

Enquête sur la santé dans les collectivités canadiennes (ESCC)

Composante annuelle

Guide de l'utilisateur

Fichiers de microdonnées de 2019-2020

Avril 2022



Statistique
Canada

Statistics
Canada

Canada

TABLE DES MATIÈRES

1. INTRODUCTION	6
2. HISTORIQUE.....	7
3. REMANIEMENT DE L'ESCC EN 2015	8
4. STRUCTURE DU CONTENU DE L'ESCC	9
4.1 Contenu de base	9
4.2 Contenu thématique	9
4.3 Contenu optionnel.....	9
4.4 Contenu de réponse rapide.....	10
4.5 Contenu des fichiers de microdonnées	10
4.6 Nouveau module sur la COVID-19 (COV)	11
5. PLAN D'ÉCHANTILLONNAGE.....	12
5.1 Population cible	12
5.2 Régions sociosanitaires.....	12
5.3 Taille et répartition de l'échantillon.....	13
5.4 Bases de sondage et stratégies d'échantillonnage des ménages.....	14
5.4.1 Échantillonnage des ménages à partir de la base aréolaire pour la population adulte	14
5.4.2 Échantillonnage des personnes à partir de la base liste pour la population des jeunes	16
5.5 Répartition de l'échantillon par période de collecte des données	17
5.6 Échantillonnage des personnes interviewées	17
5.7 Achat d'unités d'échantillonnage supplémentaires pour l'Ontario.....	18
5.8 Suréchantillon de l'ESCC aux fins de l'Enquête canadienne sur la santé des aînés	18
5.8.1 Base de sondage utilisée pour l'ESCA.....	19
5.8.2 Détermination de la taille de l'échantillon pour l'ESCA.....	19
5.8.3 Sélection des personnes interviewées dans le cadre de l'ESCA.....	19
6. COLLECTE DES DONNÉES	20
6.1 Interviews assistées par ordinateur	20
6.2 Développement des applications de l'ESCC	21
6.3 Formation des intervieweurs.....	22
6.4 L'interview	22
6.5 Opérations sur le terrain.....	25
6.6 Contrôle de la qualité et gestion de la collecte.....	25
7. TRAITEMENT DES DONNÉES	26
7.1 Vérification	26
7.2 Flux.....	26
7.3 Codage	27
7.4 Création de variables dérivées	28
7.5 Pondération.....	28
7.6 Variables du revenu pour 2019-2020 - couplage et imputation	28
7.7 Imputation du code postal.....	29
8. PONDÉRATION	29
8.1 Aperçu	30
8.2 Pondération de l'échantillon de la base aréolaire.....	33
8.3 Pondération de l'échantillon de la base de l'ECSCA	35
8.4 Ajustement des poids en fonction de la combinaison de la base aréolaire et de la base de l'ECSCA	36
8.5 Pondération de l'échantillon de la base de l'ACE	38
8.6 Ajustement des poids finals en fonction de la combinaison de la base aréolaire et de la base de l'ACE	40
8.7 Création d'un poids partagé.....	41
8.8 Particularités de la pondération pour les trois territoires	42
8.9 Modification de la pondération dans les trois territoires en raison de la COVID-19	42
8.10 Pondération d'un fichier de deux ans.....	44
9. QUALITÉ DES DONNÉES.....	45
9.1 Taux de réponse.....	45

9.2	Erreurs dans les enquêtes.....	45
9.2.1	Erreurs non dues à l'échantillonnage.....	45
9.2.2	Erreurs dues à l'échantillonnage.....	46
10.	LIGNES DIRECTRICES POUR LA TOTALISATION, L'ANALYSE ET LA DIFFUSION ...	47
10.1	Lignes directrices pour l'arrondissement.....	47
10.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation	48
10.2.1	Définitions des catégories d'estimations : de type nominal par opposition à quantitatives	48
10.2.2	Totalisation d'estimations de type nominal.....	49
10.2.3	Totalisation d'estimations quantitatives.....	49
10.3	Lignes directrices pour l'analyse statistique	50
10.4	Lignes directrices pour la diffusion	51
11.	TABLEAUX DE LA VARIABILITÉ D'ÉCHANTILLONNAGE APPROXIMATIVE	52
11.1	Comment utiliser les tableaux de CV pour les estimations de type nominal	53
11.2	Exemples d'utilisation des tableaux de CV pour des estimations de type nominal	55
11.3	Comment utiliser les tableaux de CV pour calculer les limites de confiance	58
11.4	Exemple d'utilisation de tableaux de CV pour obtenir des limites de confiance.....	59
11.5	Comment utiliser les tableaux de CV pour effectuer un test Z	59
11.6	Exemple d'utilisation des tableaux de CV pour effectuer un test Z.....	59
11.7	Variances ou coefficients de variation exacts	60
12.	FICHIERS DE MICRODONNÉES : DESCRIPTION, ACCÈS ET UTILISATION.....	61
12.1	Fichiers maîtres.....	61
12.1.1	Centre de données de recherche.....	61
12.1.2	Totalisations personnalisées.....	61
12.1.3	Télé-accès.....	62
12.1.4	Système d'accès à distance en temps réel.....	62
12.2	Fichiers de partage.....	62
12.3	Fichiers de microdonnées à grande diffusion	62
12.4	Utilisation des fichiers de l'ESCC : fichier annuel ou fichier deux ans ?.....	66
12.5	Utilisation de la variable de pondération	66
12.6	Convention appliquée pour nommer les variables	67
12.7	Dictionnaires de données	69
12.8	Différences dans le calcul des variables du contenu de base et du contenu thématique utilisant différents fichiers.....	69
ANNEXE A - L'ENQUETE SUR LA SANTE DANS LES COLLECTIVITES CANADIENNES (ESCC) - APERÇU DU CONTENU - 2020.....		70
ANNEXE B - L'ENQUETE SUR LA SANTE DANS LES COLLECTIVITES CANADIENNES (ESCC) - PROBLEMES DE SANTE CHRONIQUES - 2020.....		74
ANNEXE C - GÉOGRAPHIE DISPONIBLE SUR LE FICHIER MAÎTRE ET DE PARTAGE ET CODES CORRESPONDANTS: CANADA, PROVINCES, RÉGIONS SOCIOSANITAIRES ET GROUPES HOMOLOGUES		76
ANNEXE D - REPARTITION DE L'ECHANTILLON PAR PROVINCE/TERRITOIRE (ESCC 2020 ET ECSA 2020).....		83
ANNEXE E - TAUX DE REPONSE PAR PROVINCE/TERRITOIRE (ESCC 2020 ET ECSA 2020)		85

QUOI DE NEUF DANS L'ENQUÊTE SUR LA SANTÉ DANS LES COLLECTIVITÉS CANADIENNES (ESCC) 2020?

La pandémie de la COVID-19 a eu des répercussions majeures sur la collecte de données de l'ESCC 2020. La collecte s'est arrêtée à la mi-mars, soit vers la fin de la première période de collecte, et n'a repris qu'en septembre. Les deuxième, troisième et quatrième échantillons trimestriels couvraient des périodes de collecte très courtes, soit d'environ cinq semaines chacune, de septembre à décembre. L'impossibilité de mener des entrevues en personne, les périodes de collecte plus courtes et les problèmes de capacité de collecte ont entraîné une baisse significative des taux de réponse.

Comme pour les cycles de l'ESCC précédents, les poids de sondage ont été ajustés pour minimiser tout biais dû à la non-réponse : des rajustements pour la non-réponse et un calage utilisant les informations auxiliaires disponibles ont été appliqués, et sont reflétés dans les poids de sondage fournis avec le fichier de données. Des validations approfondies des estimations tirées de l'enquête ont également été effectuées et examinées selon une perspective d'analyse du biais. Malgré ces ajustements et validations rigoureux, le taux élevé de non-réponse augmente le risque de biais résiduel et l'ampleur avec laquelle un tel biais pourrait affecter les estimations produites à l'aide des données de l'enquête. Par conséquent, il est conseillé aux utilisateurs d'utiliser les données de l'ESCC 2020 avec prudence, particulièrement lors d'estimations pour de petites sous-populations ou lors de comparaisons avec les autres années de l'ESCC.

Contenu

Le texte qui suit fait part de certains des changements apportés au questionnaire de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) – composante annuelle en 2020.

- **Expériences vécues dans l'enfance (CEX), Consultations au sujet de la santé mentale (CMH), Couverture d'assurance de santé (INS), Expériences comme patient (PEX), Provisions sociales (SPS), Sources de stress (STS) et Contacts avec des professionnels de la santé dentaire (CP3)** ont été offerts comme modules optionnels dans les provinces (ils étaient thématiques en 2019) et ont été gardés comme modules thématiques biennaux (2019-2020) dans les territoires.
- **Utilisation du Guide alimentaire canadien (FGU), Activités physiques - adultes 18 ans et plus (PAA), Activités physiques pour les jeunes (PAY) et Comportements sexuels (SXB)** ont été demandés comme modules thématiques dans les provinces seulement (ils étaient optionnels en 2019).
- **Consommation de fruits et légumes (FVC)** a été demandé comme module thématique dans les provinces (il était optionnel en 2019) et a été gardé comme optionnel dans les territoires.
- **Perte de productivité (LOP)** a été ajouté en 2020 comme module thématique dans les provinces seulement.
- **Obstacles à l'activité physique (BPA) et Médicaments antidouleur (PRM)** ont été ajoutés comme modules optionnels.

- Un nouveau module sur la **COVID-19 (COV)** a été ajouté comme module thématique pour la période de collecte de septembre à décembre 2020 dans les provinces seulement. Voir la section 4.6 pour une note importante sur l'utilisation des données.
- Le module **Sécurité alimentaire (FSC)** faisait partie de la composante de contenu optionnel pour 2019-2020, toutes les provinces et tous les territoires l'ayant sélectionné à l'exception de la Colombie-Britannique. Afin de soutenir l'analyse des impacts de la pandémie sur l'insécurité alimentaire, ce module a été ajouté au plan de collecte de septembre à décembre 2020. Voir la section 4.3 pour une note importante sur l'utilisation des données.

Pour obtenir de plus amples renseignements sur le contenu de l'ESCC, veuillez consulter l'annexe A pour accéder à du contenu optionnel selon la province ou le territoire.

Collecte

En raison de la pandémie de la COVID-19, la période de collecte 2020 s'est arrêtée en mars. La collecte a repris en septembre, jusqu'en décembre, avec des protocoles de sécurité stricte et des pratiques exemplaires. Tous les intervieweurs, y compris ceux qui travaillaient dans les centres de collecte centralisés, ont travaillé à domicile. Les entrevues en personne ont également été suspendues, et uniquement menées par téléphone.

Plan d'échantillonnage

En 2020, puisque les entrevues en personne n'étaient pas permises pour une bonne partie de l'année dans le contexte de la pandémie de COVID-19, la couverture dans les territoires a exceptionnellement été limitée aux trois capitales (Whitehorse, Yellowknife et Iqaluit) lors des trois dernières périodes de collecte.

Documentation

- Dans le but de refléter les changements apportés par le gouvernement fédéral en juillet 2016 au programme des Prestations fiscales canadiennes pour enfant (PFCE), nous référons maintenant au programme actuel Allocation canadienne pour enfant (ACE) dans la documentation.
- Errata : Les erreurs relevées dans les fichiers et produits de données de l'ESCC après la diffusion sont communiquées aux utilisateurs dans un document des erreurs cumulatives intitulé « ESCC Errata ». L'errata fournit des renseignements sur :
 - les produits touchés (p. ex. le fichier maître, les fichiers partagés ou le FMGD);
 - les années touchées;
 - les corrections suggérées (si possible);
 - les étapes à suivre pour effectuer la correction.

Depuis la diffusion de l'ESCC de 2019, il y a eu 0 ajout aux documents des erreurs.

Pour votre commodité, le document des erreurs renferme un index à la première page. Celui-ci fournit une liste des modules et des années touchés. Pour vérifier si une variable particulière est affectée, suivre les hyperliens menant aux numéros des erreurs connexes.

