bootstrap de Rao-Wu. Les sections 7.2 et 7.3 décrivent comment les échantillons bootstrap de l'EPA et les poids bootstrap sont générés. La section 7.4 explique comment les poids bootstrap sont utilisés pour calculer les estimations de la variance.

7.1 La méthode bootstrap de Rao-Wu

L'EPA s'appuie sur la méthode bootstrap de Rao-Wu, comme le proposait les publications de Rao et Wu (1988) et de Rao, Wu et Yue (1992). La méthode a été proposée pour les plans stratifiés à plusieurs degrés, où les unités primaires d'échantillonnage (UPE) sont sélectionnées au moyen de la méthode d'échantillonnage où la probabilité est proportionnelle à la taille avec remise (PPTAR). Pour l'EPA, les UPE sont en fait sélectionnées au moyen de la méthode d'échantillonnage où la probabilité est proportionnelle à la taille sans remise (PPTSAR). La publication de Särndal, Swensson et Wretman (1992, p. 154) énonce que l'estimateur de la variance pour l'échantillonnage à plusieurs degrés où les UPE sont sélectionnées sans remise peut être estimé par l'estimateur de la variance pour l'échantillonnage à plusieurs degrés, les UPE étant sélectionnées avec remise, et que l'estimation est modérée si la sélection des UPE sans remise est plus efficace que la sélection d'UPE avec remise. C'est le cas pour l'EPA.

La première étape pour appliquer la méthode bootstrap de Rao-Wu consiste à sélectionner des échantillons bootstrap. Pour chaque strate h , m_h UPE sont obtenues au moyen de la méthode d'échantillonnage aléatoire simple avec remise (EASAR) à partir de l'ensemble initial de n_h UPE échantillonnées. Pour la plupart des applications de la méthode bootstrap de Rao-Wu à Statistique Canada, y compris l'EPA, m_h est établi à $n_h - 1$. Ce processus de sélection des échantillons bootstrap a lieu à B reprises. Le nombre de fois que la j^{e} UPE est sélectionnée dans l'échantillon bootstrap de la b^{e} répétition, appelé multiplicité de l'UPE, est indiqué par $m_{hj}^{(b)}$, où $b=1, \dots, B$. Les multiplicités, $m_{hj}^{(b)}$, ont des valeurs allant de 0 à $n_h - 1$ inclusivement, et sont conformes à l'équation $\sum_{j=1}^{n_h} m_{hj}^{(b)} = n_h - 1$ pour chaque répétition bootstrap et chaque strate.

La prochaine étape consiste à produire B ensembles de poids bootstrap en appliquant un facteur de correction au poids d'enquête de départ. Ils sont calculés comme suit :

$$w_{hjk}^{(b)} = \frac{n_h}{n_h - 1} m_{hj}^{(b)} w_{hjk}, \quad (7.1)$$

où w_{hjk} correspond au poids d'enquête pour l'unité k dans l'UPE j et la strate h , et $w_{hjk}^{(b)}$ représente le poids bootstrap pour la b^{e} répétition.

Les B ensembles de poids bootstrap peuvent être utilisés pour produire des estimations de variance aux fins de diverses analyses. Pour une estimation, $\hat{\theta}$, d'un paramètre de population, θ , l'estimation de variance bootstrap est calculée comme suit. L'estimation est calculée au moyen de chaque ensemble de poids bootstrap, ce qui donne lieu à B estimations représentées comme $\hat{\theta}^{*(1)}, \dots, \hat{\theta}^{*(B)}$. Par exemple, supposons que $\hat{\theta}$ est une estimation d'un total, donnée par $\hat{\theta} = \sum_h \sum_j \sum_k w_{hjk} y_{hjk}$, où y_{hjk} est la valeur d'une variable d'intérêt y pour l'unité k dans l'UPE j et la strate h . Par conséquent, l'estimation pour la b^{e} répétition bootstrap est $\hat{\theta}^{*(b)} = \sum_h \sum_j \sum_k w_{hjk}^{(b)} y_{hjk}$. L'estimation de variance bootstrap est donnée par la variance des B estimations

$$\hat{V}_{\text{BOOT}}(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^B (\hat{\theta}^{*(b)} - \hat{\theta}^{*(\cdot)})^2 \quad (7.2)$$

$$\text{où } \hat{\theta}^{*(\cdot)} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^{*(b)}.$$

7.2 Échantillons bootstrap de l'EPA

Pour obtenir des estimations stables de la variance pour différents types d'analyses, le plus grand nombre possible de répétitions bootstrap devraient être rendues disponibles. Il faut en venir à un compromis entre la garantie de stabilité, la réduction du délai d'exécution et la taille des fichiers. L'EPA a opté pour générer 1 000 répétitions bootstrap de l'EPA chaque mois. On assure ainsi la stabilité des estimations de la variance pour les estimations clés de l'enquête.

Comme le décrit la section 7.1, la première étape pour appliquer la méthode bootstrap de Rao-Wu consiste à tirer 1 000 échantillons bootstrap au niveau de l'UPE, $n_h - 1$ UPE étant sélectionnées avec remise par strate. Un plan de sondage à deux degrés est utilisé dans toutes les provinces sauf à l'Île-du-Prince-Édouard (Î.-P.-É.), et les échantillons bootstrap sont donc sélectionnés au niveau de la grappe. Étant donné qu'un plan de sondage à un degré est utilisé à l'Î.-P.-É., les échantillons bootstrap sont sélectionnés au niveau du logement.

Le reste de la section donne des détails sur divers facteurs à considérer au sujet de la génération d'échantillons bootstrap de l'EPA. Vient ensuite la section 7.3, qui décrit la génération des 1 000 ensembles de poids bootstrap de l'EPA.

7.2.1 Les strates à une seule UPE sélectionnée

Pour estimer la variance, chaque strate devrait contenir au moins deux UPE échantillonnées. C'est habituellement le cas pour l'EPA, et c'est toujours le cas pour les strates à un degré à l'Î.-P.-É. La plupart des strates à deux degrés dans les provinces contiennent six UPE sélectionnées, une pour chaque renouvellement. Cependant, pour diverses raisons, certaines strates peuvent avoir une seule UPE échantillonnée dans le fichier final de totalisation. Cette situation peut se produire en raison du plan (petit nombre de strates à trois degrés dans le plan précédent, transition entre remaniements), ou des résultats de l'enquête (logements hors du champ de l'enquête et non répondants). Les strates à une seule UPE sont traitées d'une des trois façons suivantes.

Premièrement, on traite les strates à trois degrés dans les provinces où une seule UPE a été sélectionnée en fractionnant l'UPE sélectionnée. L'UPE est fractionnée en fonction du groupe de renouvellement ou de l'unité secondaire d'échantillonnage (USE). Pour ces strates, les échantillons bootstrap sont sélectionnés au niveau du groupe de renouvellement ou de l'USE au lieu du niveau de l'UPE.

Deuxièmement, on traite les strates à une seule UPE qui surviennent pendant la période de transition du remaniement en regroupant les strates; on en reparlera plus en détail à la section 7.2.3.

Enfin, on traite les strates restantes à une UPE en fractionnant temporairement l'UPE en deux parties, selon que l'identificateur du ménage est pair ou impair. La stratégie a été choisie parce qu'elle est facile à mettre en œuvre et qu'elle n'exige aucune intervention manuelle. Cette situation se produit suffisamment rarement pour que la stratégie utilisée n'ait pas d'incidence sur les estimations de la variance à l'échelon provincial.

7.2.2. Coordination de l'échantillon bootstrap

L'EPA produit des estimations mettant en cause plusieurs mois d'enquête, comme des estimations du changement entre des périodes et des moyennes mobiles. Le chevauchement de l'échantillon et la dépendance qui existent d'un mois à l'autre peuvent être pris en considération dans l'estimation de la variance au moyen de la méthode bootstrap coordonnée proposée par Roberts, Kovacevic, Mantel et Phillips (2001). Cette méthode tient compte de la dépendance en conservant les mêmes échantillons bootstrap des UPE d'un mois à l'autre.

En pratique, les UPE échantillonnées dans une strate ne sont pas toujours les mêmes d'un mois à l'autre, et la méthode bootstrap coordonnée doit être adaptée. Une stratégie est proposée dans Neusy (2013) et Benhin et Mantel (2012) pour adapter la méthode bootstrap coordonnée en présence de changement. Il existe potentiellement quatre situations différentes :

- i. Lorsque les UPE échantillonnées dans la strate sont les mêmes pour le mois courant et le mois précédent, l'échantillon bootstrap du mois précédent peut être utilisé pour le mois courant sans aucun travail supplémentaire.
- ii. Lorsque les UPE ne sont pas toutes les mêmes, mais que le nombre d'UPE échantillonnées dans la strate demeure le même pendant les deux mois, on peut mettre en œuvre la méthode bootstrap coordonnée en appariant chaque UPE du mois courant avec une UPE du mois précédent. Les UPE qui se trouvent dans les échantillons des deux mois sont appariées; les nouvelles UPE remplaçant les UPE éliminées sont appariées avec l'UPE qu'elles remplacent; et toutes les UPE restantes sont appariées au hasard. On obtient les échantillons bootstrap du mois courant pour la strate en transférant les multiplicités du mois précédent au mois en cours : chaque UPE du mois courant reçoit les multiplicités de l'UPE du mois précédent avec laquelle elle est appariée. Il s'ensuit un échantillon bootstrap de Rao-Wu comportant les mêmes multiplicités le mois courant que le mois précédent pour les UPE qui se trouvent dans les deux mois.
- iii. Lorsqu'il y a moins d'UPE échantillonnées dans la strate pour le mois courant que pour le mois précédent, la méthode bootstrap coordonnée est adaptée comme suit. Chaque UPE du mois courant est d'abord appariée avec une UPE du mois précédent comme décrit à ii, ce qui laisse une ou plusieurs des UPE du mois précédent non appariées. Chaque UPE du mois courant reçoit les multiplicités de l'UPE du mois précédent avec laquelle elle est appariée, ce qui donne lieu à des échantillons bootstrap provisoires pour le mois courant. La somme des multiplicités pour les échantillons bootstrap provisoires n'est pas nécessairement $n_h - 1$ pour toutes les répétitions bootstrap. Ce phénomène est attribuable au fait que les multiplicités des UPE non appariées du mois précédent ne sont pas reportées au mois courant et que n_h est plus petit qu'il ne l'était le mois précédent. Pour les répétitions bootstrap où la somme est inférieure à $n_h - 1$, les UPE sont ajoutées au hasard à l'échantillon bootstrap au moyen de la méthode d'EASAR (c.-à-d. que les multiplicités des UPE sont accrues) jusqu'à ce que la somme des multiplicités soit $n_h - 1$. À l'inverse, pour les répétitions bootstrap où la somme est supérieure à $n_h - 1$, les UPE sont éliminées au hasard de l'échantillon bootstrap (c.-à-d. que les multiplicités des UPE sont diminuées) jusqu'à ce que la somme des multiplicités soit $n_h - 1$.
- iv. Lorsqu'il y a plus d'UPE échantillonnées dans la strate pour le mois courant que pour le mois précédent, une étape supplémentaire est requise pour adapter la méthode bootstrap coordonnée. Les UPE du mois courant sont appariées, le plus possible, avec les UPE du mois précédent comme décrit à ii. Le mois courant a plus d'UPE échantillonnées que le mois précédent, ce qui fait que les UPE du mois courant ne peuvent pas toutes être appariées. Les UPE appariées du mois courant reçoivent leurs multiplicités des UPE du mois précédent avec lesquelles elles sont appariées. Les multiplicités

des UPE non appariées du mois courant (nouvelles UPE) sont générées au moyen de la distribution Binomial $\left(n_h^* - 1, 1/n_h^*\right)$, où n_h^* est le nombre d'UPE échantillonnées au cours du mois précédent. On garantit ainsi que les multiplicités prévues des UPE non appariées sont les mêmes que les UPE appariées. Les multiplicités pour les UPE appariées et non appariées du mois courant forment collectivement les échantillons bootstrap provisoires pour le mois courant. La somme des multiplicités pour les échantillons bootstrap provisoires n'est pas nécessairement $n_h - 1$ pour toutes les répétitions bootstrap. Les UPE sont ajoutées ou éliminées de façon aléatoire dans les échantillons bootstrap, comme décrit à iii, jusqu'à ce que la somme des multiplicités soit $n_h - 1$ pour tous les échantillons bootstrap.

Les stratégies pour traiter les augmentations ou les diminutions du nombre d'UPE échantillonnées décrites à iii et à iv maintiennent les bonnes estimations transversales de la variance et offrent une certaine coordination des estimations de la variance mettant en cause plusieurs mois.

Pour l'EPA, la coordination pour les strates à deux et à trois degrés dans les provinces est mise en œuvre comme suit. Les échantillons bootstrap de l'EPA sont basés sur les UPE se trouvant dans le fichier de totalisation final du mois courant, et le nombre d'UPE dans la plupart des strates demeure inchangé d'un mois à un autre. Autrement dit, la coordination décrite à i et à ii est utilisée le plus souvent. Cependant, le nombre d'UPE peut fluctuer, habituellement parce que temporairement, une UPE n'a pas de répondants dans le fichier de totalisation final. Si le nombre d'UPE a diminué de 1, l'adaptation en fonction de la méthode bootstrap coordonnée décrite à iii est utilisée. Si le nombre d'UPE diminue de plus que 1 ou s'il augmente, de nouveaux échantillons bootstrap sont sélectionnés au hasard au moyen d'un germe aléatoire fixe attribué à chaque strate de l'EPA et conservé jusqu'au prochain remaniement. Ces germes aléatoires fixes sont utilisés de façon à ce que les mêmes échantillons bootstrap soient sélectionnés pour une strate donnée et le nombre d'UPE.

Depuis le remaniement de 2015, l'Î.-P.-É. utilise un plan de sondage à un degré, et les UPE sont au niveau du logement. Comme décrit à la section 2.5.6, les strates de l'Î.-P.-É. ont été formées en fonction des aires de diffusion (AD) du recensement, puis attribuées à un des six groupes de renouvellement. Tous les logements dans la même strate à un degré appartiennent au même groupe de renouvellement. Tous les six mois, lorsqu'un nouvel échantillon de logements est intégré à une strate, de nouveaux échantillons bootstrap sont également sélectionnés pour cette strate. Pour les cinq autres mois, les échantillons bootstrap sont coordonnés au moyen des stratégies décrites à i, ii, iii et iv, selon la situation.

7.2.3. Période de transition du remaniement

L'échantillon de l'EPA a été remanié en janvier 2015, et le nouvel échantillon a été mis en place graduellement de janvier à juin 2015. Chaque mois de la période de transition, un groupe de renouvellement de l'ancien plan a été éliminé et remplacé par un groupe de renouvellement du nouveau plan. Au bout du compte, les estimations sont basées sur un amalgame des deux plans. Le nouvel échantillon de l'EPA a été sélectionné indépendamment de l'ancien échantillon, de manière à ce que les échantillons bootstrap pour le nouveau plan soient également sélectionnés indépendamment et séparément des échantillons bootstrap pour l'ancien plan, c.-à-d. sans coordination.

À mesure que l'ancien échantillon était éliminé graduellement pendant la période de transition, le nombre d'UPE échantillonnées dans les strates de l'ancien plan diminuait tous les mois. Les échantillons bootstrap pour l'ancien plan étaient coordonnés pendant cette période au moyen de la stratégie de coordination décrite au point iii de la section 7.2.2. Au cinquième mois de la période de transition, un seul groupe de renouvellement de l'ancien plan de sondage était toujours dans l'échantillon de l'EPA, ce qui fait que bien des strates n'avaient qu'une seule UPE. Les strates à une seule UPE ont été appariées au hasard dans la province et regroupées de manière à former des strates à deux UPE. Des échantillons bootstrap provisoires pour les strates regroupées ont été générés au moyen des multiplicités des mois précédents de chaque UPE. Chaque strate regroupée comportait deux UPE, ce qui fait que les UPE ont été ajoutées ou éliminées au hasard dans les échantillons bootstrap jusqu'à ce que la somme des multiplicités soit de 1 pour tous les échantillons bootstrap de chaque strate regroupée.

On a créé les échantillons bootstrap pour le nouvel échantillon en générant d'abord des échantillons bootstrap pour l'échantillon de juin 2015, où le nouvel échantillon était entièrement en place. Les échantillons bootstrap ont été générés au moyen d'un nouvel ensemble de germes aléatoires qui sera conservé jusqu'au prochain

remaniement. Ensuite, on a généré les échantillons bootstrap pour le nouveau plan en remontant de mai à janvier 2015. Lorsque l'on retourne en arrière, le nombre d'UPE échantillonnées diminue chaque mois. La même méthodologie utilisée pour coordonner les échantillons bootstrap pour l'ancien plan en avançant dans le temps a servi à coordonner les échantillons bootstrap pour le nouveau plan en remontant dans le temps. Comme l'échantillon de janvier 2015 renfermait un seul groupe de renouvellement du nouvel échantillon, bien des strates du nouveau plan renfermaient une seule UPE. Les strates à une seule UPE ont été regroupées et traitées, comme susmentionné pour les strates de l'ancien plan au cinquième mois de la transition.

7.3 Poids bootstrap de l'EPA

Afin de bien estimer la variabilité d'échantillonnage d'un estimateur, chacune des étapes de la pondération menant au calcul des poids finaux devrait être répétée pour chaque répétition bootstrap. À l'heure actuelle, seule la dernière étape de la pondération, le calage composite (voir la section 6.3.1), est répétée pour chaque répétition bootstrap. C'était également le cas pour le système d'estimation de la variance précédent d'après la méthode du jackknife.

Les étapes suivantes sont suivies pour générer 1 000 ensembles de poids bootstrap finaux de l'EPA pour les provinces :

1. Les poids bootstrap d'origine sont générés pour chaque ménage en appliquant l'équation (7.1), au moyen des multiplicités des 1 000 échantillons bootstrap de l'EPA et des sous-poids des ménages du fichier de totalisation final de l'EPA du mois courant.
2. Un ensemble distinct de totaux de contrôle composites est requis pour chaque répétition bootstrap. Les 1 000 ensembles de totaux sont calculés à partir des poids bootstrap finaux du mois précédent. Pour ce faire, chaque ensemble de poids répétés du fichier de poids bootstrap du mois précédent est calé en fonction des totaux de contrôle démographique du mois courant. Ensuite, pour chaque ensemble de poids, les estimations au niveau provincial pour les 28 caractéristiques de la population active indiquées à l'annexe G sont calculées.
3. Les variables auxiliaires composites qui correspondent aux totaux de contrôle composites sont obtenues pour les ménages du mois courant, comme décrit à la section 6.3.1. Les caractéristiques sont déterminées au moyen du fichier final de totalisation du mois précédent pour les ménages qui se trouvent dans le mois courant et dans le mois précédent. Les variables auxiliaires des ménages manquants dans le fichier final de totalisation du mois précédent sont imputées au moyen de l'imputation par donneur dans le cas de non-répondants, et au moyen de l'imputation par la moyenne dans le cas des ménages du renouvellement naissants. L'imputation par donneur pour les non-répondants est effectuée une seule fois, tandis que l'imputation par la moyenne est effectuée séparément pour chacune des 1 000 répétitions bootstrap.
4. Les poids bootstrap initiaux générés à l'étape 1 sont calés en fonction des totaux de contrôle démographique du mois courant et des totaux de contrôle composites à l'étape 2, à partir des variables auxiliaires composites obtenues à l'étape 3. Le calage sur marges est répété pour chaque répétition bootstrap.

Il convient de souligner que si des poids négatifs sont obtenus à l'étape 4, le calage sur marges est appliqué une deuxième fois aux poids calés, les poids négatifs étant remplacés par leur valeur dans le fichier des poids bootstrap initiaux. Si, après cette deuxième série de calage sur marges, des poids négatifs subsistent, ces poids négatifs sont établis à 1, et l'on accepte que les totaux de contrôle ne seront pas satisfaits.

Des poids bootstrap mensuels de l'EPA sont générés depuis 1998 pour les 10 provinces. Ils sont maintenant générés tous les mois, dans le cadre de la production mensuelle.

7.4 Estimation de la variance

Les poids bootstrap de l'EPA sont utilisés pour calculer les estimations de la variance à l'aide de l'équation (7.2). Les estimations de la variance peuvent être produites au moyen de progiciels, comme SAS (PROC SURVEYMEANS), Stata 9 ou une version plus récente, SUDAAN et WesVar. Gagné, Roberts et Keown (2014) et Phillips (2004) donnent des conseils sur la façon d'utiliser les poids bootstrap avec ces progiciels.

Afin de réduire la taille des fichiers de poids bootstrap de l'EPA, les fichiers contiennent un enregistrement par ménage. À l'instar des poids de l'enquête, les poids bootstrap sont les mêmes pour tous les membres du ménage, ce qui fait qu'on peut générer un fichier des poids bootstrap au niveau de la personne en attribuant des poids bootstrap au niveau du ménage à chaque membre du ménage.

Comme décrit à la section 7.1, on obtient l'estimation de la variance bootstrap pour une estimation, $\hat{\theta}$, en calculant d'abord l'estimation avec chaque ensemble de poids bootstrap pour obtenir $\hat{\theta}^{*(1)}, \dots, \hat{\theta}^{*(1,000)}$, puis en appliquant (7.2).

Pour les estimations mettant en cause plusieurs mois d'enquête, chacune des valeurs $\hat{\theta}^{*(1)}, \dots, \hat{\theta}^{*(1,000)}$ devrait être calculée sur plusieurs mois d'enquête également. Par exemple, envisagez une estimation du changement sous la forme : $\hat{\theta}_C = \hat{\theta}_2 - \hat{\theta}_1$, où $\hat{\theta}_1$ est une estimation de θ_1 , le paramètre de population pour le premier mois; et $\hat{\theta}_2$ est une estimation de θ_2 , le paramètre de population pour le deuxième mois. On obtient l'estimation de la variance bootstrap de $\hat{\theta}_C$ en calculant d'abord $\hat{\theta}_C^{*(b)} = \hat{\theta}_2^{*(b)} - \hat{\theta}_1^{*(b)}$ pour $b=1, \dots, 1000$, où $\hat{\theta}_1^{*(b)}$ est une estimation de θ_1 basée sur le b e ensemble des poids bootstrap du premier mois, et $\hat{\theta}_2^{*(b)}$ est une estimation de θ_2 basée sur le b e ensemble des poids bootstrap du deuxième mois. L'équation (7.2) est alors appliquée à $\hat{\theta}_C^{*(1)}, \dots, \hat{\theta}_C^{*(1,000)}$. Étant donné que les poids bootstrap sont fondés sur des échantillons bootstrap coordonnés, cette méthode de traitement des estimations mettant en cause plusieurs périodes tiendra compte du chevauchement et de la dépendance qui existent d'un mois à un autre. En pratique, les progiciels ne sont pas habituellement conçus pour composer avec plusieurs ensembles de données de différentes périodes. Une solution à ce problème consiste à créer un fichier d'entrée renfermant les données et les poids bootstrap de tous les mois à l'étude dans le même fichier. Il pourrait être nécessaire de créer des variables nominales pour cerner les différents mois.

Chapitre 8 Qualité des données

8.0 Introduction

L'évaluation de la qualité des données désigne le processus d'évaluation du produit fini de l'enquête par rapport aux objectifs initiaux de l'enquête. Plus précisément, les évaluations se font en termes d'exactitude et de fiabilité des données. Ce genre d'information permet aux utilisateurs de procéder à une interprétation et à une utilisation mieux informées des résultats de l'enquête. Les utilisateurs doivent avoir accès à de l'information leur permettant d'évaluer dans quelle mesure les limites des données restreignent l'utilisation des données. Les évaluations de la qualité des données sont aussi à l'avantage de l'organisme statistique. Dans la mesure où les limites des données peuvent être détectées à des étapes particulières du processus de l'enquête, ces évaluations peuvent servir à améliorer la qualité des occurrences ultérieures si l'enquête est réitérée ou s'il y a d'autres enquêtes semblables.

L'*exactitude* de l'information statistique est la mesure dans laquelle l'information décrit correctement le phénomène qu'elle devait évaluer. Généralement, elle est caractérisée par l'erreur dans les estimations statistiques et est décomposée en composantes de biais (erreur systématique) et de variance (erreur aléatoire). L'exactitude peut également être décrite en fonction des sources d'erreur majeures qui peuvent mener à l'inexactitude (p. ex., erreurs d'échantillonnage et erreurs non dues à l'échantillonnage). C'est l'approche qui sera utilisée ici.

Dans une enquête par sondage, des inférences sont faites au sujet de la population visée à partir des données recueillies auprès d'une partie seulement de cette population. Les résultats divergeront probablement de ceux pouvant être obtenus d'un recensement complet de cette population dans les mêmes conditions. L'erreur causée par l'étendue des conclusions fondées sur un échantillon seulement à toute la population est appelée erreur d'échantillonnage. Au nombre des facteurs qui contribuent aux erreurs d'échantillonnage, on trouve : la taille de l'échantillon, la variabilité des caractéristiques étudiées, le plan de sondage et la méthode d'estimation.

L'erreur non due à l'échantillonnage, comme son nom l'indique, n'a rien à voir avec le processus d'échantillonnage et se produit dans un recensement aussi bien que dans une enquête par sondage. Ce type d'erreur peut survenir à n'importe quelle étape d'une enquête (planification, conception, collecte des données, codage, saisie, vérification, estimation, analyse et diffusion des données) et est principalement attribuable à des erreurs humaines. Les intervieweurs peuvent avoir mal compris les instructions, les répondants peuvent se tromper en répondant aux questions, les réponses peuvent être mal saisies sur le questionnaire et des erreurs peuvent survenir lors du traitement et de la totalisation des données. Ce sont tous des exemples d'erreurs non dues à l'échantillonnage. L'erreur non due à l'échantillonnage est également associée à d'autres types d'erreurs, comme des erreurs dans les sources d'information, les méthodes utilisées pour obtenir des projections démographiques, des erreurs de désaisonnalisation, etc.

Pour surveiller et assurer la qualité de ses données, l'EPA s'est dotée d'un programme pour mesurer la qualité des données. Une gamme d'indicateurs de qualité sont produits sur une base régulière et analysés avec soin. En présence de valeurs inhabituelles, les responsables de l'EPA sont immédiatement avisés afin d'apporter les corrections nécessaires le plus rapidement possible. Certains indicateurs sont tout simplement surveillés, puisque leur rôle est de déceler des tendances ou des effets à long terme. À titre d'exemple, certains indicateurs permettent de mesurer les conséquences de certains changements d'ordre opérationnel alors que d'autres mesurent l'effet de modifications mineures au plan de sondage. Ces renseignements à long terme au sujet de la fiabilité des données peuvent servir à apporter des changements susceptibles d'améliorer la qualité générale des résultats et d'aider les analystes et les utilisateurs de données, tant à l'interne qu'à l'externe, dans leur travail.