Pour obtenir une version à jour de l'ESCC errata, veuillez communiquer avec le Centre de données sur la santé de la population à 613-951-1746 ou à hd-ds@statcan.gc.ca.

1. INTRODUCTION

L'Enquête sur la santé dans les collectivités canadiennes (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'ESCC a fait l'objet d'un important remaniement ayant commencé en 2012 et ayant été mis en œuvre au niveau de la collecte en 2015. Des détails sur ce remaniement sont fournis à l'intérieur du chapitre 3.

Les objectifs de l'enquête ont été révisés comme suit :

- soutenir les programmes de surveillance en santé en produisant des données sur la santé à l'échelle nationale, provinciale et à l'échelle des régions sociosanitaires;
- offrir une source unique de renseignements pour la recherche sur la santé des petites populations et sur des caractéristiques rares;
- diffuser de l'information facilement accessible à une communauté diversifiée d'utilisateurs dans un temps opportun; et
- proposer un instrument d'enquête flexible qui inclut une option de réponse rapide pour répondre à des questions émergentes liées à la santé de la population.

Les données de l'ESCC sont recueillies auprès de personnes âgées de 12 ans et plus demeurant dans des logements privés dans plus de 100 régions sociosanitaires couvrant toutes les provinces et les territoires. Sont exclues de la base de sondage les personnes vivant sur les réserves indiennes et les terres de la Couronne, les résidents d'institutions (institutions de santé, prisons, institutions religieuses, couvents, etc.), les membres à temps plein des Forces canadiennes, les jeunes âgés de 12 à 17 ans demeurant dans des foyers d'accueil, et les personnes vivant dans certaines régions éloignées. L'ESCC couvre environ 98 % de la population canadienne âgée de 12 ans et plus.

Le présent document a pour but de faciliter la manipulation des fichiers de microdonnées de l'ESCC et de décrire la méthodologie utilisée. L'enquête produit 3 types de fichiers de microdonnées : des fichiers maîtres, des fichiers de partage et des fichiers de microdonnées à grande diffusion (FMGD). Les caractéristiques de chacun de ces fichiers sont présentées dans ce guide. Diffusé au bout de deux ans de collecte des données, le FMGD contient deux années de données. Le prochain FMGD (dont la date de diffusion reste à déterminer) comprendra les données recueillies au cours des années 2019 et 2020.

Pour toute question concernant les ensembles de données ou leur utilisation, adressez-vous à :

Service d'aide aux utilisateurs des produits électroniques:	1-800-949-9491
Totalisations spéciales ou renseignements généraux sur les données:	
Services personnalisés à la clientèle, Centre de données sur la santé de la population:	613-951-1746
Courriel:	hd-ds@statcan.gc.ca

Renseignements sur le télé-accès:	613-951-1746
Courriel:	cchssm-esccds@statcan.gc.ca
Télécopieur:	613-951-0792

2. HISTORIQUE

En 1991, le Groupe de travail national sur l'information en matière de santé a relevé plusieurs questions et problèmes posés par le système d'information sur la santé. Selon ses membres, les données étaient fragmentées, elles étaient incomplètes, elles ne pouvaient être partagées facilement et elles n'étaient pas analysées aussi pleinement que possible; en outre, les résultats des études réalisées n'atteignaient pas de façon régulière la population canadienne.¹

En 1999, pour résoudre ces problèmes, l'Institut canadien d'information sur la santé (ICIS), Statistique Canada et Santé Canada ont conjugué leurs efforts en vue de créer un Carnet de route de l'information sur la santé. L'Enquête sur la santé dans les collectivités canadiennes (ESCC) a été conçue à partir de ce mandat. Le format, le contenu et les objectifs de l'enquête ont été définis après avoir mené des consultations approfondies auprès de spécialistes et d'intervenants fédéraux, provinciaux et de régions sociosanitaires communautaires en vue de déterminer leurs exigences en matière de données². La collecte de données dans le cadre de l'enquête a débuté en 2000.

Afin de combler les nombreux besoins en données, le cycle de collecte des données de l'ESCC s'étendait sur deux années. Jusqu'en 2007, la première année du cycle d'enquête, désignée par « .1 », a constitué une enquête sur la santé de l'ensemble de la population qui visait à fournir des estimations fiables à l'échelon des régions sociosanitaires. La deuxième année du cycle d'enquête, désigné par « .2 », reposait sur un échantillon plus modeste et visait à fournir des résultats à l'échelle provinciale sur des sujets particuliers relatifs à la santé.

À partir de 2007, la composante régionale du programme de l'ESCC a commencé à faire l'objet d'une collecte continue. Les cycles x.1 de l'ESCC ont été renommé « composante annuelle » de l'ESCC et ont désormais fait l'objet d'une collecte chaque année. Le titre complet d'un cycle annuel donné est « Enquête sur la santé dans les collectivités canadiennes – Composante annuelle, 20XX » et le titre abrégé est simplement « ESCC – 20XX ».

La composante axée sur le contenu de l'enquête a été désignée par le nom de l'enquête suivi du sujet des thèmes visés par chaque enquête (p. ex. « Enquête sur la santé dans les collectivités canadiennes portant sur le vieillissement en santé » ou « ESCC – Vieillissement en santé »).

En 2012, un remaniement en profondeur de l'enquête a commencé, qui a modifié l'échantillonnage et la stratégie de collecte. Le contenu de l'enquête a fait l'objet d'essais qualitatifs visant à assurer la validité et la fiabilité en préparation à un nouveau plan octennal du contenu débutant en 2015.

¹ *Carnet de route de l'information sur la santé : Répondre aux besoins*, Santé Canada, Statistique Canada. 1999 p.3.

² *Carnet de route de l'information sur la santé : Entreprendre le voyage*. Institut canadien d'information sur la santé; Statistique Canada. 1999. ISBN 1-895581-70-2. p. 19.

3. REMANIEMENT DE L'ESCC EN 2015

En 2012, le remaniement a commencé par de vastes consultations menées avec des experts et des intervenants clés aux niveaux fédéral, provincial et des régions sociosanitaires dans l'ensemble du Canada. L'objectif consistait à recueillir des suggestions et propositions de modifications ainsi que des renseignements détaillés au sujet des besoins en données et des produits de nos divers partenaires.

Voici quels sont les principaux changements résultants du remaniement de l'ESCC :

- Des changements ont été apportés au contenu de la majorité des modules, mais principalement à celui des modules de base et des modules thématiques. Environ 70 % des modules qui existaient avant le remaniement ont fait l'objet d'un certain degré de changement. Les changements en question varient des raffinements mineurs ou changements majeurs, des concepts, du vocabulaire, ou catégories de réponse. Certains nouveaux modules ont été créés pour le traitement de nouveaux concepts ou pour combler des lacunes statistiques signalées par des partenaires partageant des données, comme Santé Canada, l'Agence de la santé publique du Canada et les ministères provinciaux et territoriaux responsables de la santé.
- Les règles relatives aux variables ont changé en 2015. Les règles d'affectation des noms utilisés pour les variables dans l'ESCC de 2015 utilisent un maximum de huit caractères. Comme auparavant, les trois premiers caractères correspondent au nom du module et de la section du questionnaire, tandis que le quatrième caractère indique le type de variable (soulignement, D, F, C ou G). Les éléments d'enquête à l'intérieur du questionnaire (p. ex. E, C, D, Q, N) sont désormais assujettis à des règles uniformes qui augmentent par incréments de 5 (p. ex. CCC_005, CCC_010, CCC_015).
- Toutes les variables dérivées ont été passées en revue, et les spécifications mises à jour conformément au nouveau nom des variables. Les modifications ayant trait aux concepts et au contenu ont elles aussi été passées en revue, ce qui a donné lieu à la création de nouvelles variables dérivées correspondant aux nouveaux modules de contenu. La stratégie annuelle de collecte des données a été modifiée en 2015. Avant 2015, la collecte se déroulait dans le cadre de six périodes de collecte de deux mois. À partir de 2015, la collecte pour l'enquête a été divisée en quatre périodes de trois mois qui ne se chevauchent pas.
- À partir de 2015, la collecte des données dans le Nord a été modifiée de sorte qu'environ la moitié des collectivités de chaque territoire sont visitées chaque année. En conséquence, les données ne sont représentatives des territoires qu'au bout de deux ans.

4. STRUCTURE DU CONTENU DE L'ESCC

En plus de données sociodémographiques et administratives, le contenu de l'ESCC comporte trois composantes de contenu qui répondent chacune à des besoins différents : contenu commun (contenu de base et contenu thématique), contenu optionnel et le contenu de réponse rapide. L'annexe A présente la liste des modules qui ont été inclus dans le questionnaire de 2019-2020 et leur type de composante.

La durée moyenne d'une entrevue de l'ESCC est estimée entre 40 et 45 minutes.

Tableau 4.1 Durée de l'enquête selon les composantes

COMPOSANTES DE L'ESCC		Durée
Contenu commun	Contenu de base	19 minutes
	Contenu thématique	8 minutes
Contenu optionnel		8 minutes
Contenu de réponse rapide (recouvrement des coûts)		2 minutes

4.1 Contenu de base

La composante de contenu de base de l'ESCC comprend des questions posées aux répondants de toutes les provinces et territoires (à moins d'exception). Ces questions demeureront relativement stables dans le questionnaire durant une période d'environ six ans, soit jusqu'en 2021.

4.2 Contenu thématique

Le contenu thématique, demandé à tous les participants à l'enquête, se compose de modules et, dans certains cas, de groupes de questions faisant partie de ces modules, ayant trait à un sujet précis. Des données sont recueillies chaque année pour deux types de thèmes : les modules thématiques biennaux, demandés pendant une période de deux ans, et les modules thématiques annuels, demandés pendant une période d'un an. Les thèmes annuels changent chaque année. Le contenu des deux types de thèmes (annuels et biennaux) correspond à environ huit minutes du temps de l'interview. Certains thèmes peuvent être réintroduits dans l'enquête à tous les deux, quatre ou six ans au besoin. Cette composante permet de mieux planifier le contenu de l'ESCC à moyen terme.

4.3 Contenu optionnel

La composante de contenu optionnel est conçue pour permettre aux provinces et aux territoires de sélectionner du contenu correspondant à leurs priorités en matière de santé publique. Chaque province et territoire sélectionne des modules d'une durée maximale de 8 minutes. On demande aux participants à l'enquête le contenu optionnel qui figure dans les modules sélectionnés pour la province ou le territoire où ils résident. Le contenu optionnel est sélectionné pour deux ans. Il convient de souligner que contrairement aux modules qui font partie du contenu thématique ou de base, les données recueillies dans le cadre d'un module de contenu optionnel ne doivent pas être

extrapolées à l'ensemble du Canada³.

L'annexe A présente les résultats de la sélection du contenu optionnel pour 2020 par province ou territoire de résidence.

En 2020, le module sur la Sécurité alimentaire (FSC) a été offert comme optionnel et a été sélectionné par toutes les provinces et tous les territoires, à l'exception de la Colombie-Britannique. Étant donné que ce contenu représente une préoccupation croissante dans la recherche de la santé publique, et les risques d'insécurité alimentaires attendus en raison de la pandémie de la COVID-19, ce module a été ajouté au questionnaire de la Colombie-Britannique pour la période de collecte de septembre à décembre afin de permettre l'estimation d'un taux de sécurité alimentaire national.

Note importante: Pour faciliter l'analyse des impacts de la pandémie sur l'insécurité alimentaire, les réponses des répondants de la Colombie-Britannique pour les périodes de collecte de septembre à décembre ont été incluses dans le fichier annuel 2020, mais pas dans les fichiers pour 2019-2020.

4.4 Contenu de réponse rapide

La composante de réponse rapide est offerte par recouvrement des coûts aux organisations désirant obtenir des estimations nationales sur un sujet émergent ou particulier lié à la santé de la population. La réponse rapide est formée d'un maximum de deux minutes de contenu. Les questions apparaissent au questionnaire pour une ou deux périodes de collecte (trois à six mois) et sont posées à tous les répondants de l'ESCC au cours de cette période, à l'exception de ceux des territoires.