Les indicateurs de la qualité produits par l'EPA sont décrits ci-dessous. La section 8.1 présente des indicateurs liés aux erreurs d'échantillonnage. Les indicateurs liés aux erreurs non dues à l'échantillonnage sont décrits à la section 8.2. La section 8.3 décrit les comités chargés de surveiller divers aspects de l'EPA afin d'assurer la qualité des données diffusées. La section 8.4 informe les utilisateurs d'autres ressources disponibles au sujet de la qualité des données de l'EPA.

8.1 Indicateurs de la qualité qui se rattachent aux erreurs d'échantillonnage

L'erreur d'échantillonnage a été définie auparavant comme une erreur découlant de l'estimation d'un paramètre de population selon laquelle une partie de la population est mesurée, au lieu de la population au complet. L'effet des

erreurs d'échantillonnage sur les estimations de l'enquête dépend de plusieurs facteurs, y compris la taille de l'échantillon, le plan de sondage, la méthode d'estimation et la variabilité de la caractéristique d'intérêt.

Tous les autres facteurs étant constants, l'erreur d'échantillonnage devrait diminuer à mesure que la taille de l'échantillon augmente. Ce phénomène traduit le fait que l'erreur d'échantillonnage devrait tomber à zéro une fois la population complète échantillonnée. Pour une taille donnée de l'échantillon, l'erreur d'échantillonnage est couplée à l'efficacité relative de diverses caractéristiques du plan. La stratification, la répartition et la méthode de sélection à chaque étape ont toutes une certaine incidence sur l'ampleur de l'erreur d'échantillonnage. La méthode d'estimation utilisée joue également un rôle important pour un plan de sondage donné. Par exemple, la méthode d'estimation composite utilisée par l'EPA réduit considérablement les erreurs d'échantillonnage (voir le chapitre 6).

Enfin, l'erreur d'échantillonnage varie d'une variable à une autre puisque le degré de variabilité diffère également d'une variable à une autre. Ces erreurs sont généralement plus grandes pour les caractéristiques qui sont relativement rares, ainsi que lorsque la caractéristique d'intérêt n'est pas répartie uniformément dans la population. Ainsi, bien qu'elles se fondent sur le même échantillon, les estimations relatives au chômage comportent généralement une erreur d'échantillonnage plus élevée que les données relatives à l'emploi.

Pour les enquêtes-échantillons probabilistes, comme l'EPA, il existe des méthodes pour calculer les erreurs d'échantillonnage. La mesure appliquée le plus souvent pour quantifier l'erreur d'échantillonnage est la variance d'échantillonnage. Les méthodes employées pour l'estimation de la variance dans le cas de l'EPA ont été présentées au chapitre 7.

Trois mesures clés sont obtenues de la variance d'échantillonnage : l'erreur type (ET), le coefficient de variation (CV) et l'effet de plan.

8.1.1 Erreur type

L'erreur type, définie comme la racine carrée de la variance d'échantillonnage, peut être utilisée pour calculer l'intervalle de confiance associé à une estimation. L'intervalle de confiance repose sur l'estimation obtenue et sa largeur dépend de l'erreur type et d'un paramètre lié au niveau de confiance.

À titre illustratif, examinons l'exemple suivant. En mai 2015, l'estimation de l'EPA pour le taux de chômage de la population canadienne de 15 ans et plus se situait à 6,8 %, et l'erreur type associée à cette estimation se chiffrait à 0,001395. Un intervalle de confiance approximatif de 68 % pour le taux de chômage réel est alors donné par $0,068 \pm 1 \times (0,001395)$, ou de 6,66 % à 6,94 %. Le niveau de confiance signifie que si le même processus de sélection et d'estimation était répété à plusieurs reprises (menant à différents échantillons et à différentes estimations), 68 % des intervalles de confiance conçus de cette manière contiendraient la valeur vraie de la population.

Les estimations du changement d'un mois à l'autre ont pris une grande importance pour les utilisateurs au fil du temps. À cet égard, la diffusion mensuelle de l'EPA indique maintenant les erreurs types (ET) pour les changements provinciaux et nationaux d'un mois à l'autre pour les personnes ayant un emploi et les chômeurs.

Compte tenu de leur stabilité, les ET incluses dans la publication mensuelle de l'EPA ne sont pas mises à jour chaque mois. Au lieu de cela, une estimation de l'ET correspondant à la moyenne des ET des 12 mois précédents est fournie. Ces estimations sont mises à jour deux fois l'an (habituellement en janvier et en juillet). Le tableau ci-dessous indique les ET observées pour la variation d'un mois à l'autre des estimations de l'emploi et du chômage, pour la population canadienne de 15 ans et plus.

Tableau 8.1
Erreur type (ET) de la variation d'un mois à l'autre, personnes ayant un emploi et chômeurs

Province	Travailleurs	Chômeurs
	millier	
Terre-Neuve-et-Labrador	2,1	2,1
Île-du-Prince-Édouard	0,6	0,6
Nouvelle-Écosse	2,7	2,5
Nouveau-Brunswick	2,3	2,1
Québec	15,9	13,7
Ontario	19,3	17,0
Manitoba	2,6	2,1
Saskatchewan	2,7	2,1
Alberta	9,8	8,1
Colombie-Britannique	10,6	8,5
Canada	29,5	25,3

8.1.2 Coefficient de variation

Le coefficient de variation (CV), qui est défini comme l'erreur type divisée par l'estimation, est une mesure relative de la variation et est habituellement exprimée en pourcentage. Dans l'exemple utilisé précédemment, le CV pour le taux de chômage de mai 2015 est de 2,05 % $((0,001395/0,068) \times 100 \%)$. Cela donne une idée de l'incertitude associée aux estimations. De petits CV sont souhaitables car ils indiquent que la variabilité due à l'échantillonnage est petite par rapport à l'estimation.

Afin d'obtenir les CV, les utilisateurs ont accès à des tableaux des CV approximatifs. Ces tableaux indiquent les CV approximatifs en fonction des valeurs observées des estimations, pour divers domaines. Les valeurs sont prudentes puisque, si plusieurs estimations étaient produites pour le même domaine, environ 75 % des CV approximatifs obtenus à partir des tableaux seront supérieurs aux CV réels qui seraient calculés si les méthodes précises étaient utilisées. Cependant, 25 % des CV approximatifs seront un peu plus faibles que le calcul précis. L'effet net se traduit par la production d'indicateurs de qualité qui montrent une qualité des estimations de l'enquête inférieure à la réalité – les intervalles de confiance sont plus larges et les tests statistiques révèlent moins de différences significatives. Ces CV approximatifs sont mis à jour chaque année et fournis dans le Guide de l'Enquête sur la population active (71-543-G).

8.1.3 Effet de plan

Une troisième mesure dérivée de la variance d'échantillonnage est l'effet de plan, une mesure relative que l'on calcule en divisant la variance d'échantillonnage d'une estimation obtenue au moyen du plan d'enquête par la variance d'échantillonnage d'un échantillon aléatoire simple (EAS) de la même taille. Cette mesure peut également être utilisée pour comparer l'efficacité d'un plan de sondage par rapport à un autre. Dans le cas de l'EPA, elle est particulièrement utile en tant qu'indicateur de la détérioration du plan de sondage au fil du temps, ou en tant que comparaison indiquant le gain ou la perte d'efficacité découlant du remaniement de l'enquête ou de la modification de certaines composantes du plan.

Différents types d'effets de plan de sondage peuvent être calculés, chacun dépendant des données utilisées pour les établir. Dans les lignes qui suivent, le terme effet du plan non ajusté sera employé pour désigner les effets de plan de sondage basés sur les poids non calés, c'est-à-dire sans l'ajustement qui tient compte des chiffres de population et des totaux estimatifs. Le terme effet de plan ajusté sera employé pour désigner les effets de plan de sondage qui sont basés sur les poids finaux, après calage composite. Par conséquent, les effets du plan de sondage non ajustés indiquent l'efficacité du plan de sondage, tandis que les effets du plan de sondage ajustés apportent une évaluation plus générale de la stratégie globale adoptée en combinant toutes les caractéristiques du plan d'enquête (stratification, échantillonnage à plusieurs degrés, poststratification et estimation). Plus l'effet de plan est faible, plus le plan est efficace en ce qui concerne la variance de l'échantillonnage. Il convient de souligner que les effets du plan non ajustés (plan de sondage) sont généralement plus importants que les effets du plan ajustés (plan d'enquête) fondés sur les poids finaux, puisqu'ils ne profitent pas du gain de précision qu'apporte le calage sur marges.

Le tableau suivant présente quelques valeurs représentatives des effets du plan ajustés et non ajustés pour les caractéristiques emploi et chômage à l'échelle nationale et provinciale, selon les données d'enquête de janvier à août 2015.

Tableau 8.2
Effets du plan, employés et chômeurs, 2015

Province	Travailleurs		Chômeurs	
	Ajustés	Non ajustés	Ajustés	Non ajustés
Terre-Neuve-et-Labrador	0,40	1,78	1,08	1,00
Île-du-Prince-Édouard	0,31	1,28	1,00	1,03
Nouvelle-Écosse	0,35	1,85	1,08	1,17
Nouveau-Brunswick	0,36	2,20	1,17	1,17
Québec	0,50	2,70	1,66	1,96
Ontario	0,42	2,92	1,39	1,64
Manitoba	0,32	3,03	1,01	1,19
Saskatchewan	0,34	4,87	1,10	1,11
Alberta	0,48	4,25	1,44	1,66
Colombie-Britannique	0,44	3,52	1,44	1,59
Canada	0,54	3,73	1,77	2,08

Dans l'EPA, les effets non ajustés du plan, ainsi que d'autres renseignements, sont utilisés pour déterminer les régions où le plan de sondage a perdu une importante partie de son efficacité au fil du temps. Dans certains cas, un léger remaniement est effectué dans ces régions pour remédier à ce problème.

8.2 Indicateurs de la qualité qui se rattachent aux erreurs non dues à l'échantillonnage

Les erreurs non dues à l'échantillonnage sont des erreurs qui surviennent pendant à peu près toutes les activités d'enquête, mis à part l'échantillonnage. L'effet sur les estimations peut se manifester sous forme de biais et/ou de variabilité accrue dans les estimations. Si ces erreurs sont des erreurs aléatoires, leurs effets seront à peu près annulés dans un secteur suffisamment grand, ce qui n'entraînera rien de plus qu'une variabilité accrue. Par contre, l'effet peut demeurer important lorsqu'il s'agit de petits secteurs ou lorsque les caractéristiques à l'étude sont rares. Si les erreurs sont systématiques, en ce sens qu'elles ont tendance à aller dans la même direction, il s'ensuivra des résultats finals biaisés. De plus, contrairement aux erreurs aléatoires, le biais associé aux erreurs systématiques ne peut être réduit en augmentant la taille de l'échantillon.

Les sources d'erreurs non dues à l'échantillonnage les plus fréquentes sont l'erreur de couverture, la non-réponse, les erreurs de mesure ou de réponse et les erreurs de traitement. Nous y reviendrons individuellement dans les sections qui suivent.

8.2.1 Erreurs de couverture

Les erreurs de couverture sont des omissions, des inclusions erronées, des répétitions et des erreurs de classification d'unités dans la base de sondage. Dans le cas de l'EPA, ces erreurs peuvent se produire lorsque la liste des logements associée à une UPE est établie ou chargée, lorsque la liste est mise à jour pour cerner la croissance, lorsque les logements et/ou les personnes à inclure dans l'enquête sont contactés, ou lorsque des données sont recueillies et traitées. Dans l'EPA, trois grands indicateurs sont utilisés pour mesurer et surveiller les erreurs de couverture : le taux de glissement, le taux d'inoccupation et l'évaluation du rendement de l'UPE.

Le taux de glissement est la différence relative entre les estimations de la taille de la population établies à partir des poids avant le calage et les dernières estimations démographiques utilisées comme totaux pour le calage.

Les estimations démographiques utilisées dans la détermination du taux de glissement peuvent également comporter des erreurs, et ces erreurs sont en fait un des facteurs qui contribuent au glissement. Dans l'EPA, le sous-dénombrement est habituellement observé, comme l'indique un taux de glissement positif. Pour réduire le biais obtenu le plus possible, le poids de chaque répondant est modifié par le facteur de correction par calage composite (voir le chapitre 6).

Le sous-dénombrement découle de l'omission de logements ou de personnes dans la population cible. Il se peut qu'un logement occupé ne soit pas inscrit dans la liste des UPE pour diverses raisons : omission lors de l'établissement de la liste, immeuble en construction durant la dernière vérification, erreurs dans les délimitations

de la grappe ou encore logement classifié vacant par erreur. Il est également possible que des personnes soient oubliées dans un ménage, soit parce que le répondant ne révèle pas sa présence ou encore qu'on lui a attribué un lieu de résidence habituel ailleurs que dans le ménage échantillonné. Les étudiants sont souvent oubliés puisqu'ils résident ailleurs durant leurs études, quoique leur résidence habituelle soit dans l'échantillon. Des erreurs peuvent donc se glisser dans les estimations de l'enquête, si les caractéristiques des personnes non incluses dans l'enquête diffèrent de celles des personnes incluses. Par exemple, si l'enquête n'inclut pas une partie de la population qui est jeune et grandement mobile, qui affiche un taux de chômage plus élevé que celui de la population du même âge dans l'enquête, le glissement biaise les estimations du chômage à la baisse.

Le glissement est également touché par les corrections pour tenir compte de la croissance démographique et de la non-réponse. La population s'accroît entre les remaniements, généralement à des endroits spécifiques et non pas de manière uniforme. L'échantillon sélectionné peut surestimer ou sous-estimer cette croissance ou en rendre compte de façon précise. Par exemple, les UPE sélectionnées dans une région peuvent n'afficher aucune croissance, mais d'autres UPE dans la base de sondage de la même région pourraient enregistrer une croissance significative. En pareil cas, la croissance serait sous-estimée par l'échantillon sélectionné, et si les estimations projetées de population cadrent avec la croissance réelle, les taux de glissement augmenteraient pour cette région.

Les corrections pour tenir compte de la non-réponse (voir les chapitres 5 et 6) peuvent également influencer le glissement. Par exemple, si les ménages non répondants ont moins de membres mais qu'ils sont représentés dans l'échantillon, au moyen de l'imputation ou de facteurs de correction de la non-réponse, par les gros ménages, le taux de glissement peut être touché.

Finalement, comme mentionné précédemment, les estimations de population ont également un rôle à jouer en ce qui concerne le glissement. Plus elles sont précises, plus les taux de glissement sont utiles.

Tous les mois, les taux de glissement sont analysés rigoureusement. Ils sont produits tous les mois à l'échelle nationale (sauf les territoires) et provinciale et pour 12 groupes d'âge-sexe (15 à 19 ans, 20 à 24 ans, 25 à 29 ans, 30 à 39 ans, 40 à 54 ans et 55 ans et plus). Le tableau qui suit présente les taux de glissement moyens pour l'année civile 2015.

Tableau 8.3
Taux de glissement moyens – Canada par groupe d'âge et province, 2015

Canada	%
Tous les âges	11,7
15 à 19 ans	8,2
20 à 24 ans	21,3
25 à 29 ans	21,3
30 à 39 ans	16,3
40 à 54 ans	9,8
55 ans et plus	7,0
Terre-Neuve-et-Labrador	11,6
Île-du-Prince-Édouard	16,2
Nouvelle-Écosse	12,3
Nouveau-Brunswick	11,7
Québec	8,6
Ontario	12,0
Manitoba	9,5
Saskatchewan	13,7
Alberta	15,1
Colombie-Britannique	12,8

Les logements identifiés correctement comme étant vacants ou invalides n'introduisent aucun biais dans les estimations de l'EPA. Toutefois, la variance de l'estimation s'en trouve plus élevée puisque l'échantillon compte un nombre moins élevé de ménages valides. Les intervieweurs de l'EPA retournent visiter les logements vacants sélectionnés tous les mois afin d'interviewer les personnes ciblées par l'enquête qui peuvent avoir emménagé depuis le mois précédent. Les logements inexistantes sont tout simplement retirés de la base de sondage. Une attention particulière doit être accordée à la détermination des logements qui sont vacants, qui influencent directement deux autres indicateurs. En effet, si un logement est codé vacant alors que les occupants sont temporairement absents, le taux de non-réponse produit pour l'EPA sera sous-estimé. Par ailleurs, le taux de glissement s'en trouve surestimé puisque ce logement mal codé aurait dû être considéré lors de la détermination

de ce taux. Les intervieweurs se doivent donc de faire un travail très minutieux pour déterminer si un logement est vacant, et par conséquent, hors du champ de l'enquête, ou tout simplement occupé par un ménage temporairement absent et donc dans le champ de l'enquête. Les taux d'inoccupation sont également produits et surveillés tous les mois.

Le tableau suivant présente les taux d'inoccupation moyens et les valeurs minimum et maximum pour 2015 à l'échelle provinciale et nationale.

Tableau 8.4
Taux d'inoccupation (non pondéré), Canada et les provinces, 2015

Province	Moyenne	Maximum	Minimum
	%		
Terre-Neuve-et-Labrador	15,4	16,4	14,5
Île-du-Prince-Édouard	21,1	23,4	19,8
Nouvelle-Écosse	18,1	18,9	17,4
Nouveau-Brunswick	17,1	18,1	16,2
Québec	12,8	13,3	12,2
Ontario	11,5	11,8	11,2
Manitoba	14,1	15,5	12,0
Saskatchewan	14,3	15,8	12,6
Alberta	14,6	15,2	13,7
Colombie-Britannique	12,3	13,2	11,9
Canada	13,7	14,1	13,0

Pour cet indicateur de la qualité, une certaine variabilité est observée entre les provinces. Ce phénomène est lié à la proportion de logements saisonniers possédés variant d'une province à une autre. Les logements saisonniers sont toujours considérés comme vacants, parce qu'ils ne sont pas le lieu de résidence habituel des occupants.

Le rendement des UPE est surveillé tous les mois pour détecter les différences importantes entre le nombre de logements sondés sur le terrain et le nombre de logements anticipés par le plan de sondage. Par conséquent, tout écart significatif, comme 50 % (positif ou négatif), entre un extrait du FUL et les résultats de l'enquête sur le terrain, est examiné. D'abord, toutes les grappes ayant un rendement inattendu sont portées à l'attention de l'unité responsable du contrôle de l'échantillon à Ottawa, qui vérifie les frontières de la grappe et le nombre de logements attendu. Si l'écart ne peut être expliqué au bureau central, la grappe est acheminée au bureau régional concerné pour être analysée en détail. Toutes les causes expliquant les écarts sont répertoriées pour consultation future.

Ce contrôle joue un rôle important puisque si la taille de l'échantillon nécessite des changements, il est essentiel de connaître quelles régions sont sous-échantillonnées ou suréchantillonnées. De plus, les écarts enregistrés peuvent révéler des problèmes pour l'enquête qui pourraient entacher la qualité des données de l'EPA.

Tous ces indicateurs (taux de glissement, taux d'inoccupation et rendement de l'UPE) servent à détecter les problèmes potentiels en ce qui concerne la couverture de l'échantillon et à participer à la prise de mesures appropriées. À titre d'exemples de mesures possibles, mentionnons la création d'outils de formation pour les intervieweurs afin d'enrichir leurs connaissances sur les règles de composition du ménage, la distribution d'un bulletin expliquant le glissement ou le concept des logements multiples, ou l'établissement d'un programme pour tenir à jour la liste d'un certain nombre d'UPE réputées en croissance.

8.2.2 Non-réponse

Chaque mois, durant la semaine d'enquête, les intervieweurs s'affairent à déterminer quels sont les logements sélectionnés qui contiennent des personnes admissibles à l'enquête. Les logements réputés inadmissibles pour le mois d'enquête le sont pour les raisons suivantes :

- logements hors du champ de l'enquête, c'est-à-dire les logements occupés par aucun membre de la population cible, p. ex., seulement des membres des Forces armées canadiennes;
- logements vacants, c'est-à-dire les logements qui sont inoccupés, saisonniers ou en construction;
- logements invalides, c'est-à-dire les logements qui ont été démolis, convertis en locaux commerciaux, déménagés (p. ex., maison mobile), abandonnés ou entrés par erreur au départ.

Lorsqu'un logement est réputé admissible à l'enquête, il n'est pas toujours possible de réaliser une interview. On parle alors de non-réponse des ménages, un phénomène qui peut être attribuable à diverses raisons, notamment les suivantes : personne à la maison, absence temporaire, interview impossible (mauvaises conditions météorologiques, circonstances inhabituelles dans le ménage, etc.), problèmes techniques ou refus.

L'ampleur du biais attribuable à la non-réponse est habituellement inconnue, mais elle est directement liée aux caractéristiques divergentes entre les groupes d'unités répondantes et les groupes d'unités non répondantes. Étant donné que l'effet de ce biais s'accroît à mesure que le taux de non-réponse augmente, on s'efforce de maintenir le taux de réponse le plus haut possible pendant la collecte.

Le tableau suivant présente les taux de non-réponse moyens, ainsi que les taux minimum et maximum pour l'année 2015.

Tableau 8.5
Taux de non-réponse (non pondérés), Canada et les provinces, 2015

Province	Moyenne	Maximum	Minimum
	%		
Terre-Neuve-et-Labrador	11,2	13,0	9,9
Île-du-Prince-Édouard	10,9	12,2	9,0
Nouvelle-Écosse	10,3	11,3	9,7
Nouveau-Brunswick	11,4	12,5	10,5
Québec	10,2	11,9	8,2
Ontario	13,9	15,4	12,6
Manitoba	11,7	12,8	10,3
Saskatchewan	11,9	12,8	11,1
Alberta	12,8	14,0	11,5
Colombie-Britannique	11,7	12,6	10,6
Canada	12,0	13,1	11,2

Tous les mois, l'EPA produit les taux de non-réponse en fonction de la cause (simple refus, aucun contact, absence temporaire, problème technique, ou autre raison) ainsi que par mode de collecte. Ces taux sont analysés avec soin afin de déceler les causes majeures de la non-réponse et d'apporter les correctifs requis.

Les taux de refus pour l'EPA sont habituellement très bas, les taux canadiens mensuels oscillant de 1 % à 2 %. Les taux de refus sont ordinairement semblables d'une province à une autre, mais ils peuvent descendre aussi bas que 0,5 % ou monter aussi haut que 3 %. Dans une certaine mesure, le système de collecte rend possible la collecte de renseignements supplémentaires sur la raison du refus, ce qui permet de suivre l'évolution de l'attitude des répondants à l'égard de l'enquête au fil du temps.

8.2.3 Erreurs de mesure ou de réponse

Les erreurs de mesure ou de réponse peuvent être attribuables à la conception du questionnaire, à la formulation des questions, à la compréhension du répondant, à la façon dont l'interview est menée ou aux conditions générales dans lesquelles l'enquête est réalisée. Elles peuvent se produire au moment où les renseignements sont fournis, reçus ou entrés dans l'ordinateur. Toutefois, le mode de collecte informatisé permet de réduire certaines de ces erreurs, puisque certaines règles de vérification sont intégrées à l'instrument de collecte et que les conflits doivent être résolus au moment même de l'interview. Il se peut cependant que le répondant interprète mal la question, qu'il ne sache pas la réponse, qu'il ait oublié ou qu'il préfère déformer les faits pour des raisons qui lui sont personnelles. De plus, les intervieweurs peuvent réinterpréter involontairement les réponses. Comme pour les autres catégories d'erreurs, les erreurs de réponse peuvent donner lieu à une augmentation de la variance et/ou à la présence d'un biais.

Les réponses par personne interposée fournies par un membre du ménage lorsque l'information est recueillie au sujet d'un autre membre du ménage peuvent également entraîner des erreurs de réponse. Cependant, ces erreurs sont considérées comme préférables aux erreurs de non-réponse qu'il faudrait régler si les réponses étaient acceptées uniquement par le répondant pour lui-même. À l'heure actuelle, environ 60 % des renseignements de l'EPA sont fournis par une personne interposée, et ce taux demeure relativement stable au fil du temps.

Dans les enquêtes répétées, où l'échantillon est constitué d'un certain nombre de panels ou de groupes de renouvellement, l'espérance mathématique des estimations varie légèrement d'un groupe de renouvellement à un autre. Il se produit alors ce qu'on appelle un biais de renouvellement. En ce qui concerne l'EPA, ce biais atteint son plus haut niveau pour le sixième de l'échantillon qui en est à sa première interview. Il est possible de calculer l'effet du renouvellement en calculant le ratio entre une estimation calculée pour la partie de l'échantillon participant à l'enquête un certain nombre de fois (premier mois, deuxième mois, etc.) et l'estimation calculée pour l'échantillon entier.

Brisebois et Mantel (1996) ont calculé un indice de renouvellement modifié qui tient compte des différences des effets des erreurs dues à l'échantillonnage pour les six groupes de renouvellement. Leur étude a révélé plusieurs écarts significatifs entre les groupes de renouvellement, mais l'effet général a été réputé mineur.

8.2.4 Erreurs de traitement

Des erreurs de traitement peuvent se produire à diverses étapes de l'enquête, comme à la saisie, à la validation, à la vérification, au codage, à l'imputation, à la pondération et à la totalisation des données.

La méthode de collecte informatisée aide à prévenir les erreurs de cheminement pendant la saisie des données, puisque l'application détermine l'ordre des questions. De même, certaines règles de vérification sont intégrées au système de collecte, ce qui permet de détecter et de corriger certaines divergences au moment de l'interview.

Les variables « profession » et « secteur d'activité » sont codées en fonction des normes de classification au bureau central. Au premier mois des interviews, l'intervieweur recueille de l'information décrivant avec précision le genre d'entreprise, de secteur d'activité ou de service où la personne travaille, ainsi que de l'information indiquant clairement et précisément le genre de travail ou la nature des fonctions. Le premier type d'information servira à déterminer le secteur d'activité, tandis que le second permettra d'identifier la profession. Une des premières étapes du traitement au bureau central consiste à coder l'information descriptive recueillie pour les variables « profession » et « secteur d'activité » selon la classification type pour ces variables, la CNP et le SCIAN. Des processus de contrôle de la qualité mensuels sont en place pour évaluer la précision de ce processus de codage.

Le taux d'imputation est également un indicateur de qualité relatif au traitement des données. Chaque mois, des diagnostics évaluant les résultats du processus d'imputation sont produits et examinés rigoureusement. Les diagnostics renseignent sur le nombre d'enregistrements traités par chaque méthode d'imputation et à chaque niveau de regroupement (voir le chapitre 5). Les profils respectifs des enregistrements non imputés et des enregistrements imputés sont comparés, ainsi que leur contribution respective aux estimations clés de l'enquête. On peut ainsi contrôler la qualité de l'imputation et prendre les mesures qui s'imposent.