4.5 Contenu des fichiers de microdonnées

Différents fichiers de données sont produits à partir des données de l'enquête :

- Fichier de données sous-annuel;
- Fichier de données d'un an;
- Fichier de données combinées de deux ans;
- Fichier de données de réponse rapide (le cas échéant).

Le tableau 4.2 fournit des précisions concernant la disponibilité du contenu dans les fichiers de l'ESCC 2019 et 2020.

Fichiers de données sous-annuels

En 2020, la collecte de l'ESCC a été interrompue à la fin mars en raison de la pandémie de COVID-19. La collecte n'a repris qu'en septembre. L'ESCC a produit deux fichiers sous-annuels : le premier couvre la période de collecte de janvier à mars et le deuxième, de septembre à décembre.

³ À moins qu'un module optionnel soit sélectionné par l'ensemble des provinces et territoires du Canada au cours d'une même période de collecte, ce qui n'est jamais produit à ce jour.

Fichiers de données d'un an

L'enquête produit des fichiers tous les ans. Le fichier annuel 2020 comprend les personnes interrogées dans le cadre de la collecte des données de 2020, les variables du contenu thématique et principal ainsi que du contenu optionnel. Comme l'échantillon des territoires n'est représentatif qu'au bout de deux ans de collecte, ceux-ci sont exclus du fichier de données annuel.

Fichier à période combinées de deux ans

Chaque publication d'un fichier de données portant sur une année paire, par exemple 2018, comprend un fichier combinant les données biennales. La prochaine diffusion du fichier de données combinées de deux ans est prévue en 2021 et comprendra les années de référence 2019 et 2020 de collecte.

Le fichier de deux ans inclut tous les répondants ainsi que les questions qui figuraient dans l'enquête au cours de cette période. À moins d'indication contraire, il englobe les questions du contenu de base annuel et du contenu thématique de deux ans, ainsi que le contenu optionnel sélectionné pour la période de deux ans. Le contenu thématique d'un an et le contenu optionnel choisi pour une année seulement ne sont pas disponibles dans le fichier de données de deux ans.

Tableau 4.2 Composantes de contenu pour les fichiers de données de 2019 et 2020

Files		Contenu commun annuel	Contenu thématique 1 an 2019	Module COVID-19 2020	Contenu thématique 1 an 2020	Contenu thématique 2 ans 2019-2020	Contenu optionnel ⁴
2019	Principal	Oui	Oui ⁵	S/O	S/O	Oui ⁵	Oui
2020 Janvier à mars	Sous-annuel	Oui	S/O	Non	Oui ⁵	Oui ⁵	Oui
2020 Septembre à décembre	Sous-annuel	Oui	S/O	Oui	Oui ⁵	Oui ⁵	Oui
2020	Principal	Oui	S/O	Oui ⁶	Oui ⁵	Oui	Oui
2019 - 2020	Principal	Oui	Non	Non	Non	Oui	Oui

4.6 Nouveau module sur la COVID-19 (COV)

Lorsque la collecte a repris en septembre, un nouveau module a été ajouté avec du contenu en lien avec la pandémie : changements au niveau de la santé mentale, précautions prises pour se protéger contre la COVID-19, expérience de symptômes, se faire tester pour la COVID, et la volonté de se

⁴ Le contenu optionnel sera inclus dans le fichier de 2019-2020 s'il a été obtenu des répondants d'une province ou d'un territoire pendant la période de deux ans. Autrement, il sera inclus uniquement dans le fichier de l'année pendant laquelle il a été recueilli. Il est à noter que si un module de contenu thématique d'une année est choisi dans le contenu optionnel d'une province ou d'un territoire au cours de la deuxième année, le module sera inclus dans le fichier de données de deux ans et sera traité comme contenu optionnel.

⁵ Le contenu thématique d'une année n'est pas accessible dans le cas des territoires à moins qu'ils aient sélectionné ce module dans leur contenu optionnel pour l'autre année.

⁶ Le nouveau module sur la COVID-19 n'a été demandé que de septembre à décembre. Les répondants de la période de collecte de janvier à mars verront ces variables définies comme « Ne s'applique pas ».

faire vacciner.

Note importante: les réponses du module COV pour les périodes de collecte de septembre à décembre ont été incluses dans le fichier annuel 2020, mais non sur les fichiers de 2019-2020. Pour les répondants de la période de collecte de janvier à mars 2020, les variables COV ont été codées comme « Ne s'applique pas ». Puisque le poids au niveau de la personne (WTS_M) dans le fichier 2020 est conçu pour l'année complète, les utilisateurs ne devraient pas produire de totaux de population pour le contenu du module COV. Les utilisateurs devraient choisir d'utiliser le fichier sous-annuel de septembre à décembre pour cette analyse.

5. PLAN D'ÉCHANTILLONNAGE

5.1 Population cible

L'ESCC vise la population âgée de 12 ans et plus vivant dans les dix provinces. Les trois territoires sont couverts au cours d'un intervalle de deux ans de collecte. Les personnes exclues de la couverture de l'enquête comprennent : les personnes habitant sur des réserves et d'autres établissements autochtones dans les provinces; les membres à temps plein des Forces canadiennes; les jeunes âgés de 12 à 17 ans demeurant dans des foyers d'accueil; les personnes placées dans des établissements et les personnes habitant dans les régions sociosanitaires du Québec du Nunavik et de la région des Terres-Cries-de-la-Baie-James. En tout, ces exclusions représentent moins de 3 % de la population cible.

En 2020, puisque les entrevues en personne n'étaient pas permises pour une bonne partie de l'année dans le contexte de la pandémie de COVID-19, la couverture dans les territoires a exceptionnellement été limitée aux trois capitales (Whitehorse, Yellowknife et Iqaluit) lors des trois dernières périodes de collecte. Considérant les régions qui n'ont pas pu être visitées, la couverture de la population cible pour l'ESCC 2019-2020 est estimée à 82 % au Yukon, 70 % aux Territoires du Nord-Ouest et 56 % au Nunavut⁷.

5.2 Régions sociosanitaires

À des fins administratives, chaque province est divisée en plusieurs régions sociosanitaires (RS) et chaque territoire est considéré comme formant une RS unique. En collaboration avec les provinces, Statistique Canada est parfois appelé à modifier les limites de certaines RS afin qu'elles correspondent aux données géographiques du recensement ou encore afin de mieux tenir compte des besoins en données sur la santé selon de nouvelles bornes géographiques. Dans le cadre de l'ESCC de 2019-2020, des données ont été recueillies pour plus de 100 RS dans les dix provinces, ainsi que pour une RS par territoire (Annexe C).

⁷ La couverture de la population cible a été estimée à partir des comptes de population totale du Recensement de 2016.

5.3 Taille et répartition de l'échantillon

Afin de produire des estimations fiables pour chaque RS et compte tenu du budget accordé pour l'ESCC, il a été établi que l'enquête devait être réalisée auprès d'un échantillon d'environ 130 000 personnes sur une période de 2 ans. Depuis 2015, les échantillons de la population des jeunes âgés de 12 à 17 ans et de la population adulte âgée de 18 ans et plus sont traités séparément. Suivant les tailles historiques des échantillons, 120 000 des 130 000 répondants ont été rattachés à la population adulte et environ 10 000, à la population des jeunes.

La production d'estimations fiables pour chaque RS était l'objectif principal, mais la qualité des estimations pour certaines caractéristiques importantes pour les provinces a été jugée importante également. Par conséquent, la stratégie de répartition de l'échantillon, qui comporte trois étapes, accorde une importance plus ou moins égale aux RS et aux provinces. À la première étape, on impose une taille minimum de 500 répondants par RS, considérée comme étant le minimum requis afin d'obtenir un niveau de qualité de données raisonnable. Par contre, pour des raisons de fardeau de réponse, on impose une fraction de sondage maximum de 1 sur 20 logements afin d'éviter d'échantillonner trop de logements dans des petites régions qui sont aussi interpellées par d'autres enquêtes. Il est à noter que très peu de RS ont une taille de moins de 500 répondants en raison de la limite de fractionnement de l'échantillonnage. Les provinces et les territoires sont traités séparément dans un premier temps. Un échantillon de 117 000 répondants est rattaché aux provinces et 3 000 répondants sont rattachés aux territoires. À l'intérieur de chaque groupe, l'échantillon est réparti en vertu d'un taux de répartition de 0,75 selon la taille de la population. Le tableau 5.1 fait état des tailles d'échantillon visées pour une période de deux ans, telles qu'établies lors du remaniement de l'ESCC en 2015. Il est à noter que la taille de l'échantillon et la répartition des territoires permettent seulement une représentativité sur une période de deux ans.

Tableau 5.1 Nombre de régions sociosanitaires et tailles d'échantillon visées selon la province ou le territoire, 2019-2020

Province ou territoire	Nombre de RS	Taille d'échantillon visée pour 2019-2020	
		Adultes	Jeunes
Terre-Neuve-et-Labrador	4	3 172	313
Île-du-Prince-Édouard	1	2 116	221
Nouvelle-Écosse	4	4 952	385
Nouveau-Brunswick	7	4 182	328
Québec	16	24 804	1 896
Ontario	35	36 354	3 032
Manitoba	5	5 804	522
Saskatchewan	11	5 128	453
Alberta	5	13 984	1 175
Colombie-Britannique	16	16 388	1 259
Yukon	1	899	166
Territoires du Nord-Ouest	1	979	166
Nunavut	1	722	166
Canada	107	119 484	10 082

On rattache ensuite à l'intérieur de chaque province l'échantillon aux RS en utilisant un taux de répartition de 0,35 en fonction de la taille de la population de la RS. Le nombre de répondants alloué à chaque RS est ensuite gonflé pour compenser la non-réponse et les unités hors du champ d'observation. On appelle la taille ainsi gonflée d'un échantillon la taille brute de l'échantillon; elle correspond au nombre de logements devant être échantillonnés pour l'obtention du nombre requis de répondants. Les tailles d'échantillon brutes provinciales déterminées lors du remaniement de 2015 ont été gardées stables lors des cycles suivants, tandis que la répartition par RS à l'intérieur de chacune des provinces est revue annuellement pour tenir compte de l'information la plus récente quant aux taux attendus d'unités hors du champ d'observation et de non-réponse. L'annexe D fournit les tailles des échantillons brutes par province/territoire pour l'ESCC 2019-2020.

5.4 Bases de sondage et stratégies d'échantillonnage des ménages

L'échantillon de l'ESCC est sélectionné au moyen de différentes bases de sondage selon le groupe d'âge. Dans le cas de la population adulte (18 ans et plus), l'échantillon de ménages est sélectionné à partir d'une base aréolaire. Dans le cas de la population des jeunes (12 à 17 ans), on utilise une liste pour sélectionner les personnes.

5.4.1 Échantillonnage des ménages à partir de la base aréolaire pour la population adulte

La base aréolaire utilisée pour l'Enquête sur la population active (EPA) est employée comme base de sondage dans le cas de la population adulte. L'EPA utilise un plan d'échantillonnage à deux

degrés⁸. Au premier degré, un échantillon d'unités primaires d'échantillonnage (UPE) est sélectionné, correspondant à des régions géographiques appelées *grappes*. Dans chacune des UPE sélectionnées, un échantillon de logements est sélectionné au second degré. Des détails au sujet de la base aréolaire et du plan d'échantillonnage de l'EPA sont accessibles à partir du lien ci-dessous : <https://www150.statcan.gc.ca/n1/fr/catalogue/71-526-x>

Échantillonnage des ménages à partir de la base aréolaire dans les provinces

Les provinces sont divisées en grappes, des régions géographiques généralement composées de 100 à 600 logements. Chaque grappe est classée dans l'un des six groupes de renouvellement de l'EPA, et les grappes sont regroupées en groupes géographiques ou socioéconomiques homogènes et mutuellement exclusifs, appelés strates. La base aréolaire contient toutes les UPE et leur strate correspondante.