Pour éviter les erreurs susceptibles de se produire aux étapes de l'estimation et de la totalisation, un outil d'évaluation pré-diffusion a été conçu. Grâce à cet outil, il est possible de faire ressortir des variables, des sous-groupes et/ou des domaines pour lesquels les estimations et/ou les erreurs types sont anormalement éloignées de leurs moyennes historiques respectives. Ces estimations peuvent être examinées de plus près pour déterminer si une erreur quelconque est à l'origine du changement soudain. En outre, des comparaisons avec d'autres sources de données sont effectuées régulièrement, afin de vérifier si les données de l'EPA sont conformes à d'autres réalités économiques.

8.2.5 Observation des procédures de collecte

L'application de collecte produit des fichiers de paradosés contenant une foule de renseignements sur les activités des intervieweurs sur le terrain et dans les centres d'appels. À partir de ces fichiers, il est possible de produire des indicateurs de qualité des activités des intervieweurs. L'EPA analyse régulièrement les appels et les visites effectués par les intervieweurs. Parmi les rapports produits, mentionnons entre autres de l'information sur la durée des interviews (en personne et par téléphone), le nombre de tentatives en vue de joindre un répondant et le nombre de cas transférés d'un mode de collecte à un autre. Au moyen de cette source d'information, il est relativement facile de vérifier si les intervieweurs respectent rigoureusement les procédures de collecte et de prendre les mesures qui s'imposent pour les cas douteux. Ces indicateurs peuvent également être utilisés pour

améliorer le programme de formation pour les intervieweurs et renforcer certaines composantes, comme la planification des tâches ou l'horaire de travail.

8.3 Comités de l'EPA

L'EPA a besoin de plusieurs groupes de coordination pour veiller au bon déroulement de l'enquête. Deux comités permanents sont décrits ci-après. Leur mandat consiste entre autres à s'occuper des opérations permanentes et à évaluer l'enquête de façon régulière.

8.3.1 Comité des opérations

Ce comité a pour mandat de surveiller les activités qui surviennent pendant chaque mois d'enquête et les circonstances entourant la réalisation de l'enquête, de veiller au bon déroulement des opérations, d'examiner les changements proposés et d'en recommander ou non l'adoption. Le Comité des opérations est présidé par un membre principal de la Division de la statistique du travail et se réunit toutes les semaines.

8.3.2 Comité de la qualité des données

Le Comité, qui a été créé officiellement au printemps 1972, a pour mandat d'examiner, d'évaluer et de documenter la qualité des enquêtes mensuelles, ainsi que de donner des conseils sur les aspects de la qualité à examiner. Il entreprend et examine également des études et des recherches ponctuelles au sujet des méthodes et des procédures influant sur la qualité des données, et formule des recommandations en fonction de ses constatations. Ce comité est présidé par un membre de la Division des méthodes d'enquêtes auprès des ménages.

Pour assurer la meilleure qualité possible des données, le Comité sur la qualité des données examine périodiquement les différents indicateurs de qualité décrits précédemment. Il se réunit chaque mois pour étudier et évaluer la qualité des données mensuelles et pour faire des suggestions et des recommandations sur tout aspect susceptible d'améliorer la qualité. En suivant étroitement l'évolution des indicateurs de la qualité, le Comité peut intervenir immédiatement auprès des responsables des activités de l'EPA concernées afin de contrôler la qualité des données mensuelles. Le Comité discute également de faits nouveaux susceptibles d'influencer la qualité des données venant d'être recueillies ou devant être recueillies dans l'avenir, en particulier les changements relatifs aux méthodes de collecte ou au questionnaire, les problèmes inhabituels sur le terrain, la mise à l'essai continue des procédés et des méthodes, etc.

8.4 Ressources disponibles au sujet de la qualité des données de l'EPA

Il y a de multiples autres ressources contenant des informations sur différents aspects de la qualité des données de l'EPA. Cette section fera la description de quelques-unes d'entre elles.

8.4.1 Le Quotidien

L'Enquête sur la population active évalue l'état actuel du marché du travail canadien. Grâce aux données recueillies par l'EPA, il est possible de produire divers types d'estimations (estimations mensuelles, estimation de la variation d'un mois à l'autre, moyenne mobile sur trois mois, etc.) pour bien des caractéristiques différentes (situation vis-à-vis de l'activité, heures travaillées, titulaires d'emplois multiples, etc.), dans des milliers de domaines (national, provincial, infraprovincial, groupes d'âge-sexe, etc.). Statistique Canada publie les estimations de l'EPA à chaque mois, seulement 10 jours après la fin de la collecte des données. La diffusion des nouvelles estimations de l'EPA, qui survient généralement le premier vendredi du mois, est annoncée dans Le Quotidien, le bulletin de diffusion officielle des données de Statistique Canada, et est accompagnée d'une analyse sommaire du marché du travail actuel. La diffusion inclut aussi des informations sur des aspects précis de l'enquête, tels que les révisions à venir, les produits et rapports nouvellement disponibles, la date de la prochaine diffusion.

8.4.2 La page web de l'Enquête sur la population active

La page web de l'Enquête sur la population active, sur le site web de Statistique Canada, contient de l'information détaillée sur plusieurs aspects de l'enquête, incluant la qualité. En particulier, la page contient de l'information sur

le processus d'évaluation de qualité et sur les diverses sources de données auxquelles les estimations de l'EPA sont comparées pour déterminer si les tendances du marché du travail concordent avec la performance générale de l'économie. On y trouve aussi un sommaire des changements survenus aux données ou aux estimations au fil du temps.

8.4.3 Le Guide de l'Enquête sur la population active

Le Guide de l'Enquête sur la population active (71-543-G) est une source précieuse d'informations sur les concepts, les classifications et les définitions utilisés par l'enquête. Il contient aussi des lignes directrices et du soutien pour la comparaison des estimations de l'EPA avec les estimations produites par d'autres enquêtes (telles que l'Enquête sur l'emploi, la rémunération et les heures de travail (EERH)) ou par d'autres pays (telles que les États-Unis).

8.4.4 Accès aux données de l'Enquête sur la population active

Pour les utilisateurs intéressés par les estimations les plus courantes de l'EPA, l'information pourrait être facilement accessible dans les tableaux connexes de CANSIM. Divers types d'estimations sont fournies pour différents domaines et les règles de divulgation sont appliquées pour protéger la confidentialité.

Pour des situations plus précises, les utilisateurs peuvent utiliser le fichier de microdonnées à grande diffusion publié mensuellement (71M0001X). Ce produit s'adresse aux utilisateurs qui préfèrent effectuer leur propre analyse et leur permet de se concentrer sur des sous-groupes en particulier dans la population ou de recouper les variables qui ne sont pas dans les produits catalogués. Les utilisateurs peuvent alors envoyer leurs demandes en recouvrement des coûts pour obtenir les estimations de la variance associées à leurs besoins particuliers.

Un centre de données de recherche (CDR) donne accès aux fichiers de microdonnées confidentielles de Statistique Canada. Ils ne sont accessibles qu'aux chercheurs dont les projets ont été approuvés et qui ont prêté serment en tant que « personnes réputées être employées » de Statistique Canada. Les fichiers de microdonnées confidentielles des CDR contiennent la plus grande partie des renseignements recueillis sur le sujet dans le cadre de l'interview d'enquête, ainsi que des variables dérivées ajoutées par la suite à l'ensemble des données. Ils contiennent également les poids bootstrap qui servent à calculer les estimations de variance. Ces poids ne se trouvent que dans le fichier principal. On trouve des CDR à l'étendue du pays. Pour en savoir plus, veuillez visiter www.statcan.gc.ca/fra/cdr/index.

Le système d'accès à distance en temps réel (ADTR) vient compléter les méthodes existantes d'accès aux microdonnées confidentielles. À l'aide d'un nom d'utilisateur et d'un mot de passe sécurisés, l'ADTR offre en tout temps un accès aux résultats d'enquête à partir de n'importe quel ordinateur muni d'un accès à internet. La confidentialité des microdonnées est automatisée dans le système d'ADTR, rendant ainsi inutile l'intervention manuelle et permettant d'accéder rapidement aux résultats. Afin d'utiliser le programme d'accès direct en temps réel (ADTR), les demandeurs doivent remplir un formulaire de demande. Pour obtenir plus de renseignements, veuillez visiter www.statcan.gc.ca/fra/adtr/adtr.

Chapitre 9 Utilisation de la base de sondage ou de l'échantillon de l'EPA pour d'autres enquêtes

9.0 Introduction

Bien des enquêtes-ménages utilisent la base de sondage ou l'échantillon de l'Enquête sur la population active pour leur plan d'enquête. La section 9.1 décrit comment la base de sondage de l'EPA est utilisée par certaines autres enquêtes pour assurer la coordination avec l'EPA. La section 9.2 décrit comment l'échantillon de l'EPA est utilisé pour obtenir des échantillons pour des enquêtes supplémentaires ou des enquêtes après renouvellement de l'EPA. La section 9.3 contient des exemples. Les enquêtes qui utilisent la base de sondage ou l'échantillon de l'EPA sont des composantes importantes du programme des enquêtes-ménages de Statistique Canada et sont souvent parrainées par d'autres ministères.

9.1 Enquêtes qui utilisent la base de sondage de l'EPA

Certaines enquêtes utilisent la base de sondage de l'EPA pour sélectionner un échantillon distinct de ménages, habituellement dans des unités primaires d'échantillonnage (UPE) qui sont également actives dans l'EPA. Chaque enquête réserve un ensemble de départs aléatoires pour sélectionner des logements pour son usage exclusif. Selon la répartition souhaitée, chaque strate de l'EPA peut avoir zéro, un ou plusieurs départs réservés de cette manière. Dans certains cas, des UPE qui ne seront pas actives avant plusieurs années pour l'EPA peuvent également avoir des départs aléatoires réservés pour d'autres enquêtes. Les échantillons sont sélectionnés à l'aide de ces départs aléatoires. Si l'échantillon n'exige pas de départs complets, l'enquête peut procéder à sa propre stabilisation. Cette stratégie de sélection d'échantillons distincts réduit le fardeau du répondant parce qu'elle fait en sorte qu'un logement ne puisse pas être sélectionné par plus d'une enquête. On parle habituellement de coordination négative des logements sélectionnés.

Bien que des échantillons distincts soient sélectionnés pour les autres enquêtes qui utilisent la base de sondage de l'EPA, ils peuvent souvent partager les ressources des intervieweurs avec l'EPA, puisqu'ils se trouvent habituellement dans les mêmes UPE. L'échantillonnage de logements dans la même région pendant la même période de collecte entraîne des réductions des coûts de collecte, en particulier pour les enquêtes ayant une forte proportion d'IPAO (interviews sur place assistées par ordinateur). Cette stratégie de sélection de logements pour différentes enquêtes pour les mêmes UPE s'appelle la coordination positive des UPE sélectionnées.

9.2 Enquêtes qui utilisent l'échantillon de l'EPA

Deux types d'enquêtes utilisent l'échantillon de l'EPA : les enquêtes supplémentaires et les enquêtes après renouvellement de l'EPA. Les enquêtes supplémentaires visent les ménages qui ont également été sélectionnés pour l'EPA et qui sont encore actifs dans l'EPA. Les suppléments dépendants utilisent les ménages de l'EPA pendant qu'ils sont encore interviewés pour l'EPA, tandis que les suppléments indépendants se détachent de l'EPA pour interviewer les ménages de l'EPA à un autre moment, ou pour permettre plus de temps que ne le ferait l'EPA pour la collecte des données. Les enquêtes après renouvellement de l'EPA ressemblent aux enquêtes supplémentaires, mais elles contactent le ménage après son élimination de l'échantillon de l'EPA, c.-à-d. une fois que le ménage a terminé son sixième mois de participation à l'EPA.

Le principal avantage des enquêtes supplémentaires et des enquêtes après renouvellement de l'EPA est qu'elles peuvent utiliser les données recueillies dans le cadre de l'EPA pour présélectionner les répondants en fonction des besoins de l'enquête, ce qui peut représenter des économies considérables pour les enquêtes qui essaient de joindre des ménages ou des personnes ayant des caractéristiques particulières (p. ex., les chômeurs). Lorsqu'un ménage a été éliminé de l'échantillon de l'EPA, ce ménage demeure admissible aux enquêtes après renouvellement de l'EPA pendant une période pouvant aller jusqu'à deux ans.

La principale préoccupation dans le contexte des enquêtes supplémentaires et des enquêtes après renouvellement de l'EPA est le fardeau du répondant. Les sujets ou les questions qui risquent d'être inacceptables pour les répondants ou qui risquent d'avoir une incidence, dans une certaine mesure, sur les réponses obtenues à l'EPA le mois suivant sont évités. Selon le sujet et/ou le nombre d'enquêtes actives dans un mois donné, certains

suppléments sont bien accueillis; ils augmentent la durée d'interview, mais d'un autre côté, ils permettent de varier l'expérience des ménages qui font partie de l'échantillon de l'EPA pour une durée de six mois.

On peut utiliser chacun des six groupes de renouvellement de l'EPA pour produire des estimations. En général, ces enquêtes utilisent d'un à cinq groupes de renouvellement pour leur échantillon, selon le niveau de fiabilité nécessaire. Pour les suppléments, on évite habituellement le groupe de renouvellement naissant de l'EPA c'est-à-dire celui formé de ménages sondés pour la première fois dans le cadre de l'EPA, en raison du fardeau du répondant. L'interview initiale de l'EPA prend plus de temps à réaliser que les interviews subséquentes.

Dans certains cas, seule une partie des ménages d'un groupe de renouvellement est requise. Les logements sont retirés au hasard pour que le nombre de ménages soit réduit au nombre requis, comme c'est le cas dans le programme de stabilisation de l'EPA. Dans un logement sélectionné, l'enquête peut s'adresser à tous les répondants admissibles selon les critères de l'EPA ou à des personnes en particulier. Des particuliers distincts à l'intérieur des logements choisis peuvent être sélectionnés par le biais d'un processus de sélection aléatoire ou de présélection des répondants ayant des caractéristiques particulières du point de vue démographique ou du point de vue de l'activité à partir de l'EPA ou grâce à des questions spéciales.

Deux autres utilisateurs de l'échantillon actif de l'EPA sont la composante du contenu additionnel (CA) et la composante des effets de désastre ou catastrophe (EDC). La composante CA est un module qui permet l'ajout d'un petit nombre de questions (pas plus de cinq) au questionnaire de l'EPA. Des questions sur un sujet spécifique et ponctuel sont ajoutées pour un seul mois et sont posées à tous les répondants. Ceci permet à l'enquête d'être réalisée et aux résultats d'être diffusés très rapidement. La composante EDC est un module qui est utilisé pour mesurer l'impact économique d'événements tel que des désastres naturels. Dans les régions affectées par de tels événements, quatre questions sont ajoutées pour mesurer le nombre d'heures de travail perdues et le nombre d'heures supplémentaires travaillées. Ce module a été utilisé, par exemple, pour mesurer l'impact économique des inondations de 2013 dans la région de Calgary.

9.3 Exemples d'enquêtes qui utilisent la base de sondage ou l'échantillon de l'EPA

La liste suivante présente certaines des enquêtes ayant utilisé la base de sondage ou l'échantillon de l'EPA en 2015.

Tableau 9.1
Enquêtes utilisant la base de sondage ou l'échantillon de l'EPA en 2015

Enquête	Période de collecte des données - 2015
Enquêtes supplémentaires	
Enquête sur les voyages des résidents du Canada (EVRC)	Janvier à décembre (mensuelle)
Enquête canadienne sur le revenu (ECR)	Janvier à avril
Contenu additionnel d'Élections Canada	Novembre
Enquêtes après renouvellement de l'EPA	
Enquête sur la couverture de l'assurance-emploi (ECAE)	Avril-mai, juillet-août, novembre-décembre et janvier 2016 -février 2016
Enquêtes utilisant la base de sondage de l'EPA	
Enquête sur la santé dans les collectivités canadiennes (ESCC)	Janvier à décembre (4 périodes de collecte trimestrielles)
Enquête sur les dépenses des ménages (EDM)	Janvier à décembre (mensuelle)

9.3.1 Enquête sur la santé dans les collectivités canadiennes

L'ESCC est une enquête transversale qui vise à recueillir des renseignements sur l'état de santé, l'utilisation des services de santé et les déterminants de la santé de la population canadienne. Elle est réalisée auprès d'un grand échantillon de répondants et conçue pour fournir des estimations fiables à l'échelle de la région sociosanitaire. L'utilisation des données de l'ESCC comprend la surveillance et la recherche sur la santé de la population. Les ministères fédéraux et provinciaux de la santé et des ressources humaines, les organismes de services sociaux et autres agences gouvernementales utilisent l'information recueillie auprès des répondants afin de surveiller, planifier, instaurer et évaluer des programmes pour améliorer la santé des Canadiens. L'échantillon est divisé sur une base annuelle en quatre périodes de collecte de trois mois sans recoupement. Une grande partie de l'échantillon de l'ESCC est coordonnée positivement avec l'échantillon de l'EPA pour ce qui est des UPE

sélectionnées. Autrement dit, chaque mois, la collecte de l'ESCC se déroule dans de nombreuses UPE où l'EPA est effectuée des IPAO.

9.3.2 Enquête sur les dépenses des ménages

L'EDM est une enquête annuelle qui vise principalement à recueillir des renseignements détaillés sur les dépenses des ménages. L'EDM combine l'utilisation d'un questionnaire avec périodes de rappel adaptées à la dépense (1, 3 et 12 mois, dernier paiement, quatre semaines) et d'un journal de dépenses quotidiennes que les ménages sélectionnés remplissent pendant une période de deux semaines suivant l'interview. Les données de l'EDM sont utilisées à Statistique Canada par le Système de comptabilité nationale, surtout pour permettre le calcul du produit intérieur brut (PIB). Les données aident également à mettre à jour les proportions (poids) de l'Indice des prix à la consommation (IPC). Les données sont recueillies sur une base continue de janvier à décembre de l'année d'enquête, auprès d'un échantillon de ménages répartis sur 12 cycles mensuels de collecte. Depuis 2015, la majeure partie de l'échantillon de l'EDM est coordonnée positivement avec l'échantillon de l'EPA en ce qui concerne les UPE sélectionnées. Autrement dit, chaque mois, la collecte de l'EDM se déroule presque exclusivement dans les UPE où l'EPA est en train de réaliser des premières interviews.

9.3.3 Enquête sur les voyages des résidents du Canada

L'EVRC est parrainée par Statistique Canada, la Commission canadienne du tourisme et les gouvernements provinciaux. Elle mesure l'importance des voyages intérieurs au Canada, du point de vue de la demande. Cette enquête a pour objectif de fournir de l'information sur le nombre de voyages et les dépenses de voyage des résidents canadiens, de l'information sur l'incidence des voyages et le profil sociodémographique des voyageurs et des non-voyageurs. Il s'agit d'une enquête supplémentaire à participation volontaire réalisée mensuellement auprès des ménages répondants à l'EPA qui en sont à leur deuxième mois de participation. Une fois l'interview de l'EPA terminée, une personne de 18 ans ou plus est sélectionnée au hasard parmi les membres du ménage et la personne choisie doit répondre au questionnaire de l'EVRC.

9.3.4 Enquête canadienne sur le revenu

Le principal objectif de l'ECR est de fournir des renseignements sur le revenu et les sources de revenu des Canadiens, de même que sur leurs caractéristiques personnelles et celles de leur ménage. Dans le cadre de l'enquête, on recueille de l'information au sujet de l'activité sur le marché du travail, de la fréquentation scolaire, de l'incapacité, des paiements de pension alimentaire, des dépenses de soins aux enfants, des transferts entre ménages, du revenu personnel et des caractéristiques et des coûts du logement. Il s'agit d'une enquête supplémentaire menée de janvier à avril auprès des ménages répondants à l'EPA qui en sont à leur sixième mois de participation. Après l'interview de l'EPA, et compte tenu des contraintes opérationnelles, l'intervieweur demande au membre du ménage qui a fourni l'information dans le cadre de l'EPA de répondre au questionnaire de l'ECR pour tous les membres du ménage de 16 ans et plus.

9.3.5 Enquête sur la couverture de l'assurance-emploi

Le principal objectif de l'ECAE est d'étudier la couverture du programme de l'assurance-emploi. Elle dresse un portrait détaillé des personnes sans emploi ou en situation de sous-emploi selon qu'elles ont ou n'ont pas accès aux prestations d'assurance-emploi. L'Enquête sur la couverture de l'assurance-emploi touche également les prestations de maternité et parentales. L'ECAE est une enquête utilisant des ménages qui ne font plus partie de l'EPA et se servant des groupes de renouvellement de l'EPA qui terminent leur participation de six mois à l'EPA en mars, juin, octobre ou décembre. Les mères de quatre groupes de renouvellement additionnels (un par cycle de collecte) sont également sélectionnées pour obtenir un échantillon d'une taille adéquate. Il y a quatre périodes de collecte par année : avril-mai, juillet-août, novembre-décembre et janvier-février. Chaque cycle dure cinq semaines et commence au cours du mois suivant le mois de référence (le dernier mois de participation à l'EPA).

Références

- Alexander, C.H., Ernst, L.R. et Haas, M.E. (1982). A system for replacing primary sampling units when the units have been exhausted. *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 211-216.
- Benhin, E. et Mantel, H. (2012). Proposed Variance Estimation Method for the Labour Force Survey – from Jackknife to Bootstrap. *Article présenté au Comité consultatif sur les méthodes statistique, octobre 2012*.
- Bocci, C., et Beaumont, J.-F. (2004). Longitudinal Hot-deck Imputation for Household Nonresponse in the LFS. *Document interne, Division des méthodes d'enquêtes auprès des ménages, Statistique Canada*.
- Bocci, C., et Beaumont, J.-F. (2005). A Refinement of the Regression Composite Estimator in the Labour Force Survey. *Document interne, Division des méthodes d'enquêtes auprès des ménages, Statistique Canada*.
- Brisebois, F. et Mantel, H. (1996). Month-in-sample effects for the Canadian Labour Force Survey. *Recueil de la Section des méthodes d'enquête, Société statistique du Canada*.
- Chen, E.J., et Liu, T.P. (2002). Choices of Alpha Value in Regression Composite Estimation for the Canadian Labour Force Survey: Impacts and Evaluation. *Document de travail de méthodologie, HSMD 2002-05E, Statistique Canada*.
- Cochran, W.G. (1977). *Sampling Techniques*, 3rd edition, New York: John Wiley and Sons, Inc.
- Drew, J.D., Choudhry, G.H. et Gray, G.B. (1978). Some methods for updating sample survey frames and their effects on estimation. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 62-71.
- Fichier du réseau routier, guide de référence, 2015*. Statistique Canada n° 92-500-G au catalogue.
- Fuller, W.A., et Rao, J.N.K. (2001). Un estimateur composite de regression qui s'applique à l'Enquête sur la population active du Canada. *Techniques d'enquête, 27*, 49-56.
- Gagné, C., Roberts, G., et Keown, L.-A. (2014). Estimation pondérée et estimation de la variance bootstrap pour analyser des données d'enquête: Comment les effectuer dans certains logiciels choisis? *Le Bulletin technique et d'information des Centres de données de recherche, Hiver 2014, vol. 6 n° 1, Statistique Canada, n° 12-002-X au catalogue*.
- Gambino, J., Kennedy, B. et Singh, M.P. (2001). Estimation composite par regression pour l'Enquête sur la population active du Canada : Evaluation et application. *Techniques d'enquête, 27*, 69-79.
- Gray, G. (1973). Rotation of PSUs. *Document interne, Statistique Canada*.
- Kennedy, B. (1998). Weighting and Estimation Methodology of the Canadian Labour Force Survey. *Document de travail de méthodologie, HSMD-98-002E, Statistique Canada*.
- Keyfitz, N. (1951). Sampling with probabilities proportional to size: adjustment for changes in the probabilities. *Journal of the American Statistical Association*, 46, 105-108.
- Laflamme, G. (2003). Sélection et rotation des UPE : quelques précisions concernant la RAM, *Document interne, Division des méthodes d'enquêtes auprès des ménages, Statistique Canada*.
- Lemaître, G., et Dufour, J. (1987). Une méthode intégrée de pondération des personnes et des familles. *Techniques d'enquête, 13*, 211-220.
- Lorenz, P. (1996). Head Office Hot Deck Imputation System Specifications, Version 3. *Document interne, Division des méthodes d'enquêtes auprès des ménages, Statistique Canada*.

- Neusy, E. (2013). Adapting the Coordinated Bootstrap Method in the Presence of Change Between Survey Cycles. *Document interne, Division des méthodes d'enquêtes auprès des ménages, Statistique Canada.*
- Pandey, S., Alavi, A., et Beaumont, J.-F. (2003). Comparison of Integrated and Non-integrated Estimation Methods for GREG and Composite Estimators. *Document interne, Division des méthodes d'enquêtes auprès des ménages, Statistique Canada.*
- Phillips, O. (2004). Comment utiliser les poids bootstrap avec WesVar et SUDAAN. *Le Bulletin technique et d'information des Centres de données de recherche, Automne 2004, vo1. 6 n° 2, Statistique Canada, n° 12-002-X au catalogue.*
- Rao, J.N.K., Hartley, H.O. et Cochran, W.G. (1962). A simple procedure for unequal probability sampling without replacement. *Journal of the Royal Statistical Society, B, 24, 482-491.*
- Rao, J.N.K. et Wu, C.F.J. (1988). Resampling Inference with Complex Survey Data. *Journal of American Statistical Association, 83, 231-241.*
- Rao, J.N.K., Wu, C.F.J. et Yue, K. (1992). Quelques travaux récents sur les méthodes de rééchantillonnage applicables aux enquêtes complexes. *Techniques d'enquête, 18, 225-234.*
- Roberts, G., Kovacevic, M., Mantel, H. et Phillips, O. (2001). Cross-sectional Inference Based on Longitudinal Surveys: Some Experiences with Statistics Canada Surveys. *Proceedings of the Federal Committee on Statistical Methodology Research Conference, November 14-16, 2001.*
- Särndal, C.-E., Swensson, B., et Wretman, J. (1992). *Model Assisted Survey Sampling.* New-York: Springer Verlag.
- Section du registre des adresses et des méthodes géographiques (2015). Service de base de sondage pour les enquêtes auprès des ménages Guide de l'utilisateur. *Division des méthodes d'enquêtes sociales, Statistique Canada, document interne.*
- Singh, A.C., Kennedy, B., et Wu, S. (2001). Estimation composite par régression pour l'Enquête sur la population active du Canada avec plan de sondage à renouvellement de panel. *Techniques d'enquête, 27, 35-48.*
- Statistique Canada (2003). Méthodes et pratiques d'enquête. Statistique Canada, n° 12-587-X au catalogue.
- Statistique Canada (2012). *Dictionnaire du recensement.* Statistique Canada, n° 98-301-X au catalogue.
- White, S., et Benhin, E. (2013). Suggested Improvements for the LFS Person-Level Imputation. *Document interne, Division des méthodes d'enquêtes auprès des ménages, Statistique Canada.*

Annexe A.1 Glossaire

De l'information plus détaillée sur un bon nombre des termes dans le présent glossaire se trouve aux liens suivants.

Dictionnaire du recensement :

<http://www12.statcan.gc.ca/census-recensement/2011/ref/dict/98-301-X2011001-fra.pdf>

Catalogue de la Géographie :

<http://www.statcan.gc.ca/pub/92-196-x/92-196-x2011001-fra.pdf>

Guide de l'Enquête sur la population active :

<http://www.statcan.gc.ca/pub/71-543-g/71-543-g2015001-fra.pdf>

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. C'est la plus petite région géographique normalisée pour laquelle toutes les données du recensement sont diffusées. Les AD couvrent tout le territoire du Canada. Elles sont créées comme des unités de diffusion de données du recensement.

Base aréolaire

Une base d'unités fondée sur des régions géographiques, comme des aires de diffusion ou des unités géographiques similaires. Elle est habituellement utilisée en l'absence d'une base liste des unités ultimes de l'enquête.

Centre de population

Un centre de population contient une concentration démographique d'au moins 1 000 habitants et une densité de population de 400 habitants ou plus au kilomètre carré selon les chiffres de population du recensement actuel. Toutes les régions situées à l'extérieur des centres de population sont classées dans la catégorie des régions rurales. Ensemble, les centres de population et les régions rurales couvrent l'ensemble du Canada. La population des centres de population comprend toute la population vivant dans les noyaux, les noyaux secondaires et les banlieues des régions métropolitaines de recensement (RMR) et des agglomérations de recensement (AR) ainsi que la population vivant dans les centres de population à l'extérieur des RMR et des AR.

Chômage

Les chômeurs sont les personnes qui, au cours de la semaine de référence :

- a. n'avaient pas de travail, mais avaient cherché activement du travail au cours des quatre dernières semaines se terminant avec la période de référence et étaient disponibles pour travailler; ou
- b. avaient été mises à pied temporairement à cause de la conjoncture économique, mais s'attendaient à retourner au travail et étaient disponibles pour travailler ; ou
- c. n'avaient pas de travail, avaient un emploi devant commencer dans les quatre semaines après la période de référence et étaient disponibles pour travailler.

Côté d'îlot

Le côté d'îlot correspond à un côté de rue situé entre deux traits consécutifs qui coupent cette rue. Ces traits peuvent être d'autres rues ou des frontières de régions géographiques normalisées. Les côtés d'îlot servent à produire des points représentatifs de côté d'îlot, qui sont utilisés pour le géocodage et l'extraction de données du recensement lorsque les adresses de voirie sont connues.

Division de recensement

Terme générique qui désigne les régions établies par la loi provinciale (comme les comtés, les municipalités régionales de comté et les districts régionaux) ou leurs équivalents. Les divisions de recensement sont des

régions géographiques intermédiaires se situant entre la municipalité (subdivision de recensement) et la province/territoire.

Échantillonnage à plusieurs degrés

L'échantillonnage à plusieurs degrés est le processus de sélection d'un échantillon à deux degrés successifs ou plus. Les unités sélectionnées au premier degré sont appelées les unités primaires d'échantillonnage (UPE). Les unités sélectionnées au deuxième degré sont appelées les unités secondaires d'échantillonnage (USE) ou parfois unités ultimes d'échantillonnage s'il s'agit de la dernière étape. Les unités à chaque étape ont une structure différente et sont hiérarchiques. Dans le cas de l'EPA, les UPE correspondent aux aires de diffusion sélectionnées dans les strates. À la deuxième étape, les logements sont sélectionnés dans chaque UPE sélectionnée à la première étape.

Échantillonnage avec probabilité proportionnelle à la taille

L'échantillonnage avec probabilité proportionnelle à la taille (PPT) est une technique qui utilise de l'information auxiliaire et qui donne lieu à une probabilité inégale d'inclusion. Si les unités de la population sont de différentes tailles et si ces tailles sont connues, cette information peut servir pendant l'échantillonnage pour accroître l'efficacité statistique. Dans le cas de l'EPA, les UPE sont sélectionnées avec PPT, et la mesure de la taille utilisée est liée au nombre approximatif de ménages dans chaque UPE. Pour plus d'information sur la mesure de la taille, voir la section 6.2.1.

Échantillonnage systématique

L'échantillonnage systématique est une méthode de sélection d'un échantillon dans lequel le premier élément est sélectionné aléatoirement dans une population (origine choisie au hasard) et dont les éléments restants sont tirés à des intervalles échelonnés également (taux d'échantillonnage inverse). Avec un taux d'échantillonnage de 1 pour 10, et un point de départ au hasard à 7, la 7^e unité est sélectionnée. Par la suite chaque 10^e unité est choisie (la 17^e, la 27^e, la 37^e, etc.) jusqu'à la fin de la liste.

Effet de plan

L'effet de plan d'un estimateur est le ratio de la variance réelle d'un estimateur en vertu du plan de sondage actuel par rapport à ce qu'elle serait en vertu d'un échantillonnage aléatoire simple stratifié comportant le même nombre d'éléments.

Emploi

Les personnes occupées sont celles qui, au cours de la semaine de référence :

- a. ont fait un travail quelconque dans le cadre d'un emploi ou dans une entreprise, c'est-à-dire tout travail rémunéré accompli pour un employeur ou à son propre compte. Cela comprend aussi le travail familial non rémunéré, c'est-à-dire un travail (non rémunéré) qui contribue directement à l'exploitation d'une ferme, d'une entreprise ou d'un cabinet de professionnel appartenant à un membre apparenté du même ménage et exploité par lui; ou
- b. avaient un emploi, mais n'étaient pas au travail à cause d'une maladie ou d'une incapacité, pour obligations personnelles ou familiales, pour des vacances ou à la suite d'un conflit de travail. Cette catégorie n'inclut pas les personnes mises à pied ou entre deux emplois occasionnels, ni celles qui n'avaient pas d'emploi à ce moment-là (même si elles avaient un emploi devant commencer à une date ultérieure).

Estimation de la moyenne mobile sur trois mois

Une estimation de la moyenne mobile sur trois mois est une moyenne des valeurs de l'estimation pour chacun des trois derniers mois. Elle peut être produite tous les mois, en fonction des trois derniers mois. La moyenne est définie d'une manière légèrement différente selon que l'estimation est un total ou un taux. L'expression mathématique est donnée à la section 6.1.

Groupe du Registre des adresses (RA)

L'extrait du Fichier de l'univers des logements utilisé par l'EPA a un numéro d'identification de l'UPE et un numéro d'ordre appliqués à chaque logement. L'intégralité et l'exactitude de l'extrait au niveau de l'UPE est estimé afin d'attribuer à l'UPE un des trois numéros de groupe.

- 0 : Les UPE dans le secteur d'envoi par la poste sans listage initial. La liste de logements est réputée de bonne qualité. La sélection est effectuée à partir de la liste actuelle. Des cartes papier pour la tenue à jour de la liste sont générées sur demande seulement.
- 1 : Les UPE dans le secteur autre qu'un secteur d'envoi par la poste sans liste de départ. La liste de logements est considérée comme étant de bonne qualité. La sélection est effectuée à partir de la liste actuelle. Des cartes papier sont générées automatiquement et peuvent être utilisées pour la tenue à jour de la liste au fil du temps.
- 2 : Les UPE dans le secteur autre qu'un secteur d'envoi par la poste avec liste de départ. La liste est réputée de mauvaise qualité, et le listage de départ a lieu avant la sélection du premier échantillon. Des cartes papier sont générées automatiquement et peuvent être utilisées pour la tenue à jour de la liste au fil du temps.

Îlot

Un îlot, parfois appelé îlot de recensement ou îlot de diffusion, est un territoire dont tous les côtés sont délimités par des rues et/ou des limites des régions géographiques normalisées. Les îlots de recensement couvrent tout le territoire du Canada. Ce sont les plus petites régions géographiques pour lesquelles les chiffres de population et des logements sont stockés. Il convient de souligner que les îlots ne sont pas à la disposition du public, mais sont utilisés à l'interne en tant que plus petit niveau de géographie sur lequel s'appuient les géographies de collecte et de diffusion.

Listage

Le listage est le processus par lequel les logements appartenant à une zone (habituellement une UPE) sont enregistrés sur papier ou sous forme électronique. Cet exercice sur le terrain est nécessaire pour certaines UPE lorsqu'une liste de logements de haute qualité n'est pas disponible pour effectuer la sélection du deuxième degré des logements. Les cartes du secteur, ayant des limites claires, sont nécessaires pour déterminer où le listage est nécessaire. La plupart des listages sont séquencés selon un modèle particulier afin de permettre que tous les côtés de l'îlot soient examinés et pour pouvoir relocaliser une adresse particulière dans un autre mois ou année après la création de la liste de départ.

Logement

Désigne un ensemble de pièces d'habitation dans lequel demeure ou pourrait demeurer une personne ou un groupe de personnes. Les logements inoccupés sont dits vacants. Pour l'EPA, un logement se compose de tout ensemble de pièces d'habitation qui est structurellement isolé et qui a une entrée extérieure privée ou accessible par un vestibule commun ou une cage d'escalier à l'intérieur de l'immeuble.

Logement collectif

Un logement collectif désigne un logement de nature commerciale, institutionnelle ou communautaire où les gens résident, mais où le concept de logement unifamilial est difficile à appliquer. Il comprend les maisons de chambres et pensions, les hôtels, motels et établissements pour touristes, les maisons de soins infirmiers, les hôpitaux, les résidences pour le personnel, les casernes (bases militaires), les camps de chantier, les prisons, les foyers collectifs, etc. Il peut s'agir d'un logement occupé par des résidents habituels ou uniquement par des résidents étrangers et/ou par des personnes présentes temporairement.

Ménage

Toute personne ou tout groupe de personnes vivant dans un logement. Un ménage peut comprendre les combinaisons suivantes : une personne vivant seule, une ou plusieurs familles ou un groupe de personnes qui ne sont pas apparentées mais qui partagent le même logement. Il faut noter que les résidents étrangers et les personnes qui ont leur résidence principale ailleurs ne sont pas visés par l'enquête.

Période de référence

Période de temps utilisée dans les enquêtes dont les répondants doivent se souvenir et sur laquelle ils doivent répondre, par exemple : « combien d'heures avez-vous travaillé la semaine dernière? ». Dans l'EPA, la semaine de référence correspond habituellement à la semaine où tombe le 15 du mois.

Population active

Population civile de 15 ans et plus (à l'exclusion des pensionnaires d'établissements) qui, durant la semaine de référence de l'enquête, était occupée ou en chômage. Avant 1966, l'enquête portait sur les personnes de 14 ans et plus.

Population cible

La population cible est la population pour laquelle on souhaite obtenir de l'information. La population cible de l'EPA correspond à l'ensemble des personnes de 15 ans et plus qui résident dans les provinces du Canada, à l'exception des suivantes : les personnes qui vivent dans les réserves indiennes, les membres à temps plein des Forces armées et les pensionnaires d'établissements (par exemple, les personnes détenues dans les pénitenciers et les patients d'hôpitaux ou de maisons de repos qui ont résidé dans l'établissement pendant plus de six mois).

Région économique

Une région économique (RE) est constituée d'un groupe de divisions de recensement (DR) entières (sauf pour un cas en Ontario). Ces régions sont créées comme une unité géographique normalisée et servent à l'analyse de l'activité économique régionale. Elles ont été établies avec le concours des provinces, sauf le Québec, où les régions économiques sont désignées par la loi (« les régions administratives »). Les RE correspondent généralement aux régions utilisées par la province à des fins administratives et statistiques. Les limites utilisées actuellement sont basées sur la Classification géographique type de 2011.

Région économique de l'assurance-emploi

Ensemble de régions dans tout le pays définies par Emploi et Développement social Canada (EDSC) en vue de la répartition équitable des prestations de l'assurance-emploi (EDSC). L'EPA est responsable de produire des estimations récentes requises par EDSC afin d'établir des normes pour l'admissibilité au programme et pour la durée des prestations.

Région métropolitaine de recensement ou l'agglomération de recensement

Une région métropolitaine de recensement (RMR) ou une agglomération de recensement (AR) est formée d'une ou de plusieurs municipalités adjacentes situées autour d'un centre de population (aussi appelé le noyau urbain). Une RMR doit avoir une population totale d'au moins 100 000 habitants et son noyau doit compter au moins 50 000 habitants. Quant à l'AR, son noyau doit compter au moins 10 000 habitants. Pour être incluses dans une RMR ou une AR, les autres municipalités adjacentes doivent avoir un degré d'intégration élevé avec le noyau, lequel est déterminé par le pourcentage de navetteurs établi d'après les données du recensement précédent sur le lieu de travail.

Région rurale

Les régions rurales comprennent tout le territoire situé à l'extérieur des centres de population. Ensemble, les centres de population et les régions rurales couvrent l'ensemble du Canada. La population rurale comprend toutes les personnes qui vivent dans les régions rurales des régions métropolitaines de recensement (RMR) et des agglomérations de recensement (AR) ainsi que les personnes qui vivent dans les régions rurales à l'extérieur des RMR et des AR.

Registre des adresses (RA)

Le Registre des adresses est une base de données d'adresses résidentielles tenu à jour pour appuyer le recensement et d'autres enquêtes-ménages. La base de données contient plus de 15 000 000 adresses et a une couverture nationale, bien qu'elle soit plus précise dans les centres de population. Les trois principaux produits dérivés du RA sont le Fichier de l'univers des logements (FUL), le Fichier de numéros de téléphone résidentiels (FTR) et le Fichier socio-économique (FSE).

Renouvellement

Le renouvellement de l'échantillon est le remplacement périodique d'une unité par une autre. L'EPA

- effectue le renouvellement des logements (dans une UPE) après six mois dans l'enquête
- effectue le renouvellement des UPE après de deux à 50 ans dans l'enquête, la moyenne se situant aux alentours de 10 ans. Dans bien des cas, un remaniement de l'enquête aura lieu avant la date de renouvellement prévue pour l'UPE.

L'ensemble de logements (ou les UPE qui les contiennent) qui se renouvelle le même mois est appelé panel de renouvellement ou groupe de renouvellement. Chaque panel se compose d'un sixième de l'échantillon. Ainsi, chaque mois, il y a un mélange de logements qui ont leur première, deuxième, troisième, quatrième, cinquième et sixième interview.

Répartition

La répartition s'entend du processus consistant à diviser une taille fixe de l'échantillon en plusieurs zones provinciales et/ou infraprovinciales afin de satisfaire à diverses contraintes en ce qui concerne les coûts de la collecte et/ou la fiabilité des estimations.

Situation vis-à-vis de l'activité

Chaque enquêté de 15 ans ou plus est classé dans une catégorie de la population active (personne occupée, chômeur ou inactif), d'après les réponses qu'il a fournies à un certain nombre de questions durant l'interview.

Stratification

La stratification est le processus qui consiste à diviser les unités primaires d'échantillonnage de l'enquête en groupes absolument exclusifs homogènes appelés strates, après quoi les échantillons sont sélectionnés indépendamment à partir de chaque strate.

Taux d'activité

Le taux d'activité représente la population active exprimée en pourcentage de la population de 15 ans et plus. Le taux d'activité d'un groupe donné (âge, sexe, etc.) correspond à l'effectif de la population active dans ce groupe exprimé en pourcentage de la population totale de ce groupe.

Taux d'échantillonnage

Le taux d'échantillonnage est le rapport de la taille de l'échantillon à la taille de la population dans la base de sondage. Un échantillon de 1 sur 20 sélectionnerait 5 % des unités pour la collecte des données et aurait un taux d'échantillonnage à 0,05 ou un taux d'échantillonnage inverse de 20.

Taux de chômage

Nombre de chômeurs exprimé en pourcentage de la population active. Le taux de chômage pour un groupe donné (par exemple, l'âge, le sexe, l'état matrimonial) correspond au nombre de chômeurs dans ce groupe exprimé en pourcentage de la population active dans ce groupe. Ce taux est l'une des statistiques clés produites par l'EPA.

Taux de glissement

Le taux de glissement est une mesure de l'écart entre une estimation de la taille d'une population (p. ex., province ou groupe âge-sexe) et la valeur correspondante fondée sur les projections du recensement. Il est égal à $1 - (\text{ratio des estimations de la projection}) \times 100 \%$.

Taux d'emploi (rapport emploi-population)

Nombre de personnes occupées exprimé en pourcentage de la population de 15 ans et plus. Le taux d'emploi d'un groupe particulier (selon l'âge, le sexe, l'état matrimonial, la province, etc.) correspond au nombre de personnes occupées dans ce groupe exprimé en pourcentage de l'effectif de ce groupe.

Taux d'inoccupation

Le taux d'inoccupation est la proportion de logements inoccupés. Les logements inadmissibles, comme les entreprises et les logements démolis, ne sont pas inclus dans le dénominateur.

Tenue de liste

L'opération de listage de l'EPA s'effectue en deux étapes. La seconde étape est la tenue continue de la liste préexistante. La liste était générée à l'origine par le listage initial (groupe RA 2) ou par le chargement initial (groupe RA 0 et groupe RA 1). L'étendue des changements est habituellement mineure sauf si une croissance significative se produit dans le secteur de l'UPE, ou si des erreurs significatives sont trouvées dans le travail de listage précédent ou le chargement initial. Les mises à jour sont envoyées directement au bureau central sans l'intervention des interviewers principaux, sauf si le sous-échantillonnage est requis en raison d'une croissance significative.

Unité primaire d'échantillonnage

Les unités choisies au premier degré d'échantillonnage dans un plan à plusieurs degrés sont appelées unités primaires d'échantillonnage, ou UPE. Depuis le remaniement de 2015, les UPE de l'EPA sont principalement définies comme des aires de diffusion.

Variance d'échantillonnage

La variance d'échantillonnage mesure le degré de divergence entre les estimations d'un paramètre de population, produites à partir de tous les échantillons possibles. Elle est calculée comme la valeur moyenne de la différence au carré de cette estimation et de sa moyenne pour l'ensemble de tous les échantillons possibles.

Annexe A.2 Abréviations

AD	Aire de diffusion (recensement)
AR	Agglomération de recensement
BNDG	Base nationale de données géographiques
CA	Composante du contenu additionnel
CANSIM	Principale base de données socioéconomiques de Statistique Canada
CNP	Classification nationale des professions
CNP-S	Classification nationale des professions pour statistiques
CV	Coefficient de variation
EAS	Échantillonnage aléatoire simple
ECAE	Enquête sur la couverture de l'assurance-emploi
ECR	Enquête canadienne sur le revenu
EDC	Composante des effets de désastre ou catastrophe
EDM	Enquête sur les dépenses des ménages
EDSC	Emploi et Développement social Canada
EPA	Enquête sur la population active
ESCC	Enquête sur la santé dans les collectivités canadiennes
ET	Erreur type
EVRC	Enquête sur les voyages des résidents du Canada
FSI	Fraction de sondage inverse
FTR	Fichier de numéros de téléphone résidentiels
FUL	Fichier de l'univers des logements
IPAO	Interview sur place assistée par ordinateur
IPC	Indice des prix à la consommation
ITAO	Interview téléphonique assistée par ordinateur
IWAO	interview Web assistée par ordinateur
LSI	Logements sélectionnés par les intervieweurs

PCT	Premier contact par téléphone
PIB	Produit intérieur brut
PPT	Probabilité proportionnelle à la taille
PPTAR	Probabilité proportionnelle à la taille avec remplacement
PPTSR	Probabilité proportionnelle à la taille sans remplacement
QE	Questionnaire électronique
RA	Registre des adresses
RE	Région économique
REAE	Région économique de l'assurance-emploi
RHC	Rao-Hartley-Cochran (méthode des groupes aléatoires)
RMR	Région métropolitaine de recensement
SCIAN	Système de classification des industries de l'Amérique du Nord
SDR	Subdivision de recensement
SGC	Système généralisé de cartographie
SIHD	Système d'imputation hot deck
STBC	Système de traitement du bureau central
UPE	Unité primaire d'échantillonnage
USE	Unité secondaire d'échantillonnage

Annexe B Caractéristiques de la base de sondage et du plan de sondage

Tableau B.1
Nombre de ménages inclus dans la base de sondage et taille des échantillons provinciaux

Province	Ménages compris dans la base de sondage	Ménages exclus de la base de sondage	Échantillon financé par Statistique Canada	Échantillon financé par EDSC	Échantillon total
	nombre				
Terre-Neuve-et-Labrador	211 361	7 878	1 852	157	2 009
Île-du-Prince-Édouard	59 402	277	1 421	0	1 421
Nouvelle-Écosse	408 853	4 281	2 965	0	2 965
Nouveau-Brunswick	335 516	5 536	2 623	187	2 810
Québec	3 620 953	23 468	5 108	5 077	10 185
Ontario	5 108 516	49 030	7 036	7 936	14 972
Manitoba	467 146	18 399	3 911	295	4 206
Saskatchewan	400 608	25 012	3 904	218	4 122
Alberta	1 487 737	25 183	3 690	810	4 500
Colombie-Britannique	1 881 831	57 057	3 507	1 920	5 427
Canada	13 981 923	216 121	36 017	16 600	52 617

Tableau B.2
Nombre de ménages dans les strates éloignées, les strates à grosses grappes et les strates à un degré, par province

Province	Strates éloignées		Strates à grosses grappes		Strates à un degré	
	Strates	Ménages	Strates	Ménages	Strates	Ménages
	nombre					
Terre-Neuve-et-Labrador	0	0	0	0	0	0
Île-du-Prince-Édouard	0	0	0	0	138	59 402
Nouvelle-Écosse	0	0	0	0	0	0
Nouveau-Brunswick	0	0	0	0	0	0
Québec	0	0	0	0	0	0
Ontario	5	45 370	4	105 612	0	0
Manitoba	5	12 751	0	0	0	0
Saskatchewan	1	4 100	0	0	0	0
Alberta	1	3 847	0	0	0	0
Colombie-Britannique	5	45 366	0	0	0	0
Canada	17	111 434	4	105 612	138	59 402

Tableau B.3.1
Statistiques pour les strates à revenu élevé

Région métropolitaine de recensement (RMR)	Strates à revenu élevé	Ménages dans les strates à revenu élevé	Prévalence ¹ des ménages à revenu élevé ² dans les strates à revenu élevé	Prévalence des ménages à revenu élevé dans la RMR	Ménages à revenu élevé dans la RMR qui sont dans une strate à revenu élevé
	nombre		%		
St. John's	1	5 998	41,1	14,5	19,8
Halifax	2	14 173	36,1	11,0	26,4
Saguenay	1	5 436	22,8	6,7	24,7
Québec	1	20 720	34,9	8,1	24,2
Sherbrooke	1	6 224	23,7	5,7	26,2
Trois-Rivières	1	5 118	24,4	5,5	29,7
Montréal	3	117 935	37,2	9,1	28,1
Ottawa-Gatineau	2	44 015	48,4	17,5	22,8
Kingston	1	3 722	40,9	12,0	18,6
Oshawa	1	11 087	41,6	15,8	21,3
Toronto	6	159 618	51,2	16,9	23,0
Hamilton	1	17 896	47,1	13,8	20,6
St. Catharines - Niagara	1	12 805	28,3	8,9	24,1
Kitchener - Waterloo	1	12 927	45,0	13,0	23,4
London	1	13 367	40,8	10,6	25,3
Windsor	1	11 131	33,0	10,0	28,0
Greater Sudbury	1	4 201	41,7	13,1	18,3
Thunder Bay	1	3 300	35,6	10,8	20,2
Winnipeg	4	24 268	37,9	10,0	30,1
Regina	1	5 138	54,2	16,2	18,8
Saskatoon	2	9 023	45,1	14,6	25,3
Calgary	1	23 880	64,6	23,4	13,0
Edmonton	1	24 449	56,0	19,6	14,2
Abbotsford	1	4 845	28,2	10,4	20,4
Vancouver	3	66 687	40,5	14,3	19,4
Victoria	1	9 662	32,7	11,2	17,1

1. La prévalence déclarée est la prévalence d'après le fichier sur la famille T1 de 2011.

2. Un ménage à revenu élevé est un ménage ayant un revenu du ménage annuel déclaré de plus de 150 000 \$.

Tableau B.3.2
Statistiques pour les strates d'immigrants

Province	Strates des immigrants	Ménages dans les strates d'immigrants	Prévalence des ménages d'immigrants ¹ dans les strates d'immigrants	Prévalence des ménages d'immigrants dans la province	Ménages d'immigrants dans la province qui sont dans une strate d'immigrants
	nombre		%		
Manitoba	2	11 861	16,1	6,3	6,5

1. Un ménage d'immigrants est un ménage pour lequel au moins un membre a déclaré avoir immigré au Canada au cours des 10 dernières années, selon l'Enquête nationale auprès des ménages de 2011.

Tableau B.3.3
Statistiques pour les strates autochtones

Province	Strates autochtones	Ménages dans les strates autochtones	Prévalence de ménages autochtones ¹ dans les strates d'Autochtones	Prévalence de ménages autochtones dans la province	Ménages autochtones dans la province qui sont dans une strate d'Autochtones
	number		%		
Saskatchewan	3	13 743	36,5	10,9	11,5
Alberta	8	123 679	20,2	5,9	28,3
Colombie-Britannique	9	143 433	17,1	5,3	24,7

1. Un ménage autochtone est un ménage où au moins un membre a déclaré avoir un statut autochtone d'après l'Enquête nationale auprès des ménages de 2011.