Au premier degré d'échantillonnage, un échantillon de grappes est sélectionné dans chaque strate. Au deuxième degré, des listes de logements sont dressées pour chaque grappe sélectionnée, et les logements sont sélectionnés dans ces listes par échantillonnage systématique.

Chaque grappe est divisée en un ensemble d'origines pour permettre l'échantillonnage systématique des logements au sein de chaque grappe. Par exemple, une grappe de 300 logements pourrait être divisée en 20 origines. Sur une liste dans laquelle les 300 logements sont classés selon leur emplacement, la première origine correspondrait aux logements no 1, no 21, no 41..., no 281; la deuxième origine correspondrait aux logements no 2, no 22, no 42..., no 282; et ainsi de suite. La taille de l'échantillon fournie par une origine donnée (15 logements dans le cas présent) est appelée le rendement de l'origine.

La division des grappes en origines a été effectuée au moment de la dernière refonte de l'EPA en 2015, et le rendement prévu de chaque origine a été calculé à ce moment-là. Lorsqu'une grappe est sélectionnée une première fois pour faire partie d'une enquête, une liste actuelle des logements se trouvant dans cette grappe est dressée. En appliquant le taux d'échantillonnage qui avait été déterminé lors de la refonte de l'EPA au plus récent dénombrement des logements, on peut déterminer de façon plus précise le rendement prévu pour les origines de la grappe.

Puisqu'on connaît le rendement prévu de chaque origine, le processus de réservation de l'échantillon de l'ESCC a pour objectif de déterminer combien d'origines doivent être sélectionnées dans chaque strate de l'EPA de sorte que le rendement prévu soit suffisamment important pour obtenir la taille d'échantillon attribuée à chacune des strates de l'ESCC. Au moment de choisir les origines dans une strate de l'EPA en particulier, il faut s'assurer que le même nombre d'origines a été sélectionné dans chacun des six groupes de renouvellement de l'EPA. Autrement dit, le nombre d'origines sélectionnées doit toujours être un multiple de six⁹.

⁸ À l'exception de l'Île-du-Prince-Édouard, pour laquelle un échantillonnage aléatoire simple stratifié est utilisé.

⁹ Quelques strates de l'EPA ont deux ou trois grappes pour chaque groupe de renouvellement, ce qui signifie que les origines sont présentées par groupes de 12 ou de 18.

Les strates de l'EPA et celles de l'ESCC ont des limites géographiques différentes, ce qui signifie que les origines sélectionnées à partir d'une certaine strate de l'EPA peuvent fournir un échantillon pour différentes strates de l'ESCC. Étant donné que nous veillons à ce que le rendement total prévu à l'intérieur de chaque strate de l'ESCC corresponde au moins à la taille de l'échantillon requis, et que les origines soient sélectionnées par groupes de six, le rendement total prévu obtenu à partir des origines sélectionnées dépassera généralement, sauf exception, le minimum requis pour chacune des strates de l'ESCC. Le cas échéant, un sous-échantillon de logements est sélectionné dans chaque strate de l'ESCC au moyen d'un processus appelé stabilisation, de sorte que la taille de l'échantillon final corresponde à la répartition de l'échantillon cible pour chaque strate de l'ESCC.

Le processus de stabilisation n'a pas été exécuté lors des deux dernières périodes de collecte de 2020, pour augmenter la taille d'échantillon étant donné la baisse attendue des taux de réponse.

Échantillonnage des ménages à partir de la base aréolaire dans les trois territoires

Pour des raisons opérationnelles, le plan d'échantillonnage à partir de la base aréolaire de l'EPA pour les trois territoires est différent. Pour chaque territoire, les collectivités les plus peuplées ont chacune leur propre strate, alors que les collectivités les moins peuplées sont groupées en strates en fonction de diverses caractéristiques (population, données géographiques, proportion d'Inuit ou d'Autochtones et revenu médian du ménage). L'EPA a défini six strates pour le Yukon, dix pour les Territoires du Nord-Ouest et dix pour le Nunavut. Pour les strates regroupant plus d'une collectivité, le premier degré d'échantillonnage consiste à sélectionner aléatoirement une collectivité avec probabilité proportionnelle à la taille de la population dans chaque strate. Puis, à l'intérieur de chaque collectivité, le deuxième degré d'échantillonnage consiste à sélectionner des ménages de façon identique à la stratégie d'échantillonnage décrite plus haut. L'ESCC a sélectionné son échantillon à partir des mêmes collectivités sélectionnées par l'EPA tout en s'assurant de sélectionner des logements différents. Si trop ou pas assez de logements étaient disponibles pour une collectivité à l'intérieur d'une strate, une autre collectivité était choisie pour l'ESCC. Pour les collectivités les plus peuplées ayant leur propre strate, seulement un plan à un degré était nécessaire pour sélectionner directement les ménages de façon identique à la stratégie d'échantillonnage décrite plus haut.

Il convient de mentionner que la base de sondage de l'ESCC couvrait 94 % de la population ciblée au Yukon, 96 % de celle des Territoires du Nord-Ouest et 93 % de celle du Nunavut. Pour l'ESCC 2020, puisque les entrevues en personne n'étaient pas possibles après la mi-mars, la couverture dans les territoires a été limitée aux trois capitales (Whitehorse, Yellowknife et Iqaluit) lors des trois dernières périodes de collecte, résultant en une diminution de la couverture par rapport à un cycle de deux ans usuel de l'ESCC.

5.4.2 Échantillonnage des personnes à partir de la base liste pour la population des jeunes

Pour l'échantillonnage des membres de la population des jeunes âgés de 12 à 17 ans, l'ESCC utilise une base liste créée à partir des dossiers de l'Allocation canadienne pour enfants (ACE). Les fichiers de l'ACE renferment une liste de tous les bénéficiaires du programme comportant leurs

noms, adresses et numéros de téléphone. On utilise cette liste pour sélectionner directement les jeunes qui seront interviewés.

5.5 Répartition de l'échantillon par période de collecte des données

Afin d'équilibrer la charge de travail des intervieweurs et de réduire au minimum les possibles effets saisonniers sur les estimations, l'échantillon initial de chacune des bases de sondage est réparti également parmi chacune des quatre périodes de collecte.

Pour l'échantillon de la base aréolaire (adultes), chaque UPE sélectionnée dans chaque RS a été répartie au hasard à une période de collecte en tenant compte de plusieurs contraintes liées aux opérations sur le terrain ou encore à la pondération tout en conservant une taille égale par période.

Des échantillons indépendants ont été sélectionnés lors de chaque période de collecte pour ce qui est de l'échantillon des jeunes sélectionnés à partir d'une liste. Cette stratégie nous assure que chaque échantillon est représentatif de la population canadienne âgée de 12 à 17 ans admissible.

En 2020, la collecte des données a été interrompue de la mi-mars à la fin août. La première période de collecte a donc eu lieu de janvier à la mi-mars, tandis que les trois dernières périodes de collecte se sont déroulées de septembre à la mi-décembre.

5.6 Échantillonnage des personnes interviewées

Comme l'échantillon de la population de jeunes est sélectionné à partir d'une liste de personnes, la sélection des personnes interviewées est automatiquement gérée au moment de l'échantillonnage et aucune mesure supplémentaire n'est requise lors de la collecte. La sélection des interviewés au moment de la collecte est désormais seulement nécessaire dans le cas de la population adulte.

La stratégie d'échantillonnage adoptée a tenu compte des besoins des utilisateurs de données, du coût, de l'efficacité du plan d'échantillonnage, du fardeau de réponse et des contraintes opérationnelles. En ce qui concerne les ménages sélectionnés dans le cadre de l'échantillon des personnes de 18 ans et plus, une seule personne par ménage est sélectionnée au moyen de diverses probabilités tenant compte de l'âge et de la composition du ménage. Les probabilités choisies font suite à des résultats de simulations reposant sur divers paramètres dans le but de déterminer l'approche optimale sans générer de poids d'échantillonnage extrêmes en fin de compte.

Le tableau 5.2 donne les facteurs de sélection au niveau de la personne utilisés pour déterminer les probabilités de sélection des personnes dans les ménages échantillonnés selon le groupe d'âge. Par exemple, pour un ménage de trois personnes formé de deux adultes âgés entre 35 et 64 et d'un jeune de 19 ans, le jeune aura 1 chance sur 2 d'être sélectionné (c.-à-d. $20/(20+10+10)$) tandis que chacun des adultes aura 1 chance sur 4 d'être sélectionné. Afin d'éviter l'obtention de poids extrêmes, il y a une exception à la règle : si la taille du ménage est plus grande ou égale à 5, alors le facteur de sélection au niveau de la personne est égal à 1 pour toutes les personnes du ménage. Dans ce cas, toutes les personnes du ménage ont la même probabilité d'être sélectionnées.

Tableau 5.2 Facteurs de sélection au niveau de la personne, selon l'âge

Facteurs de sélection au niveau de la personne				
Âge	18-34	35-49	50-64	65+
Facteur	20	10	10	20

5.7 Achat d'unités d'échantillonnage supplémentaires pour l'Ontario

Pour 2019-2020, la province de l'Ontario est divisée en 14 réseaux locaux d'intégration des services de santé (RLISS), qui sont chacun sous-divisés en sous-RLISS (76 sous-RLISS au total). La province de l'Ontario a demandé une augmentation de l'échantillon afin de produire des estimations à l'échelle des sous-RLISS sur une période de deux ans. Pour ce faire, la stratification de l'ESCC a dû être rajustée.

Les limites des RS et des sous-RLISS de l'Ontario se chevauchent. Pour préserver la répartition initiale de l'ESCC à l'échelle des RS, la stratification rajustée a été établie de sorte que chaque strate corresponde à une seule RS (c'est-à-dire qu'aucune strate de l'ESCC ne correspond à plusieurs RS). Certaines de ces RS ont dû être subdivisées en plusieurs strates de l'ESCC, selon les intersections RS*sous-RLISS, pour un total de 96 strates.

Dans certains cas, l'échantillon habituel de l'ESCC donnait suffisamment d'unités pour obtenir le nombre cible de répondants nécessaire afin de produire des estimations à l'échelle des sous-RLISS. Dans le cas des sous-RLISS pour lesquels la cible n'était pas atteinte, l'échantillon dans la strate de l'ESCC correspondant aux sous-RLISS déficients a été augmenté dans le but d'atteindre les cibles des sous-RLISS. Pour un certain nombre de sous-RLISS, il n'était pas possible d'atteindre la cible en raison du faible chiffre de population ou des limites de la base aréolaire à fournir un échantillon suffisant dans de petites régions géographiques précises.

Pour ce qui est de l'échantillonnage à l'échelle de la personne, les unités supplémentaires sont traitées de la même façon que les autres unités de l'ESCC sélectionnées à partir de la base aréolaire. Autrement dit, les personnes interviewées sont sélectionnées en utilisant une liste et des facteurs de sélection comme décrit à la section 5.6. La principale différence dans la manière dont les deux ensembles d'unités sont traités se trouve dans le contenu de l'enquête : les unités achetées ne peuvent pas recevoir le contenu supplémentaire de l'ECSA ou de contenu de réponse rapide.

5.8 Suréchantillon de l'ESCC aux fins de l'Enquête canadienne sur la santé des aînés

L'Enquête canadienne sur la santé des aînés (ECSA), un supplément à l'ESCC de 2019 et 2020, a été conçue pour recueillir des données sur la santé chez la population canadienne âgée de 65 ans et plus. Les estimations de l'ECSA devaient être produites à l'échelle provinciale pour les personnes âgées de 65 ans et plus ainsi qu'à l'échelle nationale pour les personnes âgées de 85 ans et plus. Un sur-échantillonnage a été nécessaire dans huit provinces pour produire des estimations pour un niveau de précision donné.

Il convient de souligner que les unités sélectionnées pour ce suréchantillon ne pouvaient pas recevoir de contenu de réponse rapide.

5.8.1 Base de sondage utilisée pour l'ESCA

De nombreuses stratégies d'échantillonnage ont été envisagées pour l'ESCA. Les avantages et les inconvénients de sélectionner directement les personnes, plutôt que de sélectionner d'abord les logements puis les personnes, ainsi que d'utiliser différentes bases ont été évalués. Finalement, en raison de la disponibilité de l'échantillon, de la couverture, de l'effet de plan et de la qualité des coordonnées (dans un environnement d'ITAO seulement), trois tableaux du service de la Base de sondage pour les enquêtes auprès des ménages (BSEM) ont été choisis pour servir de fondement à une base de sondage. Les trois grands tableaux couplés du service de la BSEM sont les suivants : le Fichier de l'univers des logements (FUL), pour fournir des adresses; le Fichier des indicateurs socioéconomiques (FSE), pour obtenir des renseignements démographiques sur les résidents des logements dans le FUL; et le Fichier des numéros de téléphone résidentiels (FNTR), pour obtenir une liste de numéros de téléphone associés à chaque logement dans le FUL.

La base de sondage finale utilisée pour le suréchantillon de l'ESCC aux fins de l'ESCA (appelée la base de sondage de l'ESCA dans l'ensemble de ce document) a été créée en couplant ces trois fichiers du service de la BSEM et en supprimant tous les logements qui ne comptaient pas de personne âgée de 65 ans et plus, selon les données du FSE. La base de sondage de l'ESCA qui en a résulté renfermait une liste de logements canadiens qui étaient susceptibles de compter une personne âgée d'au moins 65 ans. Un échantillon aléatoire simple de logements dans chaque province admissible a été sélectionné, selon la répartition de l'échantillon décrite à la section 5.8.2.

5.8.2 Détermination de la taille de l'échantillon pour l'ESCA

La taille de l'échantillon pour l'ESCA devait être suffisamment grande pour permettre d'atteindre les objectifs de l'enquête et reposait sur les taux de prévalence cibles, le degré de précision attendu, les taux attendus de réponse et d'unités faisant partie du champ de l'enquête et les tailles de population provinciales. Le nombre de personnes en Ontario et au Québec âgées de 65 ans et plus qui répondent généralement à l'ESCC chaque année était suffisant pour atteindre les cibles provinciales. En conséquence, des suréchantillons étaient seulement nécessaires pour les huit autres provinces.

Les tailles d'échantillon finales sont présentées dans l'annexe D.

5.8.3 Sélection des personnes interviewées dans le cadre de l'ESCA

Comme dans le cas des unités de l'ESCC échantillonnées à partir de la base aréolaire, une liste des membres des ménages est compilée pour les logements sélectionnés à partir de la base de sondage de l'ESCA. Si un logement ne comptait finalement pas d'ainé, l'unité était alors considérée comme hors champ. Si le logement comptait au moins un aîné, un aîné devait alors être sélectionné au hasard pour répondre à l'ESCC et à l'ESCA. Cette sélection a été faite en donnant aux facteurs de sélection au niveau de la personne figurant dans la liste la valeur de 1 pour chaque personne âgée

de 65 ans et plus admissible au sein du logement, ce qui signifie que chaque aîné au sein d'un logement échantillonné pour l'ECSA avait les mêmes chances d'être sélectionné.

6. COLLECTE DES DONNÉES

6.1 Interviews assistées par ordinateur

L'ESCC utilise deux applications d'interviews assistés par ordinateur (IAO) distinctes pour la collecte des données, l'une pour les interviews téléphoniques (ITAO), l'autre pour les interviews sur place (IPAO). Cette façon de procéder permet d'adapter la fonctionnalité de chaque application au type d'interview menée. Chaque application comprend une composante d'entrée, une composante de contenu sur la santé et une composante de sortie. Les entrevues sur places (IPAO), qui sont traditionnellement tentées en personne en premier, n'ont été effectuées que pour les trois premiers mois de collecte. Toutes les activités de collecte ont été suspendues à la fin mars en raison de la pandémie de COVID-19. La collecte n'a repris qu'en septembre, date à laquelle toutes les entrevues ont été exclusivement menées par téléphone afin de respecter les bonnes pratiques et les protocoles de sécurité. Chacune de ces applications fait également partie d'une plateforme ou système de collecte ITAO ou IPAO, utilisé pour transmettre les données à partir des ordinateurs de l'intervieweur.

Entre janvier et décembre 2019, environ 64 900 interviews valides ont été effectuées par IAO. Environ 25 % de ces interviews ont été réalisées en personne par IPAO, tandis que les autres 75 % ont été réalisés au téléphone par ITAO. Au cours de l'année de collecte 2020, environ 42 132 entrevues valides ont été menées par IAO. Environ 6 % de ces cas complétés ont été menés en personne à l'aide de la plateforme IPAO. Il est à noter que ces entrevues en personne n'ont été menées qu'au cours des 3 premiers mois de l'année de collecte 2020, avant la cessation de toutes les activités liées à la collecte en raison de la pandémie de COVID-19. Les autres 94% des entrevues ont été menées par téléphone en utilisant à la fois les plateformes IPAO et ITAO.

L'IAO offre deux principaux avantages par rapport aux autres méthodes de collecte. D'abord, la technique est étayée d'un système de gestion des cas et d'une fonctionnalité de transmission de données. Le système de gestion des cas enregistre automatiquement de l'information de gestion importante sur chaque tentative effectuée au sujet d'un cas et produit des rapports de gestion de la collecte. L'IAO fournit également un planificateur d'appels automatisé, c'est-à-dire un système central pour optimiser le timing des rappels et la planification des rendez-vous utilisés pour la plateforme de collecte ITAO.

Le système de gestion des cas achemine les applications de questionnaire et les fichiers d'échantillons du bureau central de Statistique Canada aux ordinateurs des intervieweurs. Les données destinées au Bureau central sont acheminées en sens inverse. Par souci de confidentialité, les données sont chiffrées avant la transmission. Elles sont ensuite déchiffrées une fois sauvegardées sur un ordinateur sécurisé distinct, sans accès à distance.

Deuxièmement, l'IAO permet une interview personnalisée conçue à l'intention de chaque

répondant en fonction de ses caractéristiques particulières et de ses réponses à l'enquête. Notamment:

- l'application saute automatiquement les questions qui ne s'appliquent pas au répondant;
- des règles de vérification sont appliquées automatiquement pour repérer les réponses incohérentes ou non incluses dans l'éventail de valeurs permises, et des messages-guides apparaissent à l'écran en réaction à une inscription non valable. De cette façon, l'intervieweur reçoit une rétroaction immédiate et peut corriger toute incohérence;
- le texte des questions, y compris les périodes de référence et les pronoms, est personnalisé automatiquement d'après des facteurs comme l'âge et le sexe du répondant, la date de l'interview et les réponses aux questions précédentes.

6.2 Développement des applications de l'ESCC

Des composantes Entrée et Sortie ont été mises au point pour les ITAO et IPAO. Les composantes Entrée et Sortie comprennent des séries standard de questions auxquelles l'intervieweur a pu se référer pour prendre contact avec un répondant, choisir les répondants, recueillir de l'information importante sur l'échantillon et évaluer l'état des cas. La composante de Contenu sur la santé comprend les modules sur la santé et représente la plus grande partie de l'application. Il s'agit, notamment, des modules communs posés à tous les répondants et de modules optionnels, qui diffèrent selon les provinces et les territoires. Chacune des applications a fait l'objet de trois étapes de mise à l'essai : les tests modulaires, intégrés et de bout en bout.

Les tests modulaires consistent à mettre à l'essai indépendamment chaque module de contenu afin de vérifier la spécification exacte des instructions « passez à », la logique d'enchaînement et le texte, dans les deux langues officielles. À cette étape, les instructions « passez à » et la logique d'enchaînement entre modules ne sont pas testées, car chaque module est considéré comme un questionnaire autonome. Lorsque les responsables des essais ont terminé la vérification de tous les modules, ces derniers sont regroupés en applications intégrées avec les composantes Entrée et Sortie. À ce moment, les applications intégrées passent à l'étape suivante des essais.

Les essais intégrés ont lieu au moment du regroupement des modules mis à l'essai et des composantes Entrée et Sortie dans une application intégrée. Ce second stade d'essais vérifie la transmission de l'information démographique essentielle, comme l'âge, le sexe et le statut familial économique de la composante Entrée au contenu sur la santé et à la composante Sortie des applications. Il confirme également que les variables qui influent sur les instructions « passez à » et la logique d'enchaînement sont transmises correctement de module en module à l'intérieur de la composante du Contenu sur la santé. À cette étape, comme le fonctionnement des applications est essentiellement identique à ce qu'il sera sur le terrain, tous les scénarios possibles auxquels feront face les intervieweurs peuvent être simulés, afin d'en assurer la fonctionnalité rigoureuse. Ces scénarios permettent de vérifier divers aspects des composantes Entrée et Sortie, y compris la prise de contact, la collecte d'information sur le contact, la question de savoir si un cas fait partie du champ de l'enquête, le listage de ménages, la prise de rendez-vous et la sélection des répondants. Les tests servent également à confirmer que, au cours d'une interview, les modules de

contenu optionnel choisis sont activés.

Les essais de bout en bout situent les applications entièrement intégrées dans un environnement de collecte simulé. Les applications sont chargées dans des ordinateurs connectés à un serveur d'essai. Ensuite, des données sont recueillies, transmises et extraites en temps réel, comme ce serait le cas sur le terrain. Cette dernière étape permet d'expérimenter tous les aspects techniques de la saisie, de la transmission et de l'extraction des données pour chacune des applications de l'ESCC. Il s'agit, par ailleurs, d'une dernière occasion de déceler des erreurs dans les composantes Entrée, Contenu sur la santé et Sortie.

6.3 Formation des intervieweurs

Les gestionnaires de projet, les intervieweurs principaux et les intervieweurs des bureaux régionaux responsables de la collecte des données de l'ESCC ont reçu une trousse de formation pour de l'auto- apprentissage avant le début de la collecte. Ces trousse ont été préparées par l'équipe de projet de l'ESCC en vue d'être utilisées par les intervieweurs expérimentés de l'ESCC afin de renforcer leur formation antérieure. Les gestionnaires de projet et les intervieweurs principaux ont également mené, au besoin, des séances de formation pour les nouveaux intervieweurs. Enfin, des séances de formation mettant l'emphasis sur des sujets spécifiques reliés à la collecte des données de l'ESCC se sont déroulées sur une base mensuelle. L'objet des séances de formation était de familiariser les intervieweurs avec les applications de l'enquête, le contenu et les procédures d'entrevue spécifiques à l'ESCC 2020. La formation était centrée sur :

- les buts et objectifs de l'enquête, incluant une partie centrée sur le remaniement de l'enquête
- les techniques d'enquête;
- les fonctionnalités des applications;
- l'examen du contenu des questionnaires et des exercices axés sur les changements significatifs apportés au contenu;
- les techniques à appliquer par l'intervieweur pour maintenir le taux de réponse et des exercices pour réduire au minimum la non-réponse;
- la simulation d'interviews difficiles et de situations de non-réponse;
- la gestion de l'enquête;
- les procédures de transmission.

La formation visait en priorité à réduire au minimum les cas de non-réponse. À cette fin, les intervieweurs ont participé à des exercices qui consistaient à persuader des répondants réticents de participer à l'enquête. En outre, les intervieweurs principaux responsables de la conversion des cas de refus à chaque bureau régional de collecte ont participé à une série d'ateliers sur la façon d'éviter les refus.