Tableau B.4
Caractéristiques du plan de sondage par région infraprovinciale

Province	UPE	Strates	Population	Logements	Ménages		Ménages échantillonnés	
			nombre		nombre	%	nombre	%
Terre-Neuve-et-Labrador	995	37	493 776	241 034	211 361	100,0	2 009	100,0
REAE								
01	371	14	194 504	90 920	84 973	40,2	692	34,4
02	624	23	299 272	150 114	126 388	59,8	1 317	65,6
RE								
1010	512	19	259 803	124 902	112 432	53,2	971	48,3
1020	76	3	34 938	17 406	14 668	6,9	150	7,5
1030	192	7	95 199	45 846	39 773	18,8	433	21,6
1040	215	8	103 836	52 880	44 488	21,0	455	22,6
Indicateur RMR/AR-Type de plan								
St. John's (RMR)	375	14	196 966	91 949	85 892	40,6	699	34,8
Corner Brook	52	2	27 433	12 105	11 503	5,4	125	6,2
Autres régions urbaines	235	10	118 513	54 344	50 459	23,9	525	26,1
Régions non urbaines	333	11	150 864	82 636	63 507	30,0	659	32,8
Type de strate								
Ordinaire	968	36	477 623	234 782	205 363	97,2	1 960	97,6
À revenu élevé	27	1	16 153	6 252	5 998	2,8	49	2,4
Île-du-Prince-Édouard	NA	138	139 424	69 987	59 402	100,0	1 421	100,0
REAE								
03	NA	66	64 438	31 227	28 392	47,8	679	47,8
62	NA	72	74 986	38 760	31 010	52,2	742	52,2
RE								
1110	NA	138	139 424	69 987	59 402	100,0	1 421	100,0
Indicateur RMR/AR-Type de plan								
Charlottetown	NA	66	64 438	31 227	28 392	47,8	679	47,8
Summerside	NA	18	16 488	7 747	7 342	12,4	176	12,4
Autres régions urbaines	NA	7	7 447	3 925	3 286	5,5	79	5,5
Régions non urbaines	NA	47	51 051	27 088	20 382	34,3	488	34,3
Type de strate								
Un degré	NA	138	139 424	69 987	59 402	100,0	1 421	100,0
Nouvelle-Écosse	1 726	57	909 763	459 722	408 853	100,0	2 965	100,0
REAE								
04	333	12	166 885	88 754	74 769	18,3	654	22,0
05	704	23	362 358	188 303	162 738	39,8	1 242	41,9
06	689	22	380 520	182 665	171 346	41,9	1 069	36,1
RE								
1210	263	10	129 954	67 389	58 419	14,3	532	17,9
1220	304	11	152 348	81 041	69 226	16,9	546	18,4
1230	230	7	121 924	59 335	53 427	13,1	382	12,9
1240	224	7	115 445	63 490	52 170	12,8	409	13,8
1250	705	22	390 092	188 467	175 611	43,0	1 096	37,0
Indicateur RMR/AR-Type de plan								
Halifax (RMR)	705	22	390 092	188 467	175 611	43,0	1 096	37,0
Cap-Breton	195	8	97 398	48 831	44 326	10,8	404	13,6
Truro	84	3	45 041	23 038	20 688	5,1	163	5,5
New Glasgow	72	3	35 342	17 843	16 199	4,0	128	4,3
Autres régions urbaines	213	8	108 377	55 000	50 043	12,2	379	12,8
Régions non urbaines	457	13	233 513	126 543	101 986	24,9	795	26,8
Type de strate								
Ordinaire	1 657	55	868 980	445 069	394 680	96,5	2 877	97,0
À revenu élevé	69	2	40 783	14 653	14 173	3,5	88	3,0

Tableau B.4 (suite)
Caractéristiques du plan de sondage par région infraprovinciale

Province	UPE	Strates	Population	Logements	Ménages		Ménages échantillonnés	
	nombre				nombre	%	nombre	%
Nouveau-Brunswick	1 478	50	737 765	371 662	335 516	100,0	2 810	100,0
REAE								
07	750	24	388 553	188 036	174 494	52,0	1 191	42,4
08	225	9	110 386	57 003	51 141	15,2	595	21,2
09	503	18	238 826	126 623	109 881	32,7	1 024	36,5
RE								
1310	337	12	155 303	78 953	72 071	21,5	697	24,8
1320	398	12	201 789	103 053	91 456	27,3	657	23,4
1330	337	11	168 517	83 684	76 248	22,7	627	22,3
1340	247	9	133 660	66 365	59 181	17,6	440	15,7
1350	159	6	78 496	39 607	36 560	10,9	389	13,8
Indicateur RMR/AR-Type de plan								
Moncton (RMR)	273	8	138 596	67 328	62 885	18,7	408	14,5
Saint John (RMR)	256	8	127 813	62 684	57 608	17,2	392	14,0
Fredericton	165	6	92 914	44 250	41 500	12,4	291	10,3
Bathurst	75	3	33 343	17 717	16 130	4,8	156	5,6
Miramichi	55	2	26 889	13 272	12 157	3,6	118	4,2
Edmundston	45	2	21 698	12 065	11 277	3,4	120	4,3
Autres régions urbaines	169	6	82 037	41 339	38 235	11,4	379	13,5
Régions non urbaines	440	15	214 475	113 007	95 724	28,5	947	33,7
Type de strate								
Ordinaire	1 478	50	737 765	371 662	335 516	100,0	2 810	100,0
Québec	15 296	210	7 831 321	3 914 006	3 620 953	100,0	10 185	100,0
REAE								
10	281	8	135 446	71 862	64 498	1,8	471	4,6
11	1 522	19	756 546	380 465	364 288	10,1	878	8,6
12	316	12	148 215	80 176	74 992	2,1	723	7,1
13	301	13	154 167	78 491	70 221	1,9	813	8,0
14	337	11	169 087	89 244	82 807	2,3	664	6,5
15	1 021	15	520 259	256 370	240 625	6,6	721	7,1
16	7 236	58	3 791 701	1 788 374	1 699 706	46,9	2 042	20,0
17	2 035	18	1 033 672	580 860	497 902	13,8	919	9,0
18	458	13	223 026	128 442	106 911	3,0	705	6,9
19	907	14	440 313	234 837	207 541	5,7	744	7,3
20	570	18	305 660	148 838	138 635	3,8	860	8,4
21	312	11	153 229	76 047	72 827	2,0	644	6,3
RE								
2410	186	6	91 501	47 889	43 092	1,2	352	3,5
2415	424	7	199 492	108 381	96 591	2,7	332	3,3
2420	1 403	16	699 011	364 264	341 854	9,4	756	7,4
2425	810	18	410 444	204 018	188 150	5,2	1 053	10,3
2430	620	18	310 558	170 550	150 125	4,1	1 016	10
2433	484	5	233 940	117 487	110 318	3,0	247	2,4
2435	2 743	27	1 441 851	665 695	636 527	17,6	1 160	11,4
2440	3 682	32	1 886 479	964 739	902 456	24,9	1 162	11,4
2445	722	6	401 555	169 067	163 577	4,5	200	2,0
2450	903	6	468 048	222 711	203 375	5,6	239	2,3
2455	1 055	9	556 460	286 272	248 081	6,9	370	3,6
2460	693	20	366 212	189 489	167 141	4,6	1 004	9,9
2465	288	8	141 432	74 126	68 367	1,9	383	3,8
2470	532	13	259 237	144 550	131 193	3,6	764	7,5
2475	551	14	271 650	139 028	127 857	3,5	818	8,0
2480	167	4	79 621	39 055	36 210	1,0	232	2,3
2490	33	1	13 830	6 685	6 039	0,2	97	1,0

Tableau B.4 (suite)
Caractéristiques du plan de sondage par région infraprovinciale

Province	UPE	Strates	Population	Logements	Ménages		Ménages échantillonnés	
	nombre				nombre	%	nombre	%
Indicateur RMR/AR-Type de plan								
Saguenay (RMR)	318	11	157 063	78 373	74 477	2,1	659	6,5
Québec (RMR)	1 540	19	765 457	384 703	368 299	10,2	889	8,7
Sherbrooke (RMR)	409	14	201 890	108 641	99 134	2,7	795	7,8
Trois-Rivieres (RMR)	322	12	151 593	81 820	76 560	2,1	738	7,2
Granby	159	2	77 077	39 112	36 957	1,0	114	1,1
Saint-Jean-sur-Richelieu	184	3	92 394	43 555	42 305	1,2	130	1,3
Montréal (RMR)	7 297	58	3 824 225	1 802 903	1 713 615	47,3	2 057	20,2
Gatineau (RMR)	586	18	314 227	154 081	142 301	3,9	883	8,7
Rouyn-Noranda/Val-d'Or	142	5	73 398	39 569	37 029	1,0	207	2,0
Saint-Georges	71	3	34 642	17 816	16 622	0,5	196	1,9
Autres régions urbaines	2 333	36	1 149 527	598 274	562 191	15,5	1 828	17,9
Régions non urbaines	1 935	29	989 828	565 159	451 463	12,5	1 688	16,6
Type de strate								
Ordinaire	14 553	202	7 413 606	3 751 224	3 465 520	95,7	9 846	96,7
À revenu élevé	743	8	417 715	162 782	155 433	4,3	339	3,3
Ontario	22 236	309	12 730 148	5 477 602	5 108 516	100,0	14 972	100,0
REAE								
22	1 687	17	935 933	415 258	396 252	7,8	806	5,4
23	782	13	416 250	205 566	182 577	3,6	726	4,9
24	292	12	155 376	73 971	67 834	1,3	873	5,8
25	2 089	18	1 090 705	541 133	462 725	9,1	867	5,8
26	626	12	356 176	141 919	137 164	2,7	591	3,9
27	8 881	79	5 581 914	2 197 692	2 099 508	41,1	2 837	18,9
28	1 326	17	721 048	309 359	296 630	5,8	797	5,3
29	759	14	391 318	182 318	169 399	3,3	633	4,2
30	822	14	450 231	208 966	193 341	3,8	688	4,6
31	520	12	274 109	119 378	111 313	2,2	629	4,2
32	594	11	311 650	139 089	128 883	2,5	549	3,7
33	835	15	476 125	200 660	190 867	3,7	702	4,7
34	610	14	314 289	144 659	133 009	2,6	704	4,7
35	995	18	539 640	232 347	217 496	4,3	806	5,4
36	298	13	155 870	76 469	71 365	1,4	830	5,5
37	237	11	119 688	57 703	53 657	1,1	808	5,4
38	883	19	439 826	231 115	196 496	3,8	1 126	7,5
RE								
3510	2 285	26	1 244 910	563 913	531 768	10,4	1 338	8,9
3515	817	18	445 204	221 807	195 081	3,8	1 168	7,8
3520	653	7	342 068	180 126	149 627	2,9	303	2,0
3530	9 396	90	5 877 168	2 316 567	2 213 911	43,3	3 397	22,7
3540	2 233	33	1 210 144	523 106	483 073	9,5	1 532	10,2
3550	2 564	42	1 363 189	602 342	569 131	11,1	2 018	13,5
3560	1 153	19	629 578	283 731	264 963	5,2	917	6,1
3570	1 188	25	616 906	279 804	258 160	5,1	1 233	8,2
3580	529	6	285 597	140 919	121 284	2,4	302	2,0
3590	1 019	28	520 573	269 509	235 513	4,6	1 682	11,2
3595	399	15	194 811	95 778	86 005	1,7	1 082	7,2

Tableau B.4 (suite)
Caractéristiques du plan de sondage par région infraprovinciale

Province	UPE	Strates	Population	Logements	Ménages		Ménages échantillonnés	
		nombre			nombre	%	nombre	%
Indicateur RMR/AR-Type de plan								
Cornwall	121	2	58 957	27 624	26 541	0,5	104	0,7
Ottawa (RMR)	1 663	17	921 824	409 551	390 751	7,6	795	5,3
Kingston (RMR)	293	12	156 144	74 755	68 162	1,3	877	5,9
Peterborough (RMR)	223	4	117 610	56 754	51 745	1,0	187	1,2
Oshawa (RMR)	626	12	356 176	141 919	137 164	2,7	591	3,9
Toronto (RMR)	8 881	79	5 581 914	2 197 692	2 099 508	41,1	2 837	18,9
Hamilton (RMR)	1 326	17	721 048	309 359	296 630	5,8	797	5,3
St-Catharines/Niagara (RMR)	759	14	391 318	182 318	169 399	3,3	633	4,2
Kitchener/Camb./Wat. (RMR)	835	15	476 125	200 660	190 867	3,7	702	4,7
Brantford (RMR)	246	6	129 288	55 589	53 131	1,0	336	2,2
Norfolk	123	3	62 822	29 088	26 572	0,5	134	0,9
Guelph (RMR)	260	6	140 802	62 481	58 021	1,1	270	1,8
London (RMR)	869	15	474 237	218 577	202 713	4,0	721	4,8
Chatham-Kent	210	5	103 671	48 265	44 715	0,9	239	1,6
Windsor (RMR)	605	12	319 247	142 690	131 879	2,6	562	3,8
Sarnia	178	4	88 915	41 944	39 716	0,8	213	1,4
Barrie (RMR)	365	4	187 013	76 350	71 872	1,4	193	1,3
North Bay	126	3	64 163	31 204	29 118	0,6	149	1,0
Grand Sudbury (RMR)	306	13	160 274	78 566	73 344	1,4	853	5,7
Sault Ste. Marie	163	4	78 693	37 424	35 426	0,7	181	1,2
Thunder Bay (RMR)	239	11	120 736	58 299	54 091	1,1	815	5,4
Leamington	84	2	49 765	20 247	18 928	0,4	101	0,7
Timmins	76	2	43 165	20 284	19 217	0,4	98	0,7
Autres régions urbaines	2 079	29	1 088 446	527 628	483 428	9,5	1 476	9,9
Régions non urbaines	1 580	19	837 795	428 334	335 578	6,6	1 107	7,4
Type de strate								
Ordinaire	20 398	283	11 507 031	5 010 408	4 663 465	91,3	13 755	91,9
À revenu élevé	1 462	17	895 718	302 548	294 069	5,8	738	4,9
À grosses grappes	153	4	223 170	111 764	105 612	2,1	143	1,0
Éloignée	223	5	104 229	52 882	45 370	0,9	336	2,2
Manitoba	2 098	79	1 137 354	504 897	467 146	100,0	4 206	100,0
REAE								
39	1 316	48	723 733	318 220	303 303	64,9	2 400	57,1
40	626	22	336 826	143 707	132 184	28,3	1 092	26,0
41	156	9	76 795	42 970	31 659	6,8	714	17,0
RE								
4610	178	6	101 814	42 443	37 108	7,9	443	10,5
4620	103	4	60 285	23 032	22 102	4,7	178	4,2
4630	208	7	105 934	50 037	44 998	9,6	329	7,8
4640	83	3	44 427	17 168	16 090	3,4	129	3,1
4650	1 220	46	666 760	296 287	283 024	60,6	2 242	53,3
4660	152	6	86 657	41 411	33 835	7,2	407	9,7
4670	84	3	38 953	19 575	17 238	3,7	149	3,5
4680	70	5	32 524	14 944	12 751	2,7	329	7,8
Indicateur RMR/AR-Type de plan								
Winnipeg (RMR)	1 327	48	730 013	320 336	305 366	65,4	2 416	57,4
Brandon	101	4	53 229	24 342	23 241	5,0	170	4,0
Thompson	26	2	12 829	5 486	4 821	1,0	124	3,0
Autres régions urbaines	249	12	135 516	61 082	57 869	12,4	625	14,9
Régions non urbaines	395	13	205 767	93 651	75 849	16,2	870	20,7
Type de strate								
Ordinaire	1 847	68	998 748	452 489	418 266	89,5	3 594	85,4
À revenu élevé	126	4	73 108	24 695	24 268	5,2	192	4,6
D'immigrants	55	2	32 974	12 769	11 861	2,5	91	2,2
Éloignée	70	5	32 524	14 944	12 751	2,7	329	7,8

Tableau B.4 (suite)
Caractéristiques du plan de sondage par région infraprovinciale

Province	UPE	Strates	Population	Logements	Ménages		Ménages échantillonnés	
			nombre		nombre	%	nombre	%
Saskatchewan	1 881	67	946 824	436 428	400 608	100,0	4 122	100,0
REAE								
42	419	15	210 410	96 261	91 367	22,8	868	21,1
43	480	18	260 249	116 976	110 387	27,6	1 056	25,6
44	607	18	285 062	138 379	122 755	30,6	1 038	25,2
45	375	16	191 103	84 812	76 099	19,0	1 160	28,1
RE								
4710	561	19	280 215	129 336	120 764	30,1	1 099	26,7
4720	213	7	95 908	46 827	41 711	10,4	355	8,6
4730	577	20	306 992	138 486	129 710	32,4	1 229	29,8
4740	163	5	76 671	38 633	33 931	8,5	303	7,4
4750	347	15	174 638	78 565	70 392	17,6	1 073	26,0
4760	20	1	12 400	4 581	4 100	1,0	62	1,5
Indicateur RMR/AR-Type de plan								
Regina (RMR)	419	15	210 410	96 261	91 367	22,8	868	21,1
Saskatoon (RMR)	480	18	260 249	116 976	110 387	27,6	1 056	25,6
Moose Jaw	74	3	34 512	16 832	15 895	4,0	135	3,3
Prince Albert	78	3	42 507	17 584	16 323	4,1	249	6,0
Autres régions urbaines	360	13	179 748	83 221	78 000	19,5	841	20,4
Régions non urbaines	470	15	219 398	105 554	88 636	22,1	973	23,6
Type de strate								
Ordinaire	1 728	60	857 680	402 423	368 604	92,0	3 793	92,0
Autochtones	66	3	34 197	14 954	13 743	3,4	131	3,2
À revenu élevé	67	3	42 547	14 470	14 161	3,5	135	3,3
Éloignée	20	1	12 400	4 581	4 100	1,0	62	1,5
Alberta	6 379	88	3 557 377	1 606 911	1 487 737	100,0	4 500	100,0
REAE								
46	2 096	22	1 213 088	531 335	503 607	33,9	1 023	22,7
47	2 039	20	1 156 330	527 074	492 333	33,1	966	21,5
48	402	19	219 533	101 180	85 891	5,8	1 219	27,1
49	1 842	27	968 426	447 322	405 906	27,3	1 292	28,7
RE								
4810	486	10	263 065	120 431	110 508	7,4	458	10,2
4820	370	5	188 248	85 872	78 273	5,3	256	5,7
4830	2 267	23	1 308 684	573 157	542 207	36,4	1 099	24,4
4840	150	1	77 699	40 338	33 612	2,3	83	1,8
4850	369	5	184 575	86 584	78 714	5,3	244	5,4
4860	2 106	21	1 195 404	545 189	508 837	34,2	1 017	22,6
4870	420	10	219 245	100 116	88 673	6,0	646	14,4
4880	211	11	120 457	55 224	46 913	3,2	696	15,5
Indicateur RMR/AR-Type de plan								
Medicine Hat	139	2	71 178	34 740	32 133	2,2	93	2,1
Lethbridge	202	6	106 372	51 205	46 935	3,2	273	6,1
Calgary (RMR)	2 096	22	1 213 088	531 335	503 607	33,9	1 023	22,7
Red Deer	183	2	90 564	42 740	39 968	2,7	124	2,7
Edmonton (RMR)	2 039	20	1 156 330	527 074	492 333	33,1	966	21,5
Wood Buffalo	101	5	62 239	28 574	23 387	1,6	347	7,7
Autres régions urbaines	901	16	478 801	228 491	207 711	14,0	882	19,6
Régions non urbaines	718	15	378 805	162 752	141 663	9,5	792	17,6
Type de strate								
Ordinaire	5 594	77	3 101 778	1 418 114	1 311 882	88,2	4 008	89,1
Autochtones	551	8	302 531	134 499	123 679	8,3	344	7,6
À revenu élevé	213	2	140 572	49 991	48 329	3,2	96	2,1
Éloignée	21	1	12 496	4 307	3 847	0,3	52	1,1

Tableau B.4 (fin)
Caractéristiques du plan de sondage par région infraprovinciale

Province	UPE	Strates	Population	Logements	Ménages		Ménages échantillonnés	
			nombre		nombre	%	nombre	%
Colombie-Britannique	7 909	118	4 267 875	2 048 300	1 881 831	100,0	5 427	100,0
REAE								
50	1 238	17	634 074	337 928	299 073	15,9	790	14,6
51	283	11	169 924	69 132	64 213	3,4	663	12,2
52	4 069	41	2 301 575	1 042 496	977 381	51,9	1 578	29,1
53	676	17	338 615	179 073	165 293	8,8	829	15,3
54	1 069	18	535 082	279 895	251 482	13,4	842	15,5
55	574	14	288 605	139 776	124 389	6,6	725	13,4
RE								
5910	1 456	29	722 820	381 455	350 067	18,6	1 412	26,0
5920	4 659	58	2 630 450	1 193 330	1 111 808	59,1	2 525	46,5
5930	953	12	492 573	259 231	231 736	12,3	557	10,3
5940	285	5	141 501	78 697	67 337	3,6	233	4,3
5950	286	6	145 526	72 346	64 669	3,4	262	4,8
5960	86	2	40 732	20 105	17 610	0,9	123	2,3
5970	70	2	34 835	15 565	13 866	0,7	99	1,8
5980	114	4	59 438	27 571	24 738	1,3	216	4,0
Indicateur RMR/AR-Type de plan								
Kelowna (RMR)	308	5	171 018	88 648	79 290	4,2	249	4,6
Chilliwack	165	3	88 128	40 789	38 117	2,0	154	2,8
Abbotsford - Mission (RMR)	283	11	169 924	69 132	64 213	3,4	663	12,2
Vancouver (RMR)	4 069	41	2 301 575	1 042 496	977 381	51,9	1 578	29,1
Victoria (RMR)	676	17	338 615	179 073	165 293	8,8	829	15,3
Nanaimo	206	3	97 153	49 989	46 729	2,5	144	2,7
Prince George	160	3	83 764	40 156	36 700	2,0	149	2,7
Fort St. John	50	2	26 380	12 041	11 087	0,6	97	1,8
Autres régions urbaines	1 487	23	741 046	381 504	347 422	18,5	1 144	21,1
Régions non urbaines	505	9	250 272	144 472	115 599	6,1	420	7,7
Type de strate								
Ordinaire	6 685	99	3 606 814	1 751 931	1 611 838	85,7	4 558	84,0
Autochtones	610	9	327 646	156 267	143 433	7,6	379	7,0
À revenu élevé	393	5	225 316	85 945	81 194	4,3	206	3,8
Éloignée	221	5	108 099	54 157	45 366	2,4	284	5,2
Canada (provinces seulement)	60 136	1 153	32 751 627	15 130 549	13 981 923	100,0	52 617	100,0
Type de strate								
Ordinaire	54 908	930	29 570 025	13 838 102	12 775 134	91,4	47 201	89,7
Autochtones	1 227	20	664 374	305 720	280 855	2,0	854	1,6
À revenu élevé	3 100	42	1 851 912	661 336	637 625	4,6	1 844	3,5
D'immigrants	55	2	32 974	12 769	11 861	0,1	91	0,2
À grosses grappes	153	4	223 170	111 764	105 612	0,8	143	0,3
À un degré	138	138	139 424	69 987	59 402	0,4	1 421	2,7
Éloignée	555	17	269 748	130 871	111 434	0,8	1 063	2,0

Nota : Voir l'annexe A.2 pour les abréviations

Liste des variables utilisées pour une stratification optimale

La liste de variables utilisées pour une stratification optimale est identique à celle utilisée au cours du dernier remaniement.

Le choix des variables de stratification a été adapté à chacune des régions faisant l'objet d'une stratification optimale. Pour chaque UPE de la région, les variables qui suivent ont été obtenues à partir des données du Recensement de 2011 ou de l'ENM de 2011. Si une variable représentait moins de 2 % de la population totale, elle était écartée. Pour les catégories comme les services, si une sous-catégorie, telle les services financiers, comptait trop peu d'effectif, alors on a plutôt utilisé la variable globale. Une catégorie était considérée comme significative si elle représentait plus de 2 % de la population.

Nombre de personnes occupant un emploi dans les secteurs suivants :

- Agriculture
- Foresterie et pêche
- Exploitation minière
- Fabrication – biens de consommation
- Fabrication – caoutchouc, plastique, cuir
- Fabrication – textiles et vêtements
- Fabrication – meubles, pâte et papier, imprimerie, bois
- Fabrication – métaux et minéraux
- Fabrication – produits pétrochimiques et produits chimiques
- Construction
- Transports
- Services – commerce
- Services – finances
- Services – particuliers/entreprises
- Services – gouvernement

Total des personnes occupées

Revenu total

Population de 15 ans et plus

Population de 15 à 24 ans

Population de 55 ans et plus

Nombre de ménages d'une personne

Nombre de ménages de deux personnes

Nombre de logements possédés

Loyer brut total

Population ayant fait des études secondaires

Langue maternelle : anglais

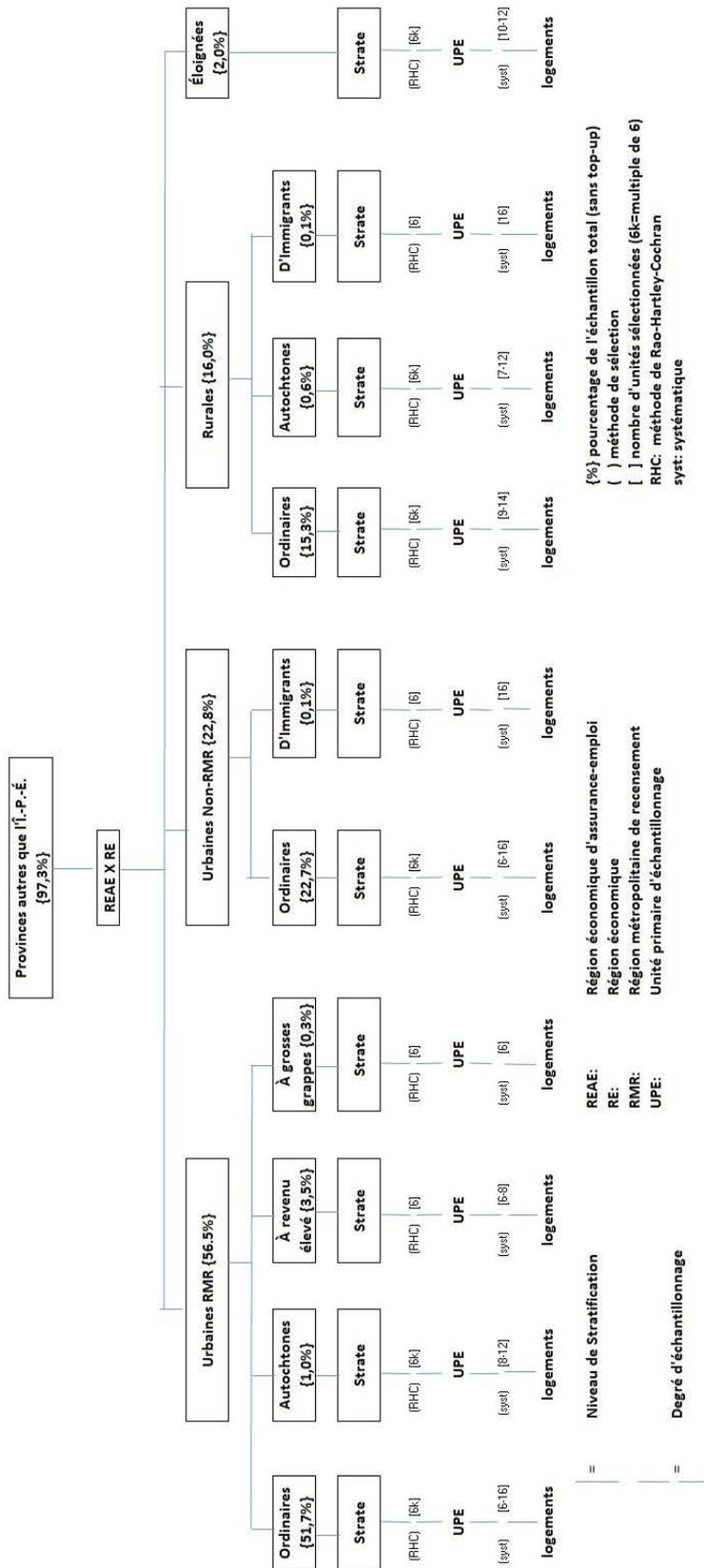
Langue maternelle : français

Langue maternelle autre que le français ou l'anglais

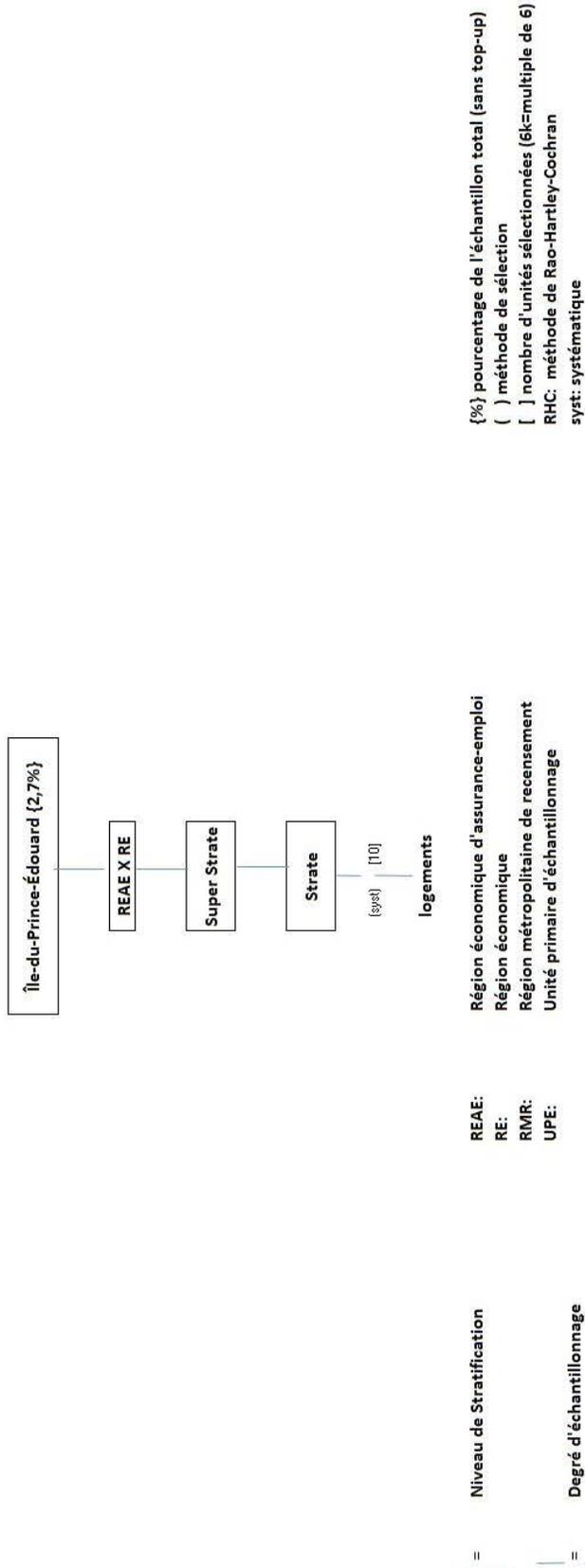
Annexe C

Plan de sondage de l'Enquête sur la population active

Plan d'échantillonnage de l'Enquête sur la population active depuis 2015

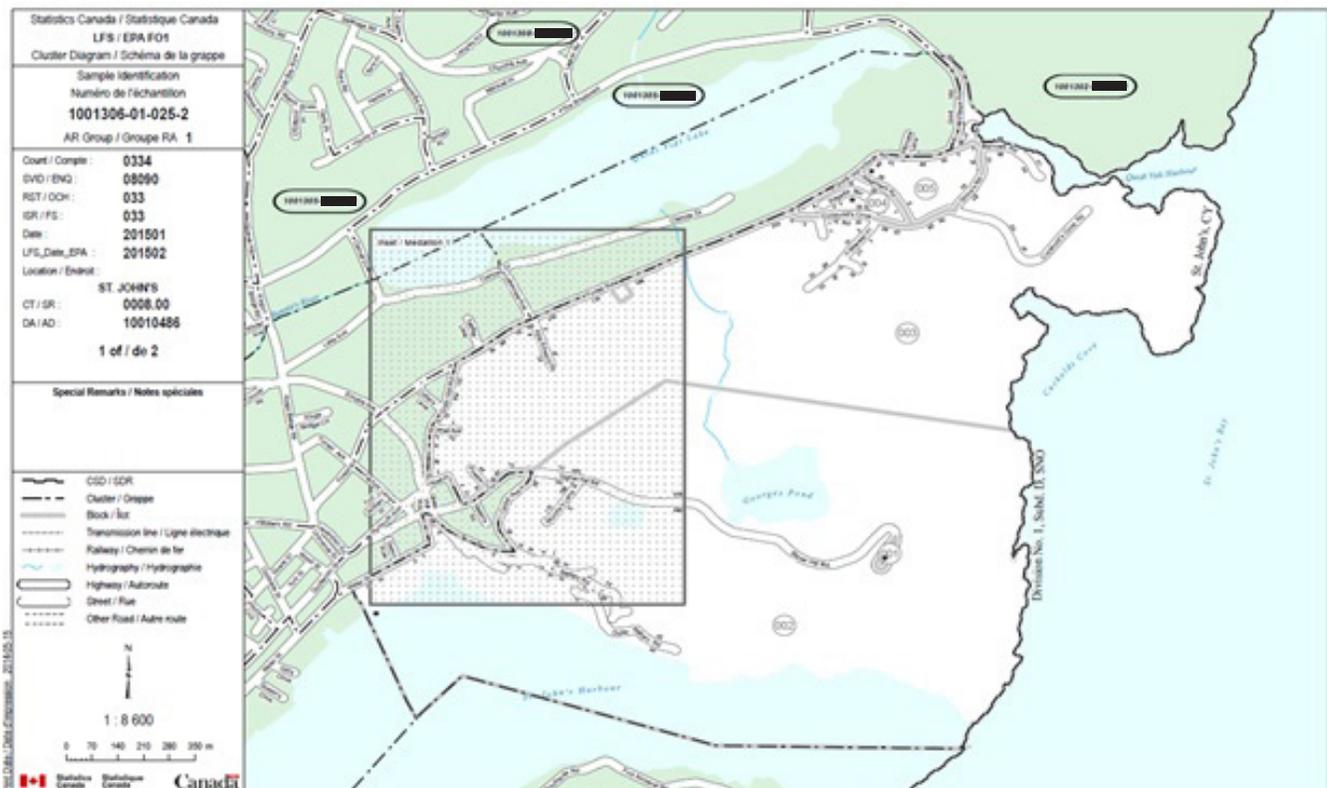


Plan d'échantillonnage de l'Enquête sur la population active depuis 2015



Annexe D Cartes des UPE (diagrammes des grappes F01)

Les cartes des UPE sont générées au moyen du logiciel Système généralisé de cartographie (SGC) en place depuis 2009. La Division de la géographie a également élaboré une application de visualisation des grappes. Il s'agit d'une carte électronique du Canada comportant plusieurs fonctions qui permettent à l'utilisateur de faire un zoom interactif sur des UPE en particulier, de modifier la portée ou l'échelle de la carte, ainsi que de créer des fichiers .pdf qui ressemblent à ceux produits par le SGC. Un tableau d'UPE, ainsi que l'information pertinente pour la légende de la carte, fait partie des principaux intrants du SGC. Toutes les cartes sont vérifiées rigoureusement avant l'envoi aux bureaux régionaux (BR). La taille des cartes et les limites des médaillons sont générées automatiquement, mais elles ne sont peut-être pas toujours pertinentes. La vérification administrative des cartes des UPE est effectuée pour améliorer le produit fini. Le SGC est doté d'outils qui permettent le rajustement de la taille de la carte, en plus de la suppression ou de la création de médaillons. Vous trouverez ci-dessous un exemple de carte d'une UPE. La légende est expliquée à la page suivante, suivie du médaillon (orientation portrait).



Des identificateurs, des dates et des chiffres fictifs ont été utilisés dans cet exemple.

Le numéro d'identification de l'échantillon est l'identificateur de l'UPE, fictif pour cet exemple, (strate 1001306; type de plan 01; UPE 030; groupe de renouvellement 2).

Groupe RA : Classe l'UPE en fonction de son appartenance au secteur d'envoi par la poste pour le recensement et de la qualité prévue du listage du RA (0 = envoi par la poste, haute qualité; 1 = envoi non postal, haute qualité, 2 = envoi non postal, faible qualité)

Compte : Compte d'origine du nombre de ménages dans l'UPE (334)

IDE : ID d'enquête de la première enquête à utiliser l'UPE (8090 est l'ESCC, 10440 est l'EPA, etc.)

OCH : Première origine choisie au hasard utilisée pour la sélection de logements

FSI : Fraction de sondage inverse au niveau de l'UPE utilisée pour la sélection de logements

Date : Date de la première utilisation de cette UPE par l'enquête

LFS_Date_EPA : Date de la première utilisation de l'EPA. Si l'IDE est 10440 = EPA, alors LFS_Date_EPA est établi à 0 par convention.

Lieu : Une municipalité de Postes Canada. Habituellement, la plus courante apparaît dans la liste de logements prospectifs.

SR : Code de secteur de recensement de la base du recensement au moment de l'introduction de l'UPE. Il s'agira au départ du Recensement de 2011, puis du Recensement de 2016 par la suite.

AD : Code d'aire de diffusion de la base du recensement au moment de l'introduction de l'UPE. Il s'agira au départ du Recensement de 2011, puis du Recensement de 2016 par la suite.

N1 of / de N2 : Numéro de la carte et nombre total de cartes pour cette UPE (1 de 2). Deux cartes d'UPE signifie qu'il y a une carte principale (numéro 1) et un médaillon (numéro 2).

Médaillon : Pour les cartes médaillons, numéro de médaillon. Le médaillon n'apparaît pas dans la légende des cartes principales.

Remarques spéciales : Renseignements pertinents supplémentaires au sujet de l'UPE – champ habituellement vide.

Légende des symboles : indique comment les caractéristiques géographiques sont affichées.

Symbole du Nord : pointe toujours vers le haut.

Échelle de la carte.

Date d'impression : Dans les cartes en format paysage, la date d'impression apparaît dans le coin inférieur gauche, à la verticale et à l'extérieur de la légende. Dans les cartes en format portrait, la date d'impression apparaît au milieu du bas de la page, au-dessous de la légende.

La surface à l'extérieur de l'UPE est en vert; la surface intérieure est en jaune.

Les îlots sont numérotés et encadrés.

Le point de départ est indiqué par un astérisque.

Les rectangles des médaillons sont en pointillé gris.

Les UPE voisines sont étiquetées. Certaines UPE en particulier ont été cachées dans l'exemple.

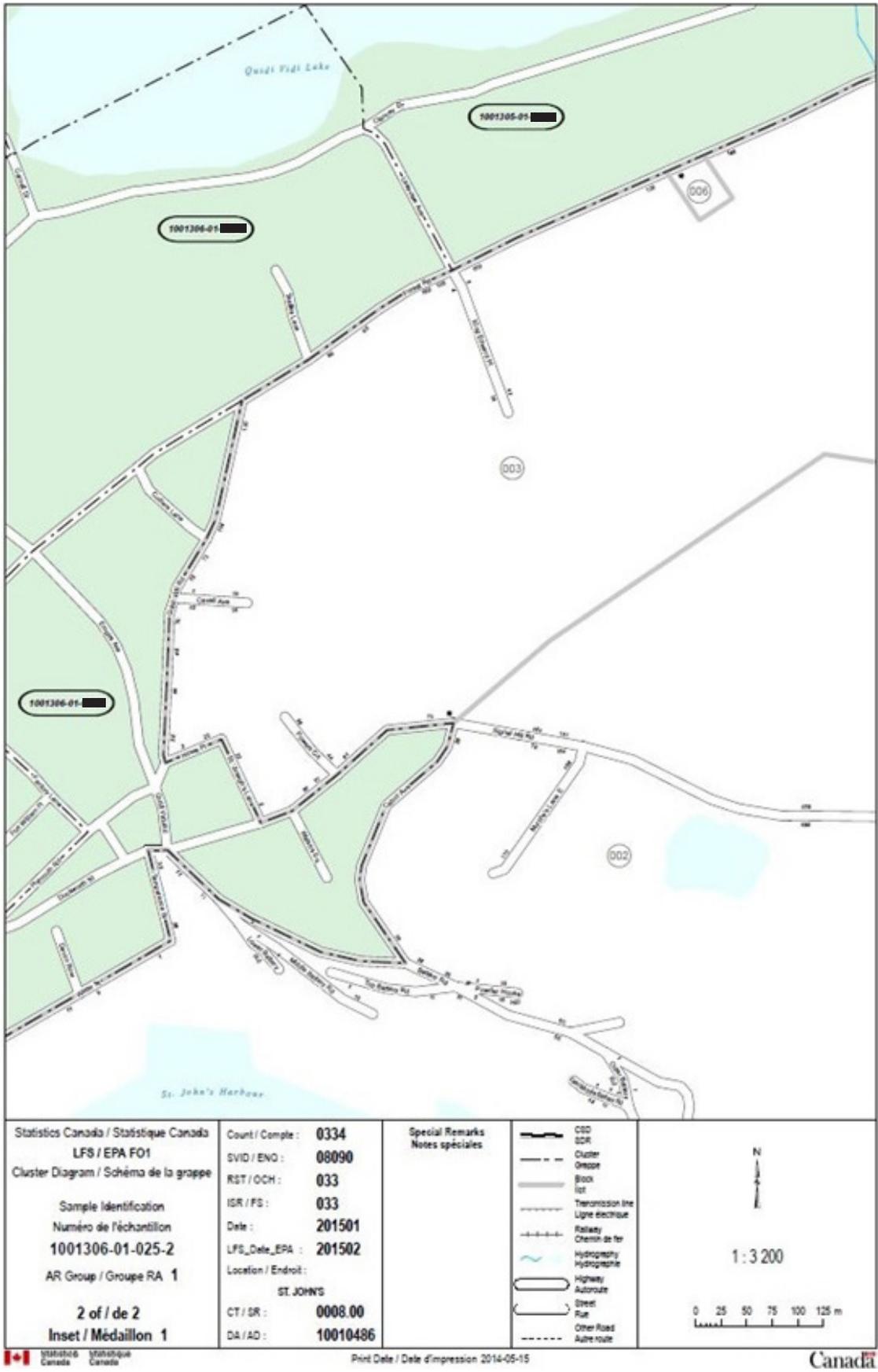
Les tranches d'adresses apparaissent le long des routes.

Les étiquettes de rues sont plus lisibles sur les cartes en tant que telles que sur ces images.

Les cartes peuvent avoir une orientation paysage (légende à gauche) ou portrait (légende au bas).

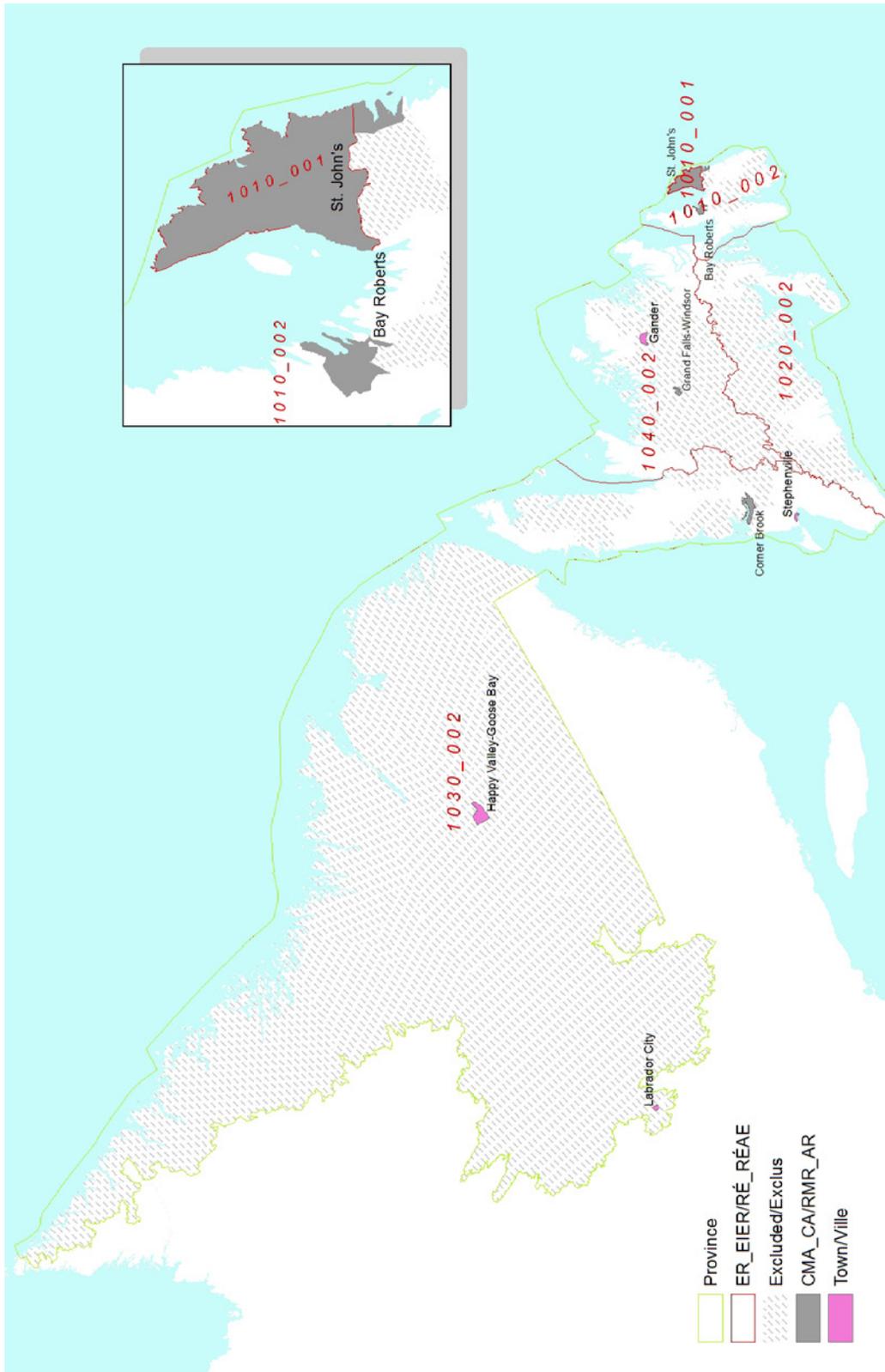
Les tailles de cartes possibles sont 11" x 17", 17" x 22", ou 22" x 34".

Tous les médaillons sont de 11" x 17".

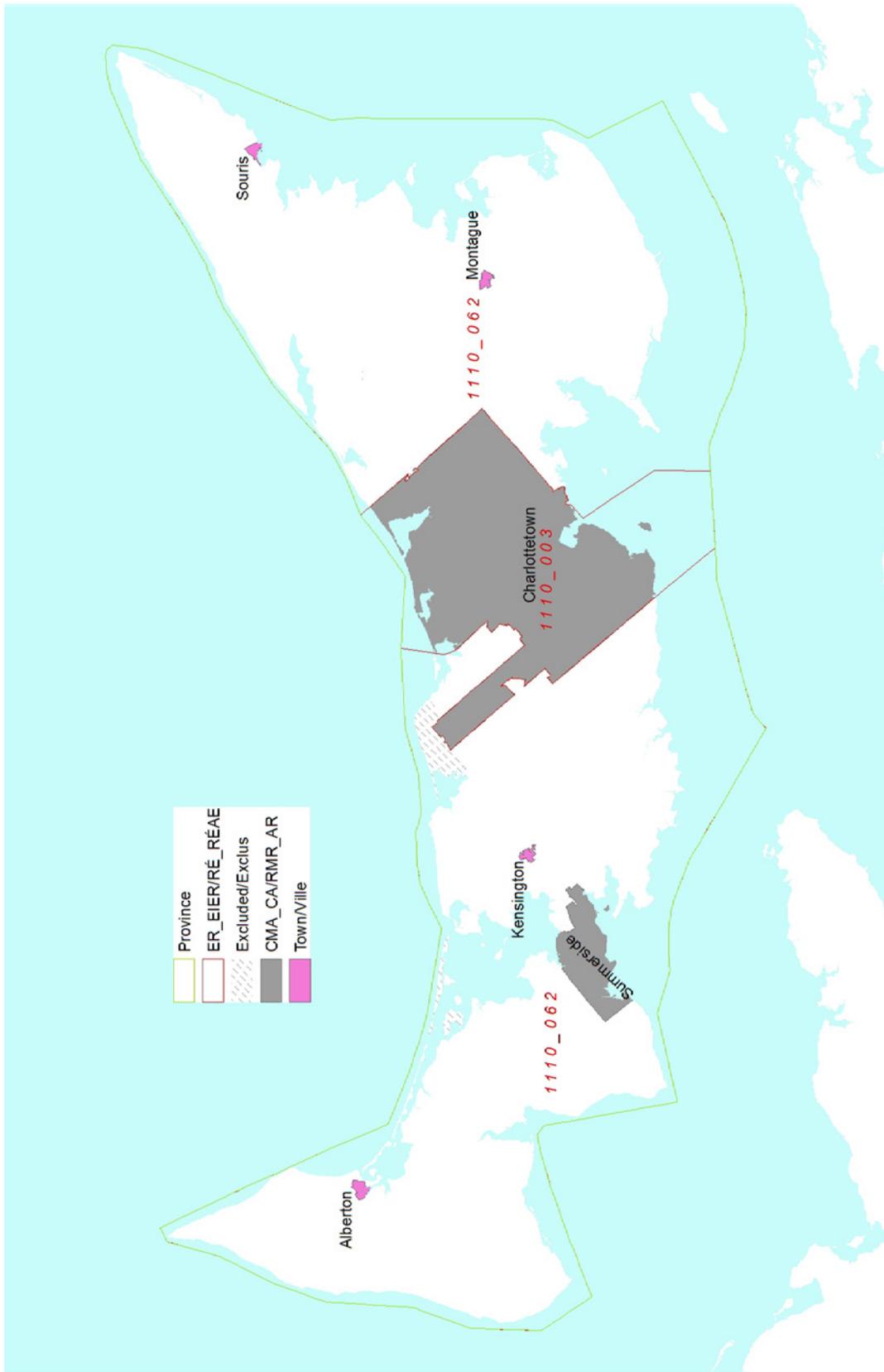


Annexe E Cartes provinciales

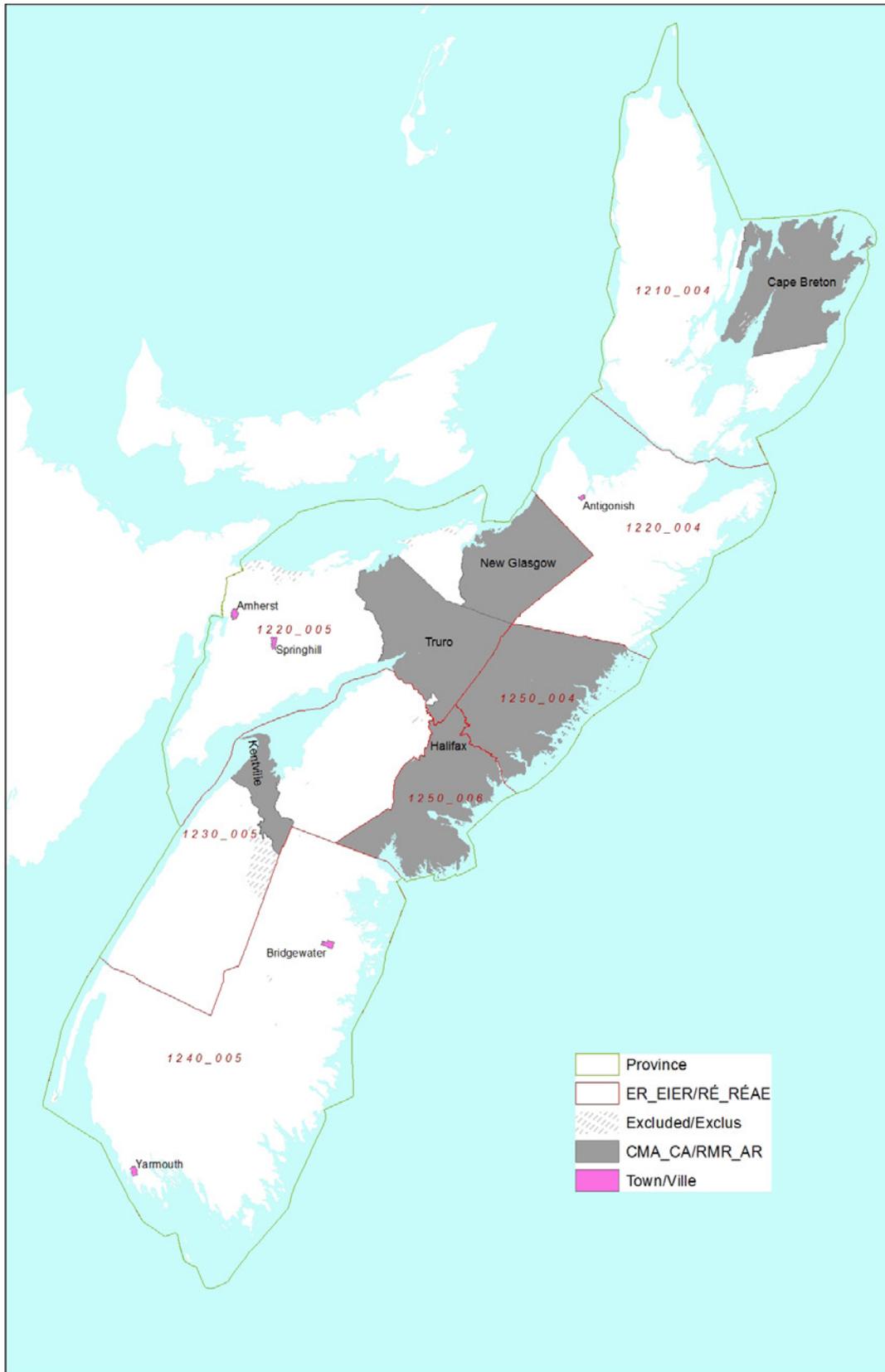
Carte 1 - Terre-Neuve-et-Labrador
RMR/AR, limites des intersections AE-REAE