6.4 L'interview

Les unités de l'échantillon ont été sélectionnées à partir d'une base aréolaire dans le cas de la population âgée de 18 ans et plus et à partir de la base de l'Allocation canadienne pour enfants

(ACE) dans le cas de la population âgée de 12 à 17 ans. Environ 60% des répondants sélectionnés à partir de la base aréolaire et tous les répondants sélectionnés à partir de la base ACE ont été interviewés en utilisant la plateforme de collecte ITAO. Environ 40% des répondants de la base aréolaire ont été interviewés avec la plateforme IPAO. Tous les intervieweurs ont effectué leur travail de façon autonome à partir de leur foyer au moyen d'ordinateurs portatifs et ils étaient surveillés à distance par des intervieweurs principaux. La variable CASEMODE dans le fichier de microdonnées indique la base de sondage de laquelle un cas a été sélectionné, ainsi que le mode de collecte utilisé. Voici la répartition des variables CASEMODE :

- CASEMODE = 0 : réalisé avec la plateforme de collecte IPAO
- CASEMODE = 1 : réalisé avec la plateforme de collecte ITAO

Avant la suspension de toutes les interviews en raison de la pandémie de COVID-19, les intervieweurs qui se servaient de la méthode de l'IPAO ont reçu la formation nécessaire pour procéder à une première prise de contact sur place avec chaque ménage échantillonné. Si la première visite se soldait par une non-réponse, un suivi par téléphone était permis. La variable ADM_N40 apparaissant dans les fichiers de microdonnées indique si l'interview par IPAO (CASEMODE = 0) a été effectuée sur place, par téléphone ou au moyen d'une combinaison des deux techniques. Lorsque la collecte a repris en septembre 2020, toutes les entrevues ont été menées par téléphone uniquement.

Dans tous les logements choisis, l'intervieweur demandait à un membre du ménage bien informé de fournir l'information démographique de base sur tous les occupants. Puis, il sélectionnait un membre du ménage pour une interview plus approfondie, appelée interview du contenu sur la santé.

Par souci d'assurer la qualité des données recueillies, les intervieweurs avaient reçu instruction de prendre tous les moyens à leur disposition pour mener en privé l'interview avec le répondant choisi. Là où la situation était inévitable, le répondant a été interviewé en présence d'une autre personne. Dans les fichiers de microdonnées, des indicateurs signalent si une personne autre que le répondant était présente à l'interview (ADM_N045) et si, de l'avis de l'intervieweur, la présence de l'autre personne a influencé les réponses du répondant (ADM_N050).

De nombreuses techniques, y compris les suivantes, ont été mises en œuvre afin de parvenir à un taux de réponse optimal.

a) Lettres d'introduction

Avant le début de chaque période de collecte, les ménages échantillonnés ont reçu des lettres d'introduction et des brochures qui expliquaient l'objet de l'enquête. Elles énonçaient, notamment, l'importance de l'enquête et offraient des exemples de l'utilisation prévue des données tirées de l'ESCC.

b) Prise de contact

On a demandé aux intervieweurs de faire tous les efforts raisonnables pour obtenir des entrevues.

De nombreux rappels ont été effectués, à divers moments et différents jours. Lorsque l'appel de l'intervieweur ne convenait pas, un rendez-vous était pris pour rappeler à un meilleur moment.

c) Conversion des cas de refus

Si une personne refusait initialement de participer à l'enquête, le bureau régional de Statistique Canada lui faisait parvenir une lettre qui soulignait l'importance de l'enquête et de la coopération du ménage. Ensuite, un intervieweur principal, un surveillant de projet ou un autre intervieweur rappelait le répondant pour faire valoir l'importance de sa participation.

d) Obstacles linguistiques

Pour parer aux problèmes de langue susceptible de nuire aux interviews, tous les bureaux régionaux de Statistique Canada ont embauché des intervieweurs qui parlaient un grand nombre de langues. Au besoin, les cas étaient transférés à un intervieweur capable de remplir le questionnaire dans la langue voulue.

e) Interviews de jeunes

Les intervieweurs devaient obtenir l'autorisation verbale des parents ou tuteurs pour interviewer des répondants sélectionnés âgés de 12 à 14 ans. Cette autorisation a été obtenue dans le bloc du consentement des parents/tuteurs (PGC) de la composante d'entrée de l'enquête. Les intervieweurs ont fait appel à plusieurs procédures pour atténuer les éventuelles inquiétudes des parents et mener à terme les interviews.

Si un parent ou tuteur demandait à voir les questions, les intervieweurs avaient pour consigne de faire en sorte que le bureau régional leur envoie sur-le-champ un exemplaire du questionnaire.

S'il se révélait impossible d'interviewer en privé au téléphone le jeune sélectionné (sans qu'une autre personne soit à l'écoute), le code de refus était attribué à l'interview.

Le bloc de la personne la mieux renseignée (PMK) a recueilli les renseignements au niveau du ménage qui sont demandés à la fin de l'enquête (Couverture d'assurance, sécurité alimentaire, Revenu et renseignements administratifs) auprès de la personne la mieux renseignée du ménage. Ce bloc est activé quand le répondant sélectionné est âgé de 12 à 17 ans. Ce bloc officialise le processus qui consiste à identifier une personne du ménage qui est susceptible de mieux pouvoir répondre à ces questions au niveau du ménage que ne l'est le jeune répondant sélectionné. Si une PMK est trouvée, l'interview passe du jeune répondant sélectionné de 12 à 17 ans à un membre du ménage qui termine le reste de l'interview après le bloc PMK.

f) Interviews par procuration

Dans les cas où le répondant sélectionné était, pour des raisons de santé physique ou mentale, incapable de se prêter à l'interview, les renseignements à son sujet ont été fournis par un autre membre bien informé du ménage. C'est ce qu'on appelle une interview par procuration. Quoique les répondants substitués aient été en mesure de donner des réponses exactes à la plupart des questions de l'enquête, les questions plus délicates ou personnelles allaient au-delà des connaissances d'un répondant substitut. Par conséquent, certaines questions posées dans le cadre

des interviews par procuration sont demeurées sans réponse. Tous les efforts ont été faits pour réduire au minimum le nombre d'interviews par procuration.

Depuis 2010, on a modifié le bloc de l'interview par procuration (GR) pour faire en sorte que l'intervieweur indique expressément si l'interview était menée par procuration en raison d'un problème de santé physique ou d'un problème de santé mentale. Dans l'un ou l'autre de ces cas, l'intervieweur devait ensuite inscrire le problème particulier. La variable ADM_PRX indique si l'interview a été réalisée par procuration ou non.

6.5 Opérations sur le terrain

L'échantillon de 2019 a été réparti en 4 périodes de collecte de 3 mois. Les bureaux régionaux de collecte ont reçu instruction de passer les 4 premières semaines de chaque période de collecte à interviewer la majorité de l'échantillon, puis de consacrer les 8 semaines suivantes aux interviews restantes et au suivi des cas de non-réponse. La première période de collecte de 2020 (janvier à mars) a été traitée de la même manière que l'échantillon de 2019. L'échantillon pour l'ESCC - 2020 septembre à décembre a été réparti en trois périodes de collecte de 6 semaines sans chevauchement. Les bureaux régionaux de collecte ont reçu instruction de passer les premières quelques semaines de chaque période de collecte à interviewer la majorité de l'échantillon, puis de consacrer les semaines suivantes à finaliser le reste de l'échantillon et au suivi des cas de non-réponse. Tous les cas d'ITAO devaient avoir fait l'objet de tentatives avant la deuxième semaine de chaque période de collecte. Les fichiers d'échantillons ont été envoyés environ deux semaines avant le début de chaque période de collecte. Chaque échantillon comprenait une série de cas fictifs dont devaient s'occuper les intervieweurs principaux afin de confirmer que les procédures de transmission de données fonctionnaient bien tout au long du cycle de collecte. Après réception des échantillons, il incombait aux surveillants de projet de planifier les tâches des intervieweurs chargés des interviews selon la méthode de l'IPAO. Quand la situation s'y prêtait, les tâches étaient limitées à 15 cas par intervieweur pour ces intervieweurs.

Pour les taux de réponse finaux, veuillez consulter l'annexe E

6.6 Contrôle de la qualité et gestion de la collecte

Durant l'année de collecte, plusieurs méthodes ont servi à assurer la qualité des données et l'optimisation de la collecte des données. Il s'agit, entre autres, de mesures internes de vérification du rendement de l'intervieweur et d'une série de rapports de contrôle des diverses cibles de collecte et de la qualité des données.

Un système de validation a été utilisé régulièrement pour valider le travail des intervieweurs par le personnel de collecte principal. Des cas choisis ont été repérés aléatoirement dans les échantillons à chaque période de collecte. Par souci de repérer promptement les problèmes, la validation avait généralement lieu au cours des premières semaines d'une période de collecte. Ensuite, les surveillants adressaient régulièrement une rétroaction aux intervieweurs.

En 2014, une nouvelle approche en matière de gestion de la collecte a été mise en œuvre pour

améliorer l'efficacité de la collecte par ITAO : le plan de collecte adaptatif. Le plan de collecte adaptatif (PCA) est une nouvelle stratégie de collecte qui permet d'adapter les approches de collecte des données tout au long de la période de collecte. Essentiellement, cela permet de regrouper des cas selon un certain nombre de facteurs, puis de cibler les groupes présentant un intérêt particulier. Dans une enquête par ITAO ordinaire, les groupes d'intervieweurs demeurent constants tout au long de la collecte. Toutefois, selon le plan de collecte adaptatif, ces groupes changent à chaque phase de la collecte, afin de maximiser les taux de réponse, la productivité et la représentativité de l'échantillon recueilli.

Une stratégie dynamique d'obtention des numéros de téléphone a été mise au point dans le cadre du remaniement de l'ESCC de 2015. À partir de 2015, les répondants de l'échantillon de plus de 18 ans peuvent avoir jusqu'à deux numéros de téléphone, tandis que les répondants des échantillons de l'ACE peuvent avoir jusqu'à quatre numéros de téléphone. Cette mesure accroît le nombre de numéros de téléphone accessibles que peuvent composer les intervieweurs pour maximiser leurs possibilités de joindre les ménages ou les répondants sélectionnés.

7. TRAITEMENT DES DONNÉES

7.1 Vérification

La vérification des données a été exécutée en grande partie par l'application d'interview assistée par ordinateur (IAO) durant la collecte des données. Les intervieweurs ne pouvaient pas entrer de valeurs hors-normes et les erreurs d'enchaînement faisaient l'objet de programmes « passez à ». Par exemple, l'IAO s'assurait de ne pas poser au répondant les questions non pertinentes.

En réponse à certaines données incompatibles ou inhabituelles, on a signalé des messages d'avertissement, mais sans prendre de mesures correctrices au moment de l'interview. On a plutôt mis au point, le cas échéant, des versions révisées à appliquer après la collecte des données au bureau central. Les incohérences ont été le plus souvent corrigées en attribuant à l'une ou aux deux variables en question la valeur « Non déclaré ».

7.2 Flux

Lorsque des données recueillies sur le terrain sont transmises au bureau central, toutes les questions sautées seront vides à l'intérieur des fichiers de données préliminaires. Durant le traitement, l'un des codes de traitement qui suivent est attribué à chaque variable vide :

« Sans objet » - S.O.

Dans le cas des variables d'une longueur de 1, la valeur du code S.O. correspondra simplement à 6. Autrement, ces valeurs figureront dans le fichier sous la forme d'une série de 9, suivis d'un 6, selon la longueur de la variable numérique ou de la zone de texte.

Le code S.O. est utilisé chaque fois qu'une question est sautée parce qu'elle ne s'applique pas aux répondants (exclusion de la population). Ce code est principalement utilisé dans deux situations :

pour les variables du module optionnel dans le cas des répondants des provinces où le contenu en question n'a pas été retenu et lorsque des questions sont sautées dans le flux du questionnaire parce qu'elles ne s'appliquent pas, comme les questions au sujet des expériences maternelles dans le cas des répondants masculins.

« Non déclaré » - ND

Dans le cas d'une variable d'une longueur de 1, la valeur de ND sera simplement 9. Autrement, ces codes figureront dans le fichier sous la forme d'une série de 9, selon la longueur de la variable numérique ou de la zone de texte.