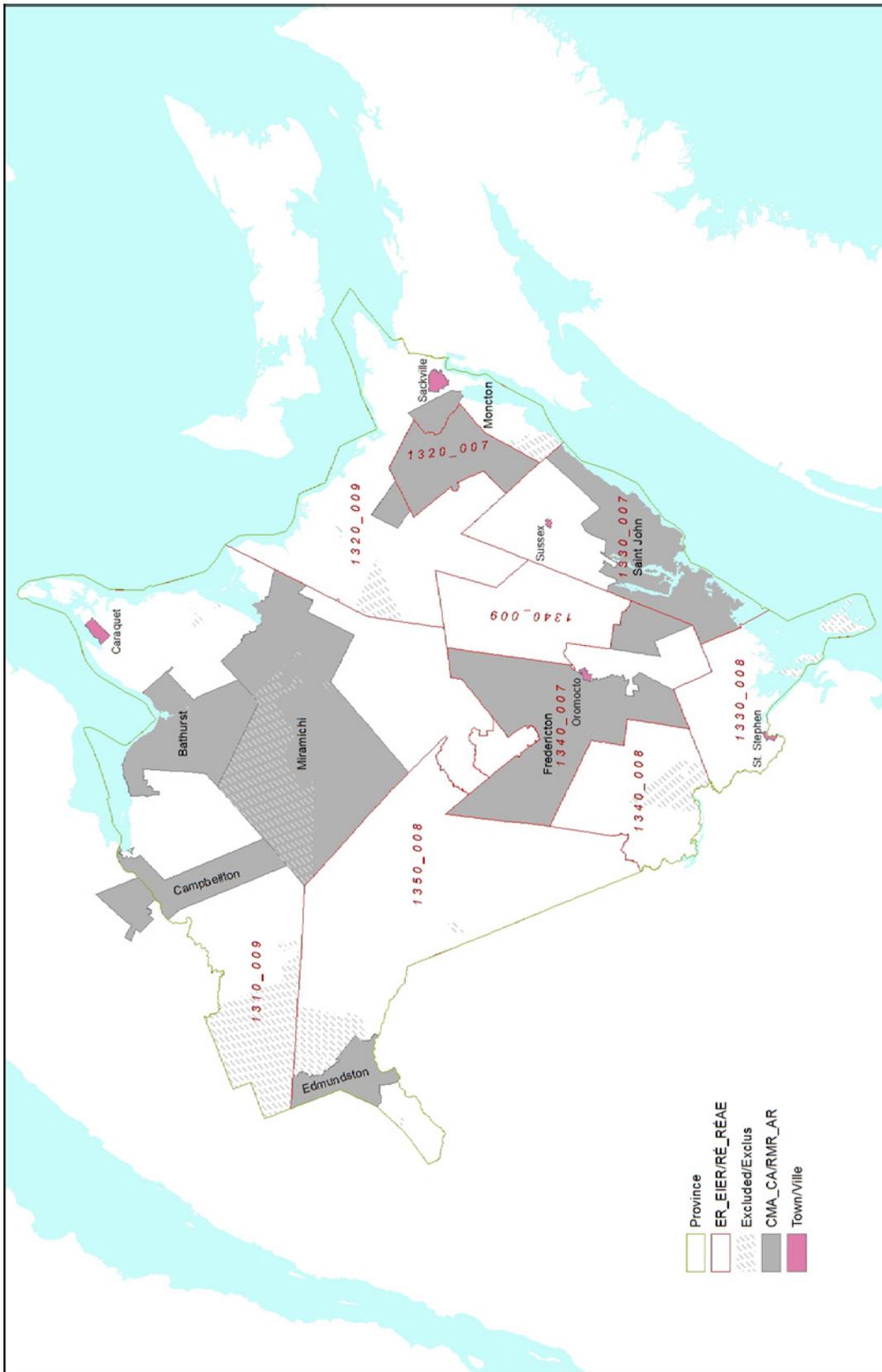
Carte 2 - Île-du-Prince-Édouard
RMR/AR, limites des intersections AE-REAE



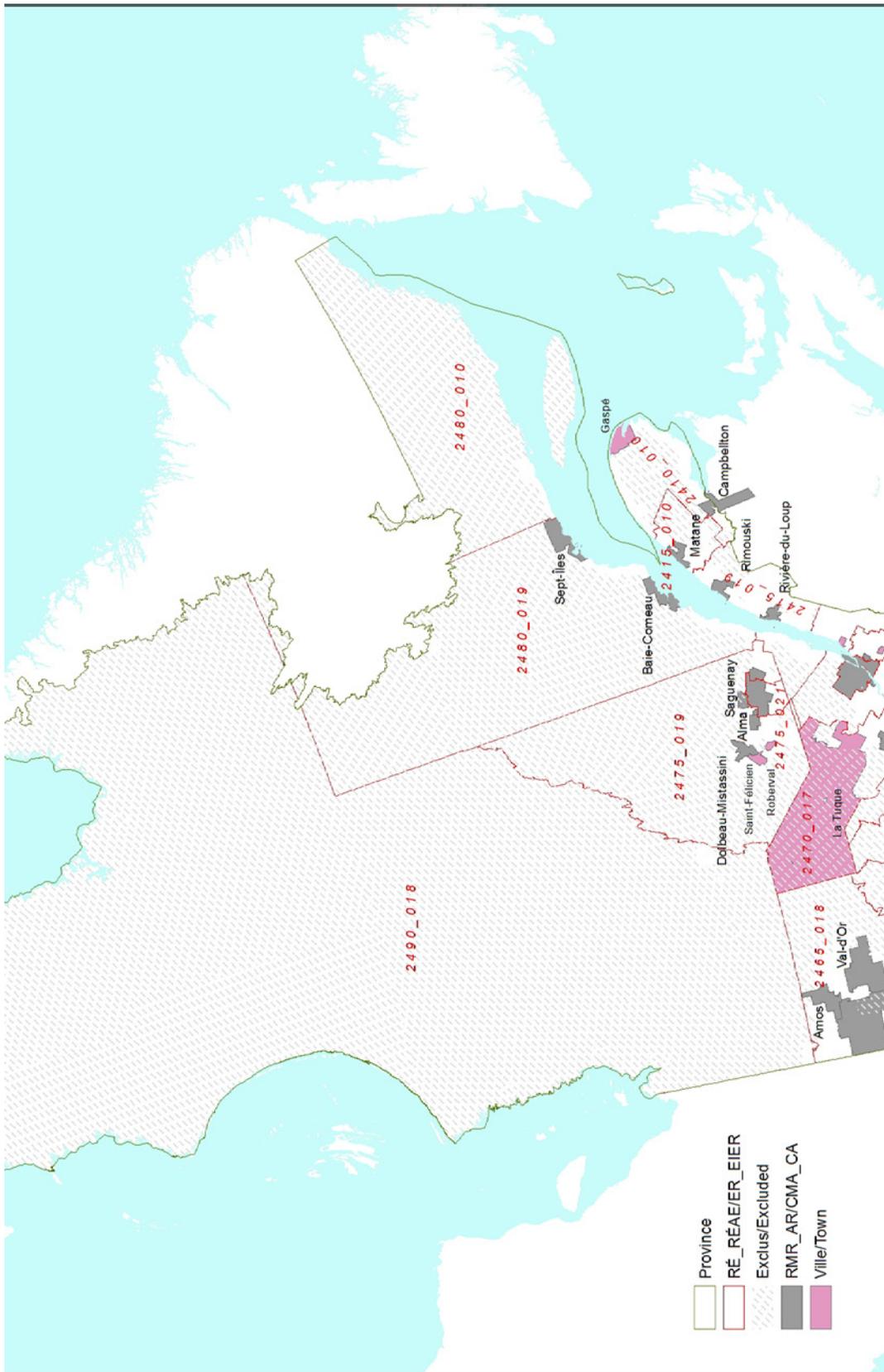
Carte 3 – Nouvelle-Écosse
RMR/AR, limites des intersections AE-REAE



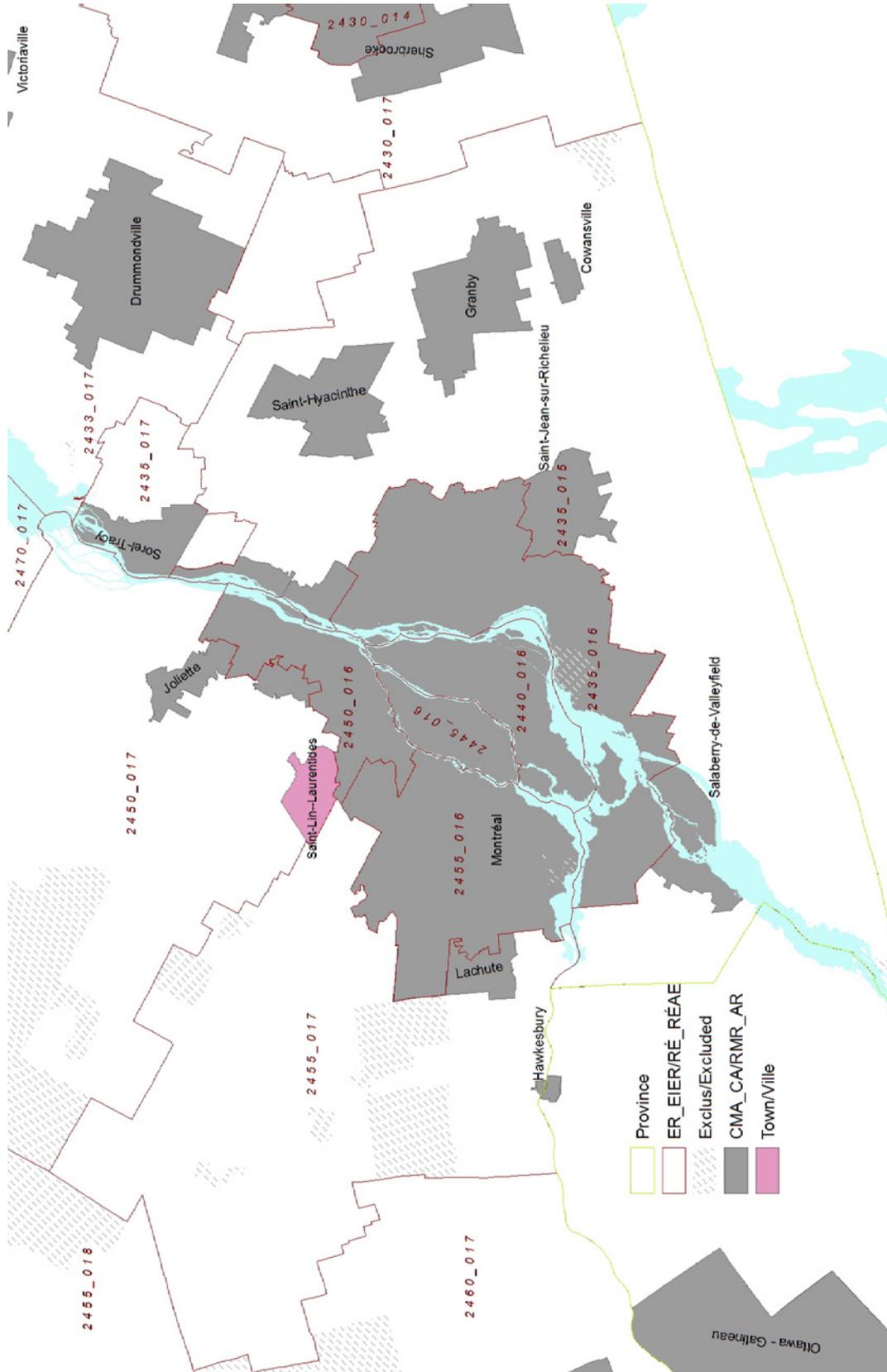
Carte 4 – Nouveau-Brunswick
RMR/AR, limites des intersections AE-REAE



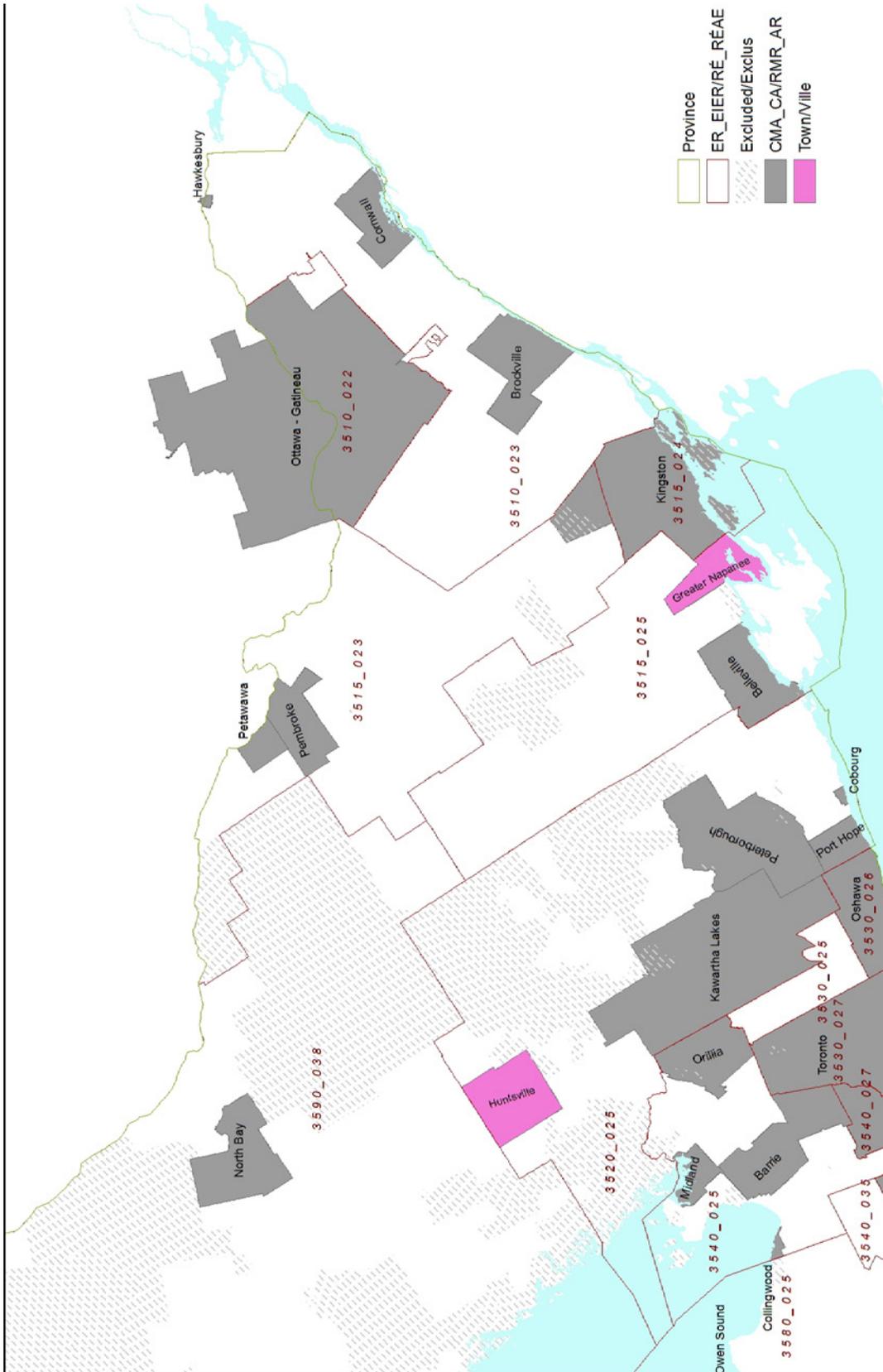
Carte 5a – Nord du Québec
RMR/AR, limites des intersections AE-REAE



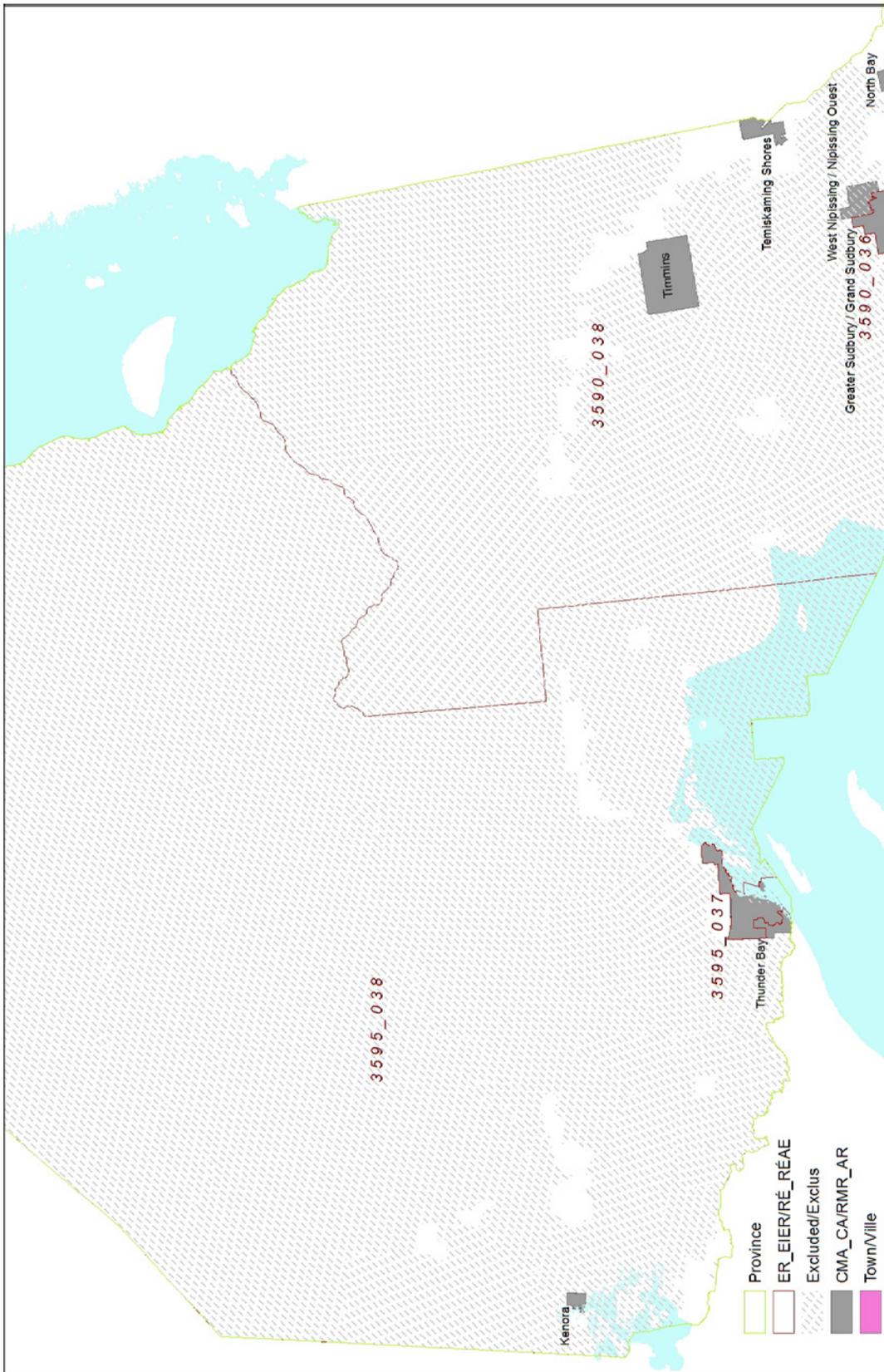
Carte 5c – Montréal
RMR/AR, limites des intersections AE-REAE



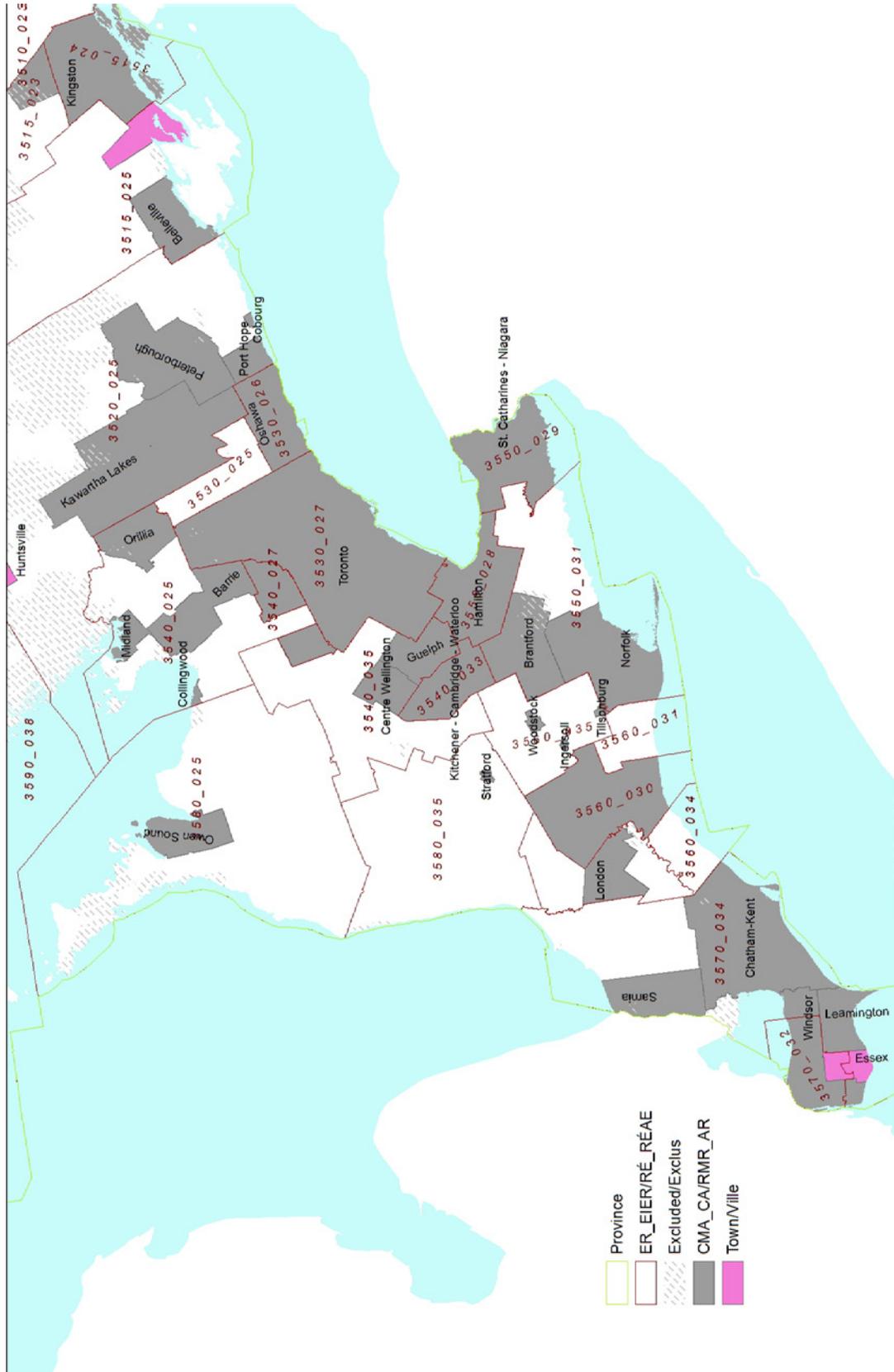
Carte 6a – Est de l'Ontario
RMR/AR, limites des intersections AE-REAE



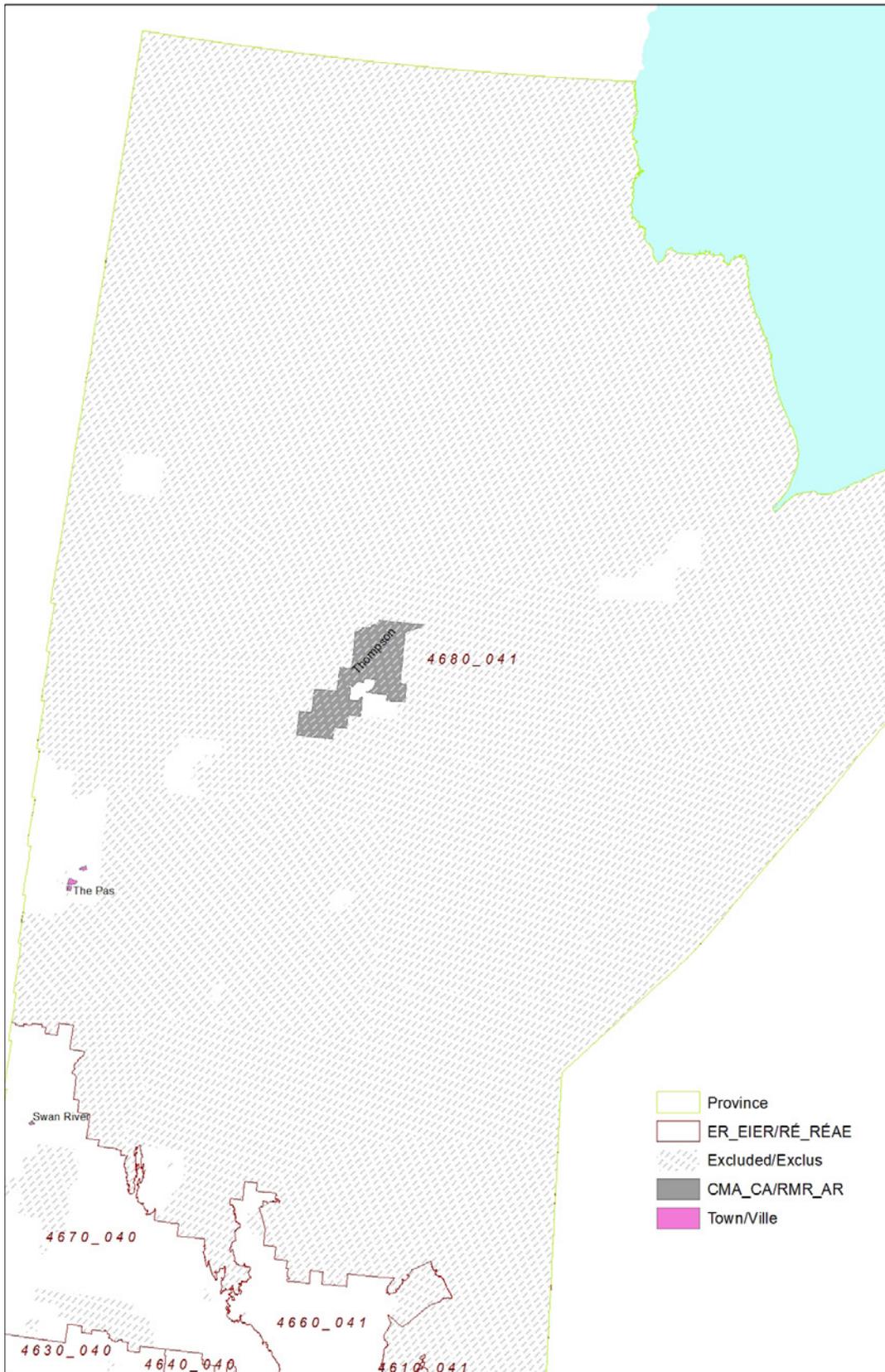
Carte 6b – Nord de l'Ontario
RMR/AR, limites des intersections AE-REAE



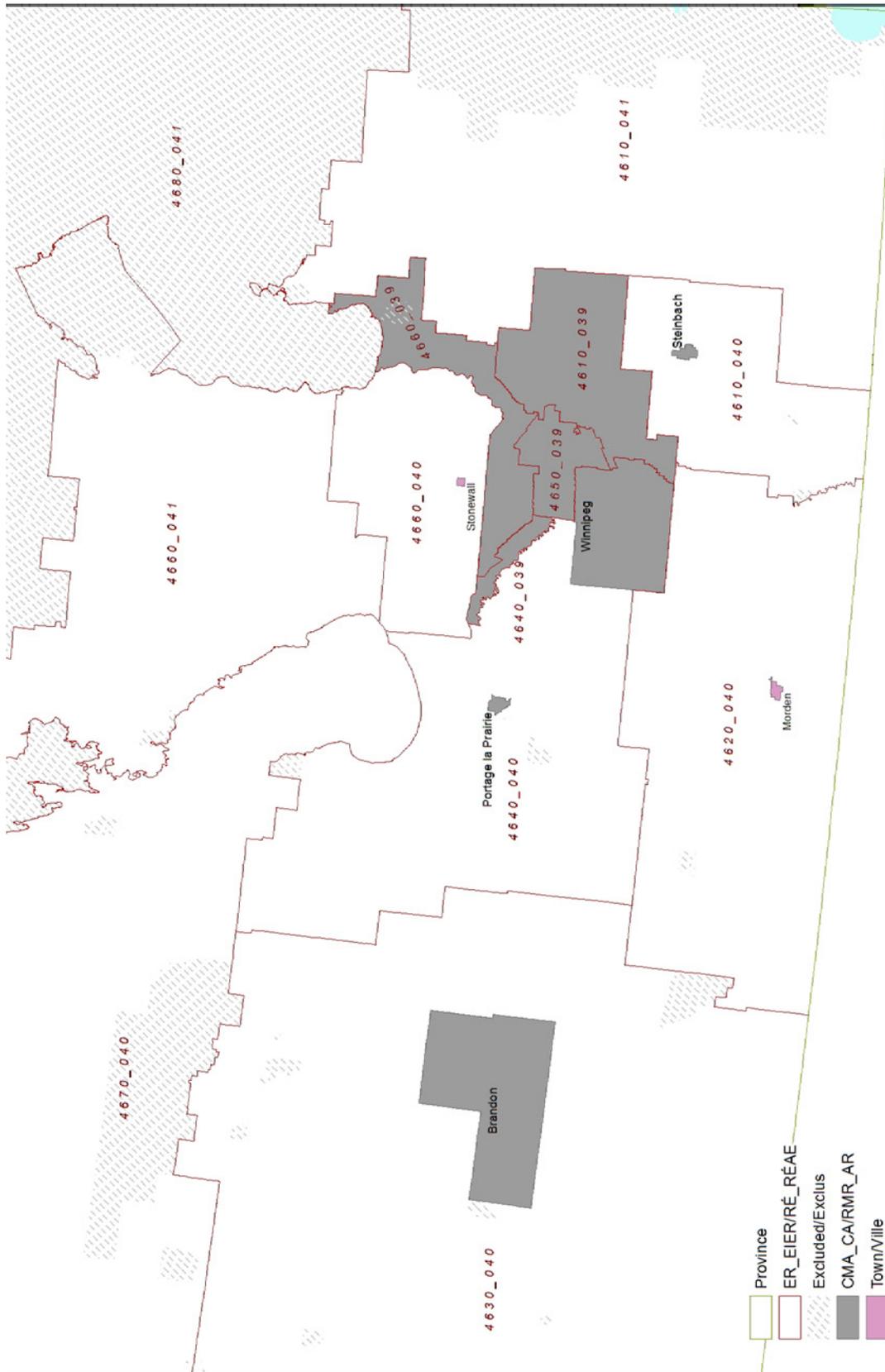
Carte 6c – Sud de l'Ontario
RMR/AR, limites des intersections AE-REAE



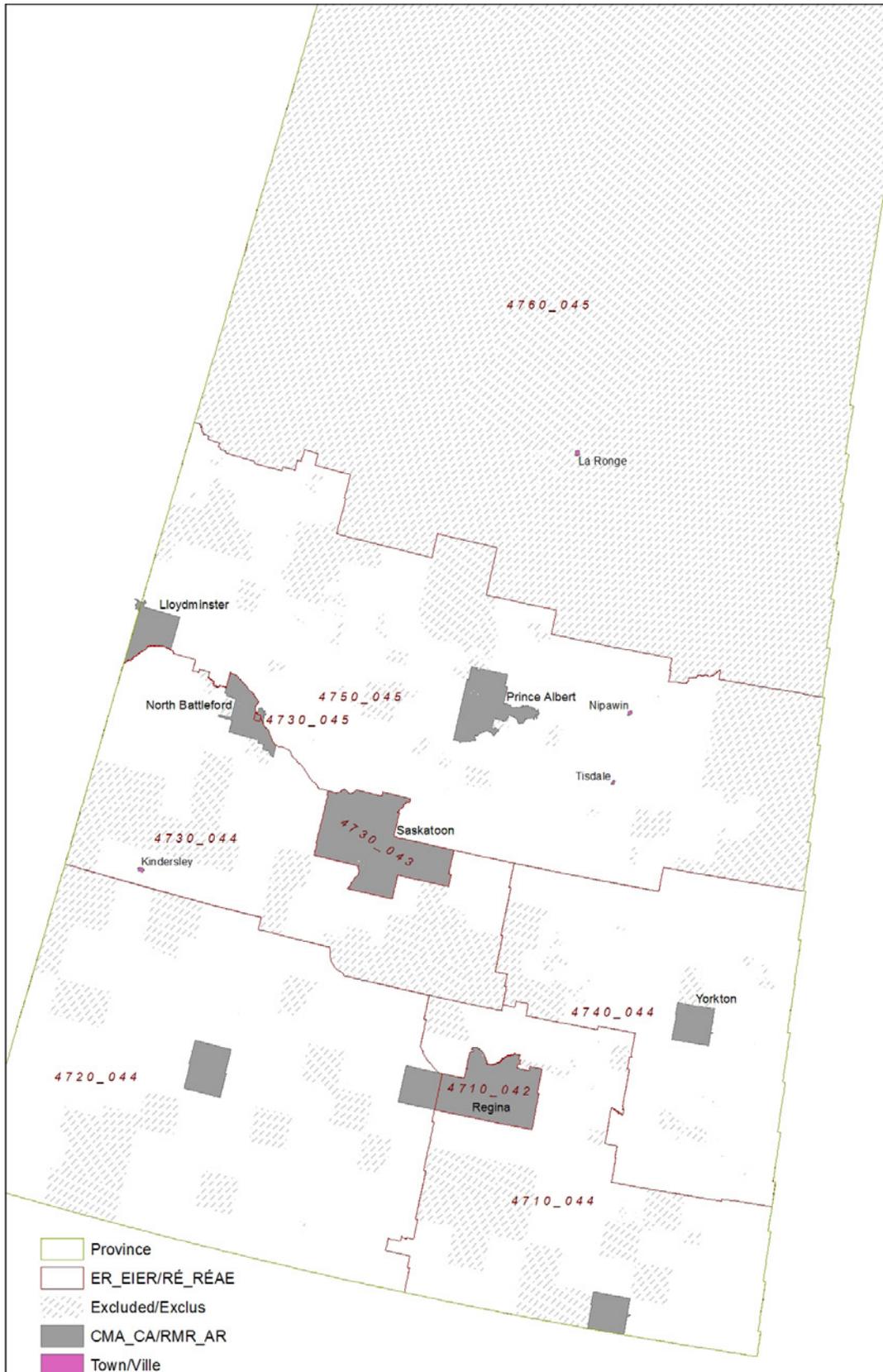
Carte 7a – Nord du Manitoba
RMR/AR, limites des intersections AE-REAE



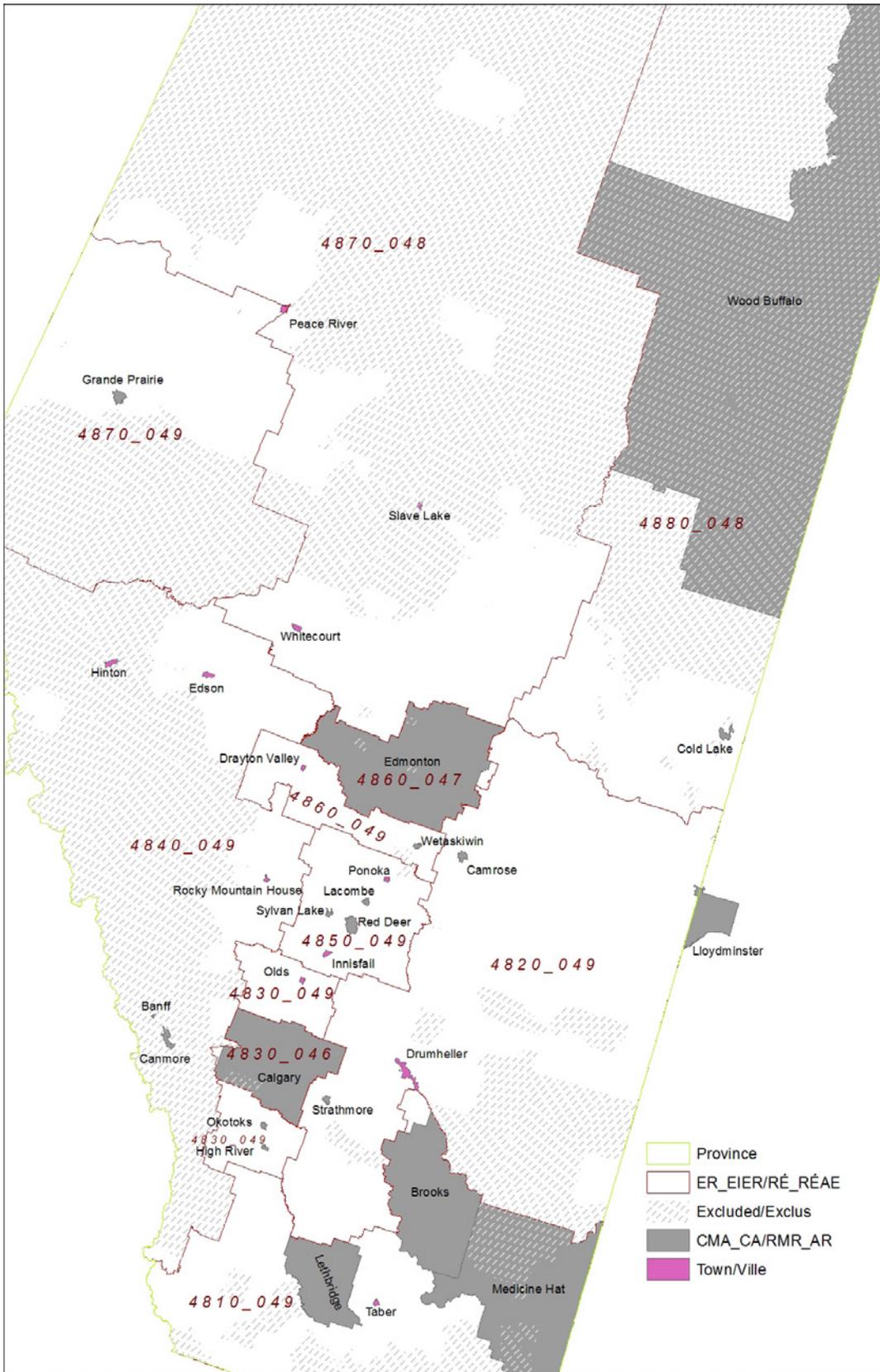
Carte 7b – Sud du Manitoba
RMR/AR, limites des intersections AE-REAE



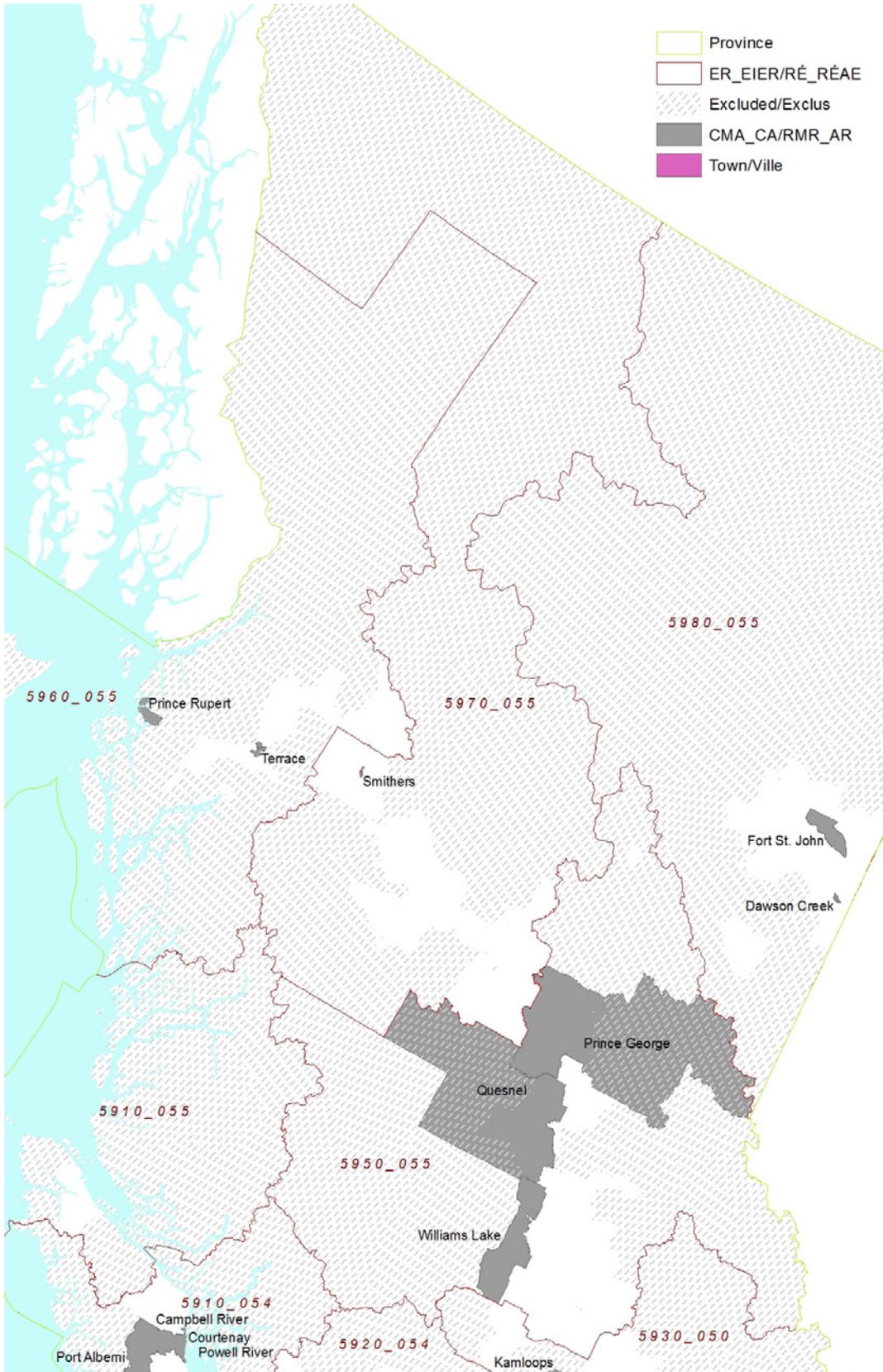
Carte 8 – Saskatchewan
RMR/AR, limites des intersections AE-REAE



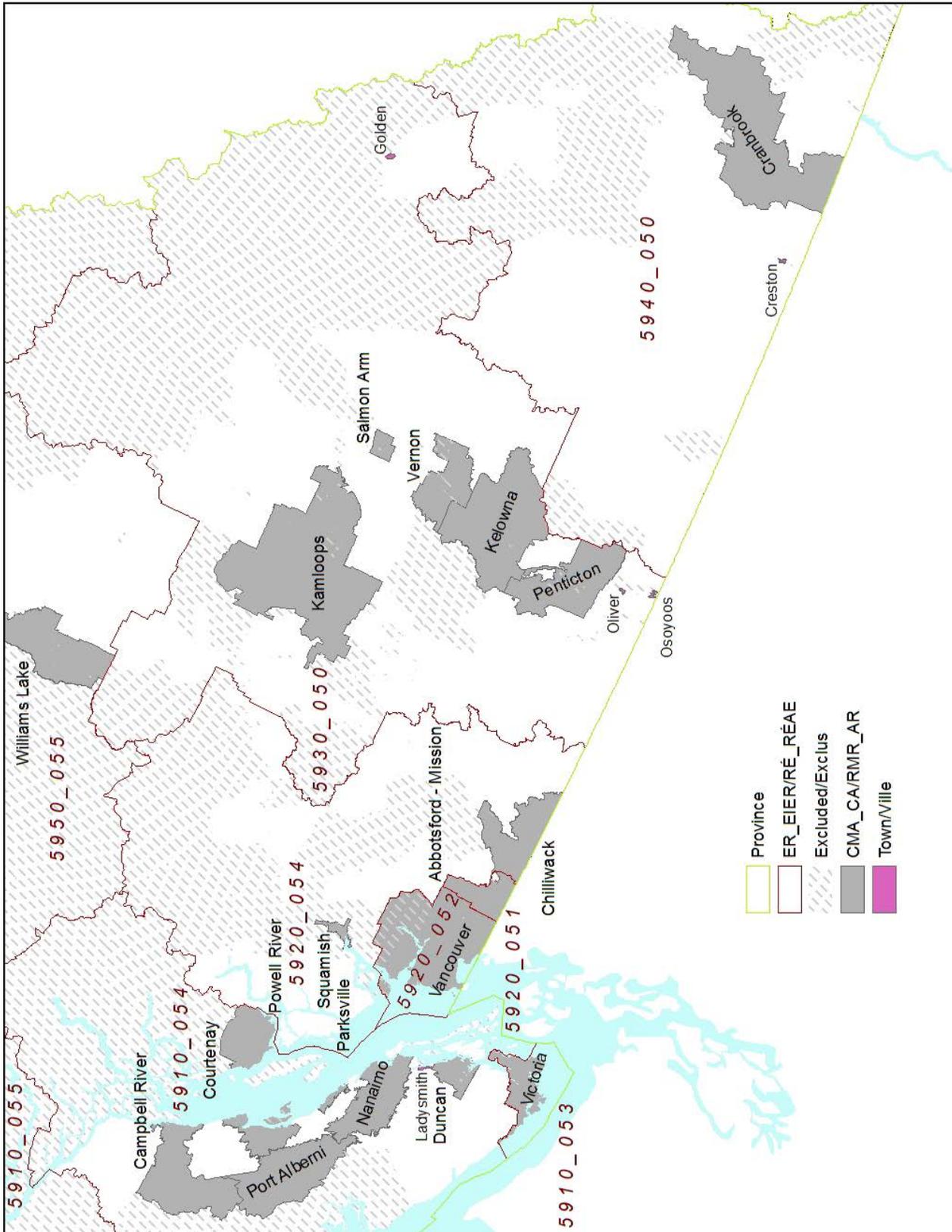
Carte 9 – Alberta
RMR/AR, limites des intersections AE-REAE



**Carte 10a – Nord de la Colombie-Britannique
RMR/AR, limites des intersections AE-REAE**



Carte 10b – Sud de la Colombie-Britannique
RMR/AR, limites des intersections AE-REAE



Annexe F Définition des variables utilisées pour former les classes d'imputation

Groupes d'âge

AGEGP1	AGEGP2	AGEGP3	Groupe d'âge
1	1	1	15 à 19 ans
2	2	2	20 à 24 ans
3	3	2	25 à 29 ans
3	4	2	30 à 34 ans
3	5	2	35 à 44 ans
3	6	2	45 à 54 ans
4	7	2	55 à 64 ans
5	8	3	65 ans et plus

Groupes de professions

OCC4	OCC10	Description
01	01	Gestion, commerce, finance et sciences naturelles et appliquées
02	02	Personnel administratif et de bureau
02	03	Santé
02	04	Enseignement, droit et services sociaux, communautaires et gouvernementaux
03	05	Arts, culture, loisirs et sports
03	06	Ventes et services
03	07	Métiers, transport et opérateurs d'équipement
03	08	Ressources naturelles et agriculture
03	09	Fabrication et services d'utilité publique
04	10	N'a jamais travaillé ou dernier emploi occupé il y a plus d'un an ou incapacité permanente

Groupes d'études

EDUC	Description
0	La personne n'a pas de diplôme d'études secondaires
1	La personne a un diplôme d'études secondaires

Catégorie de travailleur

COW	Description
1	Salarié
2	Travailleur autonome
3	Travailleur familial non rémunéré

Situation à l'égard des études

STUD	Description
0	Pas un étudiant à temps plein
1	Étudiant à temps plein

Logement possédé ou loué

DWELRENT	Description
1	Logement possédé
2	Logement loué

Province de résidence

PROV	Province
10	Terre-Neuve-et-Labrador
11	Île-du-Prince-Édouard
12	Nouvelle-Écosse
13	Nouveau-Brunswick
24	Québec
35	Ontario
46	Manitoba
47	Saskatchewan
48	Alberta
59	Colombie-Britannique

Sexe

SEX	Description
M	Homme
F	Femme

Options de parcours pour la situation vis-à-vis de l'activité (parcours temporaire)

TPATH	LFSSTAT	Description
1	1	Personne occupée au travail
2	2	Personne occupée absente du travail
3	3	Personne mise à pied temporairement
4	4	Chômeur à la recherche d'un emploi
5	5	Chômeur devant commencer un emploi à une date future
6	6	Personne hors de la population active
7	7	Personne ayant une incapacité permanente de travailler
8	2, 3, 4, 5 ou 6	
9	3, 4, 5 ou 6	
10	2, 4, 5 ou 6	
11	4, 5 ou 6	
12	5 ou 6	

Le répondant avait-il plus d'un emploi ou plus d'une entreprise la semaine dernière?

MULTJOB	Réponse
1	Oui
2	Non ou sans réponse

Pays de naissance Country of birth

IMM	Description
1	Canada
2	États-Unis
3	Autre

Identité autochtone

ABQ1	Indien de l'Amérique du Nord, Métis ou Inuit
1	Oui
2	Non

Groupes relatifs à la situation vis-à-vis de l'activité le mois dernier

LMLFS3	LMLFS7	Description
1	1	Personne occupée au travail
1	2	Personne occupée absente du travail
2	3	Chômeur mis à pied temporairement
2	4	Chômeur à la recherche d'un emploi
2	5	Chômeur devant commencer un emploi à une date future
3	6	Personne hors de la population active
3	7	Incapacité permanente de travailler

Groupe industriel du mois dernier

LMINDG	Description
1	Agriculture, foresterie, pêche et chasse
2	Exploitation minière et extraction de pétrole et de gaz
3	Services publics
4	Construction
5	Fabrication
6	Industrie du commerce de gros
7	Commerce de détail
8	Transport et entreposage
9	Industrie de l'information et industrie culturelle
10	Finance et assurances
11	Services immobiliers et services de location et de location à bail
12	Services professionnels, scientifiques et techniques
13	Gestion de sociétés et d'entreprises
14	Services administratifs, services de soutien, services de gestion des déchets et services d'assainissement
15	Services d'enseignement
16	Soins de santé et assistance sociale
17	Arts, spectacles et loisirs
18	Hébergement et services de restauration
19	Autres services (sauf les administrations publiques)
20	Administrations publiques

Annexe G Variables auxiliaires composites

Vous trouverez ci-après une liste des 28 variables auxiliaires composites qui sont utilisées pour le calage composite de l'EPA. Ces variables sont définies au niveau provincial. Un astérisque (*) indique qu'il n'est pas nécessaire de préciser la variable auxiliaire, car on peut la déduire à partir d'autres variables auxiliaires.

Caractéristiques de la population active du mois précédent (non ventilées)

Personnes occupées, 15+

Chômeurs, 15+

* *Personnes inactives, 15+*

Caractéristiques de la population active du mois précédent par groupes d'âge-sexe

Hommes occupés, 15 à 24 ans

Hommes chômeurs, 15 à 24 ans

Hommes inactifs, 15 à 24 ans

Homme occupés, 25+

Hommes chômeurs, 25+

Hommes inactifs, 25+

Femmes occupées, 15 à 24 ans

Chômeuses, 15 à 24 ans

Femmes inactives, 15 à 24 ans

* *Femmes occupées, 25+*

* *Chômeuses, 25+*

* *Femmes inactives, 25+*

Emploi pour le mois précédent par secteur d'activité

Personnes occupées – ressources naturelles, 15+

Personnes occupées – services publics, 15+

Personnes occupées – construction, 15+

Personnes occupées – secteur manufacturier, 15+

Personnes occupées – commerce, 15+

Personnes occupées – transport et entreposage, 15+

Personnes occupées – finance, assurance et immobilier, 15+

Personnes occupées – services professionnels, scientifiques et techniques, 15+

Personnes occupées – gestion d'entreprises, services administratifs et autres services de soutien, 15+

Personnes occupées – secteur d'enseignement, 15+

Personnes occupées – soins de santé et assistance sociale, 15+

Personnes occupées – information, culture et loisirs, 15+

Personnes occupées – hébergement et services de restauration, 15+

Personnes occupées – autres services, 15+

Personnes occupées – administrations publiques, 15+

* *Personnes occupées – agriculture, 15+*

Emploi pour le mois précédent par catégorie de travailleur

Personnes occupées, employés du secteur public, 15+

Personnes occupées, employés du secteur privé, 15+

* *Personnes occupées, secteur privé, travailleurs autonomes, 15+*