Le code ND est utilisé chaque fois qu'une question est sautée, mais pourrait avoir été posée. Les répondants en question font partie de l'univers, mais aucune réponse n'est présente. Les deux principales situations où ce code est employé sont celles où les questions sont sautées parce que le participant à l'interview est une personne interposée et celles où les questions sont sautées parce que la réponse à une question précédente, était « Ne sait pas », « Refuse de répondre » ou est elle-même « Non déclaré ». Par exemple, si un répondant refuse de répondre à la question CCC_Q095 (Avez-vous du diabète?), les questions l'interrogeant au sujet de l'âge du diagnostic, de l'utilisation d'insuline, etc., comporteront toutes la mention « Non déclaré », car on ne sait pas si le répondant souffre ou non du diabète.

7.3 Codage

On a fourni des catégories de réponses pré-codées pour toutes les variables appropriées. Les intervieweurs ont reçu une formation durant laquelle ils ont appris à classer les réponses recueillies dans la catégorie appropriée.

Dans les cas où la réponse donnée par le répondant ne pouvait être assignée facilement à une catégorie existante, l'intervieweur pouvait poser plusieurs questions lui permettant d'entrer une réponse en toutes lettres dans la catégorie « Autre – précisez ». Les réponses à toutes ces questions ont été examinées attentivement lors du traitement des données au bureau central. Dans certains cas, on a donné aux réponses en toutes lettres le code d'une catégorie figurant sur la liste si la réponse faisait double emploi. En 2019-2020, les variables portant le code OSLA (réponses ouvertes « Autre – Précisez ») comprennent GDR_010, SDC_037, CAN_025, CAN_035, STS_015, PG_01 et LOP_025. D'autres variables de ce cycle pour lesquelles des réponses « Autre – Précisez » sont requises n'ont pas été sélectionnées pour le codage. On tiendra compte des réponses « Autre – précisez » fournies pour toutes les questions lors du perfectionnement des catégories de réponses en vue de futurs cycles de l'enquête.

Lorsque les réponses écrites (« Autre – précisez ») sont codées dans une catégorie existante au moment du traitement au bureau central, il se peut que d'autres questions viennent s'ajouter au champ d'enquête pour le répondant. Comme ces questions n'ont jamais été posées pendant l'interview, les réponses manquantes étaient réglées à « non déclaré » pendant le traitement. Par exemple, dans le module Blessures (INJ), on demande au répondant ce qu'il faisait lorsqu'il s'est blessé (INJ_Q065). Si l'intervieweur a obtenu une réponse sous « Autre – précisez » indiquant qu'il travaille, la variable INJ_065 sera codée dans la catégorie « Occupe un emploi ou travaille

dans une entreprise (à l'exclusion des déplacements à destination et en provenance du travail) » et une valeur « 04 » y sera rattachée. Lorsque l'intervieweur utilise cette catégorie au cours de l'interview, le répondant sera invité à répondre au bloc de questions sur les accidents de travail (INW) si un tel contenu a été sélectionné comme contenu optionnel. Toutes les réponses aux questions d'INW devront alors correspondre à « Non déclaré ».

7.4 Création de variables dérivées

Pour faciliter l'analyse des données, on a dérivé un certain nombre de variables à partir des éléments disponibles sur le questionnaire de l'ESCC. Le quatrième caractère du nom des variables dérivées est en général un « D », « G » ou un « F ». Dans certains cas, les variables dérivées sont simples, donnant lieu à un regroupement des catégories de réponses. Dans d'autres cas, on a combiné plusieurs variables pour en créer une nouvelle. La *Documentation sur les variables dérivées (VD)* fournit des détails sur la façon de dériver ces variables plus complexes. Pour de plus amples renseignements concernant la nomenclature, veuillez vous référer à la section 12.6.

7.5 Pondération

Le principe de base de l'estimation dans un échantillon aléatoire comme celui de l'ESCC repose sur le fait que chaque personne représente, en plus d'elle-même, plusieurs autres personnes qui ne font pas partie de l'échantillon. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne en représente 50. Dans la terminologie en usage ici, nous dirons que nous avons attribué à chaque personne un facteur de pondération de 50.

L'étape de détermination des facteurs de pondération donne lieu au calcul du poids d'échantillonnage de chaque personne échantillonnée. Ce poids apparaît dans les fichiers de microdonnées et doit servir à extraire des estimations de l'enquête. Par exemple, si l'on doit évaluer le nombre de personnes qui fument tous les jours, on le fait en choisissant dans l'échantillon les enregistrements des personnes qui présentent cette caractéristique et en faisant la somme des facteurs de pondération que représentent ces enregistrements.

7.6 Variables du revenu pour 2019-2020 - couplage et imputation

Depuis 2016, les variables du revenu dans les fichiers maîtres de l'ESCC sont établies à partir d'une des trois sources suivantes : les dossiers fiscaux, les données fournies par les répondants et les données imputées. Lorsque les répondants ne se sont pas opposés à un couplage avec leurs données fiscales et lorsqu'un couplage avec les dossiers fiscaux appropriés pouvait être trouvé, les variables de source de revenu (INC_005A-O, INC_Q025A-O), la principale source de revenu (INC_015, INC_035), le supplément pour personnes handicapées, si reçu (INC_010, INC_030) et les montants de revenu total (INC_021, INC_041) ont été établis à partir des dossiers fiscaux. Lorsque le couplage avec les données fiscales n'était pas possible, les données des répondants ont été utilisées. Si ni les données fiscales couplées ni les données déclarées par le répondant n'étaient

disponibles, une imputation des variables du revenu personnel total et du revenu total du ménage a été effectuée. Les sources de revenu (INC_005A-O, INC_010, INC_015, INC_025A-O, INC_030, INC_035) n'ont pas été imputées.

La variable INCFIMPP indique la source utilisée pour les variables du revenu personnel et la variable INCFIMPH indique la source utilisée pour celles du revenu du ménage. À partir de 2019, toute information en provenance des données fiscales sera indiquée comme imputée. Pour de plus amples renseignements concernant les revenus dans le fichier maître, veuillez consulter le document intitulé « ESCC_2019-2020_Revenu_Fichier_Maître » accessible dans le répertoire de documentation de l'ESCC 2019-2020. Pour plus amples renseignements concernant les revenus dans le fichier de partage, veuillez consulter le document « ESCC_2019-2020_Revenu_Fichier_Partage ».

7.7 Imputation du code postal

Le code postal d'une unité échantillonnée est utilisé, de même que le fichier de conversion des codes postaux, pour dériver le reste des variables géographiques qui figurent dans le fichier de données de l'ESCC. Il est donc important que toutes les unités échantillonnées aient un code postal valide. Si le code postal d'une unité échantillonnée est manquant ou invalide, il est généralement imputé dans le cadre d'un processus d'imputation par enregistrement donneur, mais d'autres méthodes d'imputation sont aussi parfois utilisées. L'enregistrement donneur est choisi à partir de la même région géographique, avec le plus de précision possible, comme l'unité dont le code postal est manquant ou invalide.

Pour l'ESCC 2019-2020, le code postal d'environ 3,6 % des unités échantillonnées de la base aréolaire (âgés de 18 ans et plus) a été imputé. Pour ce qui est des unités échantillonnées de la base de l'ACE (âgés de 12 à 17 ans), le code postal d'environ 0,9 % de ceux-ci a été imputé pour 2019-2020. Dans le cas des unités échantillonnées de la base de l'ECSA (âgés de 65 ans et plus), le code postal de 3,8 % d'entre eux a été imputé pour 2019-2020.

8. PONDÉRATION

Pour que les estimations produites à partir de données d'enquête soient représentatives de la population couverte, et non pas seulement représentatives de l'échantillon comme tel, l'utilisateur doit incorporer les poids d'enquête dans ses calculs. Un poids d'enquête est attribué à chaque répondant inclus dans l'échantillon final. Ce poids correspond au nombre de personnes représentées par le répondant dans l'ensemble de la population de l'enquête.

Comme indiqué à la section 5, dans le cadre de l'ESCC, on utilise trois bases de sondage pour la sélection de l'échantillon : une base aréolaire dans le cas de la population canadienne âgée de 18 ans et plus, une liste supplémentaire de logements pour les personnes âgées de 65 ans et plus, et une liste de numéros de téléphone provenant des dossiers de l'Allocation canadienne pour enfants (ACE) dans le cas de la population âgée de 12 à 17 ans.

La base de sondage utilisée en vue d'obtenir un échantillon supplémentaire dans le cas des personnes âgées de 65 ans et plus (appelées « aînés » dans ce document) est destinée à l'Enquête canadienne sur la santé des aînés (ECSA), un supplément à l'ESCC. On a posé toutes les questions de l'ESCC aux personnes sélectionnées au moyen de ce suréchantillon, puis on leur a offert la possibilité de remplir le questionnaire de l'ECSA. L'ESCC inclura donc des aînés faisant partie à la fois de l'échantillon de l'ESCC et du suréchantillon de l'ECSA.

La stratégie de pondération traite de manière indépendante la base aréolaire et la base de l'ECSA jusqu'à l'ajustement lié à la non-réponse au niveau du ménage inclusivement; elle intègre ensuite les deux échantillons (avec ajustement des poids pour tenir compte du chevauchement de la population) pour aboutir à un poids au niveau de la personne pour la population adulte. La base de l'ACE est pondérée de manière indépendante des deux bases d'adultes jusqu'à la non-réponse au niveau de la personne inclusivement. Les poids au niveau de la personne de la base intégrée des adultes et de la base de l'ACE sont ensuite combinés en une seule série de poids. Ils font conjointement l'objet d'autres ajustements, notamment un appariement avec des totaux de population connus et ils deviennent les poids finaux au niveau de la personne.

8.1 Aperçu

Les sections suivantes décrivent le processus de pondération pour les provinces. La sous-section 8.2 fournit les détails de la stratégie de pondération visant la base aréolaire et la sous-section 8.3 traite de la stratégie visant la base de l'ECSA. La sous-section 8.4 fournit les détails de l'intégration de la base aréolaire et de la base de l'ECSA. La sous-section 8.5 concerne la stratégie visant la base de l'ACE. Les étapes finales de la pondération, soit la winsorisation et le calage, qui s'appliquent à la base intégrée des adultes et à la base de l'ACE sont décrites à la sous-section 8.6.

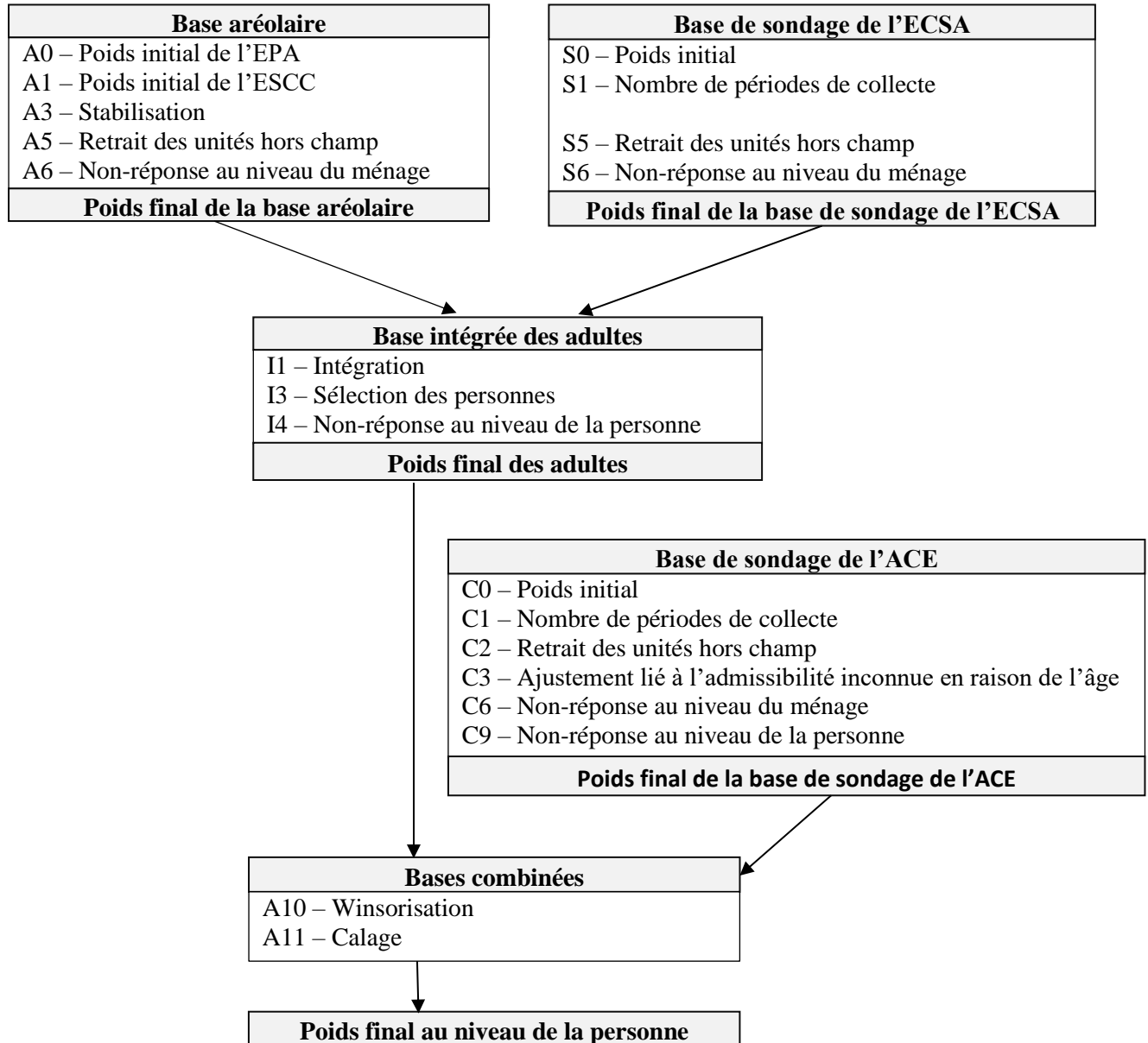
Suivant le remaniement de l'ESCC en 2015, la collecte d'un échantillon représentatif s'échelonne sur une période de deux ans en raison de contraintes opérationnelles. Cela signifie que les estimations pour les trois territoires seront seulement représentatives sur deux années. Ainsi, les répondants des territoires ne figureront pas dans les fichiers d'un an et ne seront inclus que dans les fichiers de deux ans (expliqué dans la sous-section 8.6)..

En 2020, la collecte et la méthodologie de l'ESCC ont dû être adaptées aux défis apportés par la pandémie de la COVID-19. La collecte a été interrompue de la mi-mars à août. La collecte a repris en septembre via téléphone seulement. Les taux de réponses sont également plus bas que les années antérieures. Pour ces raisons, certaines des étapes de pondération ont dû être modifiées. Ces modifications seront décrites en détails la fin de chacune des étapes décrites ci-dessous. Pour les territoires, le processus de pondération est examiné à la section 8.8, et les changements apportés au processus en raison de la pandémie sont examinés à la section 8.9. L'analyse au moyen du fichier de deux ans doit être faite avec prudence.

Plusieurs ajustements aux poids de sondage sont réalisés dans le cadre de la stratégie de pondération. Les ajustements appliqués aux unités de la base aréolaire sont numérotés de A0 à A6, les ajustements appliqués aux unités de la base de l'ECSA sont numérotés de S0 à S6, tandis que

les ajustements appliqués aux unités de la base de l'ACE sont numérotés de *C0* à *C9*. Les ajustements appliqués pour intégrer les deux bases d'adultes sont numérotés de *I1* à *I4*. Les ajustements finaux appliqués suivant la combinaison des bases d'adultes et de la base de l'ACE sont numérotés *A10* et *A11*. Le diagramme A présente un aperçu des ajustements les plus importants prévus dans le cadre de la stratégie de pondération.

Diagramme A - Aperçu de la stratégie de pondération



Les sous-sections 8.2 à 8.6 présentent en détail l'ensemble des étapes illustrées dans le diagramme A.

8.2 Pondération de l'échantillon de la base aréolaire

A0 – Poids initial de l'EPA

Le processus de pondération de l'échantillon de la base aréolaire débute avec un poids fourni par l'Enquête sur la population active (EPA). Ce poids est basé sur le plan de l'EPA parce que le plan d'échantillonnage de la base aréolaire de l'ESCC découle de celui de l'EPA. Le plan de l'EPA vise un échantillon de logements à l'intérieur de grappes sélectionnées dans des strates de l'EPA et les poids initiaux de l'EPA correspondent aux probabilités de sélection à l'intérieur des strates de l'EPA.

A1 – Poids initial de l'ESCC

Lors de l'ajustement initial A1, le poids initial de l'EPA est ajusté en fonction du fait que l'ESCC sélectionne un échantillon d'une taille suffisante pour l'estimation à l'échelon d'une région sociosanitaire (RS). L'ESCC sélectionne pour ce faire un nombre de grappes différent de l'EPA et peut répéter l'échantillonnage de logements dans les grappes sélectionnées. Le poids obtenu est appelé poids A1. Pour plus de détails sur le mécanisme de sélection, de même qu'une définition plus complète des strates de l'EPA et des grappes, voir Statistique Canada (2017)¹⁰.

A3 – Stabilisation

Dans certaines RS, l'accroissement de la taille de l'échantillon, tel que décrit dans la section 5, produit un échantillon beaucoup plus volumineux que nécessaire. On a recours à la stabilisation pour réduire la taille de l'échantillon au niveau souhaité. Pour limiter la dispersion des poids et ainsi réduire la variabilité des estimations, on effectue la stabilisation de l'une des deux façons possibles à l'intérieur de chaque RS. Si les poids initiaux dans chaque grappe d'une RS sont identiques, le processus consiste à sous-échantillonner des logements au hasard à l'échelle de la RS parmi les logements originalement sélectionnés dans chaque grappe. Si, toutefois, les poids initiaux ne sont pas identiques dans chaque grappe à l'intérieur d'une RS, on a recours à un processus itératif. On sélectionne d'abord la grappe dont le poids initial est le plus faible et on enlève une unité de cette grappe. Les poids sont ensuite recalculés aux fins d'un ajustement tenant compte de l'unité enlevée. Puis, on sélectionne encore une fois la grappe au poids le plus faible et on enlève une unité. Ce processus est répété jusqu'à l'obtention de la taille d'échantillon souhaitée. Un facteur d'ajustement représentant l'effet de la stabilisation est calculé pour la correction de la probabilité de sélection. Ce facteur, multiplié par le poids A1, produit le poids A3.

Exceptionnellement pour les périodes d'octobre 2020 et de novembre 2020, aucune stabilisation n'a été faite sur l'échantillon. Étant donné que seulement une collecte téléphonique a été utilisée pour ces périodes, ceci a été fait dans une tentative d'obtenir le plus de répondants possible. Pour ces deux périodes l'ajustement de cette étape de pondération a été de 1 pour tous les enregistrements.

A5 – Retrait des unités hors champ

Parmi tous les logements échantillonnés, une certaine proportion de ceux-ci est, lors de la collecte,

⁹ Statistics Canada, 2017. *Guide to the Labour Force Survey*, Statistics Canada. Cat. No. 71-543-G.

identifiée comme unités hors du champ de l'enquête. Les logements détruits ou en construction, les logements vacants, saisonniers ou secondaires, de même que les établissements, sont tous des exemples de cas hors champ pour l'ESCC. Les logements dans lesquels tous les résidents sont hors champ, par exemple ceux où tous les résidents sont des membres à temps plein des Forces armées ou des personnes en visite au Canada, sont également des logements hors champ. Tous les logements hors champ et les poids connexes sont simplement supprimés de l'échantillon, car on suppose que les logements hors champ pondérés à l'intérieur de l'échantillon sont représentatifs des logements hors champ dans la population. Lorsqu'on ignore si les logements sont admissibles (on n'a pas recueilli suffisamment de renseignements pour déterminer leur admissibilité), on réduit leur poids du taux d'unités hors champ de leur RS afin de considérer une certaine proportion de ces unités comme unités hors champ. L'échantillon obtenu est représentatif des logements ou ménages faisant partie du champ de l'enquête. Les logements admissibles conservent le même poids qu'à l'étape précédente, que l'on appelle maintenant poids A5.

Pour l'ESCC 2020, pour les périodes de collecte de septembre à décembre, les taux d'unités dans le champ de l'enquête étaient différents de ceux d'années antérieures. L'effet de la pandémie de la COVID-19, les taux de réponse plus bas et la collecte a été menée exclusivement au téléphone explique probablement ces différences. Il a été évalué que les taux obtenus ne reflétaient pas la réalité. Les taux d'unités dans le champ de l'enquête de 2019 ont été utilisés pour la pondération des unités avec un statut inconnu pour les périodes de septembre à décembre 2020.

A6 – Non-réponse ménage

Lors de la collecte, une certaine proportion des ménages échantillonnés aboutit inévitablement à une non-réponse. Cela survient habituellement lorsque le ménage refuse de participer à l'enquête, qu'il fournit des données inutilisables, ou qu'il ne peut être joint pour une interview. Les poids des ménages non répondants sont réaffectés aux ménages répondants à l'intérieur des groupes de réponses homogènes (GRH), formés à l'intérieur de la province. Pour créer des GRH, on utilise une méthode de cotation basée sur une régression logistique permettant de déterminer la probabilité de réponse, puis ces probabilités sont utilisées afin de diviser l'échantillon en groupes aux propriétés de réponse similaires. Les renseignements obtenus relativement aux non-répondants sont limités de sorte que les modèles de régression utilisent des caractéristiques comme le période de collecte, les données géographiques et les variables de la base des enquêtes-ménages de Statistique Canada, ainsi que des paradonnées ou des données de traitement, notamment le nombre de tentatives de contact, l'heure/le jour de la tentative et des indications précisant si le ménage a été appelé un jour de fin de semaine ou un jour de semaine. Des modèles distincts sont créés pour les logements joints par téléphone et ceux joints en personne. Le facteur d'ajustement calculé à l'intérieur de chaque GRH a été obtenu de la façon suivante :

$$\frac{\text{Somme des poids A5 relatifs à l'ensemble des ménages}}{\text{Sommes des poids A5 relatifs à tous les ménages répondants}}$$

Le poids A5 est multiplié par ce facteur d'ajustement pour l'obtention du poids A6 des ménages répondants. Les ménages non répondants sont éliminés du processus à ce stade.

Pour les périodes de collecte de septembre à décembre 2020, les périodes de collecte ont été plus courtes que les périodes régulières de trois mois. Les taux de réponses ont également été plus bas que ce qui était initialement prévu. Moins de GRH ont pu être créés et la correction du biais dû à la non-réponse n'est pas probablement pas aussi efficace que pour des cycles précédents de l'ESCC.

8.3 Pondération de l'échantillon de la base de l'ECSA

S0 – Poids initial

La sélection des logements occupés par des aînés pour le suréchantillon de l'ECSA est faite à partir de la base de l'ECSA dans chaque province.

La probabilité de sélection des logements dans chaque province correspond au ratio du nombre de logements occupés par des aînés au nombre total de logements occupés par des aînés dans la province dans la base. Le ratio repose sur l'ensemble des unités disponibles dans la base et sur le nombre d'unités sélectionnées pour la période de collecte en question. La probabilité de sélection peut par conséquent varier selon la répartition de l'échantillon et les mises à jour de la base. L'inverse de cette probabilité représente le poids initial S0.

S1 – Nombre de périodes de collecte

Des échantillons sont prélevés pour chaque période de collecte dans la base du suréchantillon de l'ECSA. À chacune des unités échantillonnées correspond un poids initial faisant en sorte que chaque échantillon est représentatif de la population échantillonnée à l'échelle provinciale. Toutefois, pour que l'échantillon total ne représente qu'une seule fois la population, il faut appliquer un facteur d'ajustement qui réduira les poids des unités dans chaque échantillon. Ce facteur d'ajustement est égal à l'inverse du nombre d'échantillons combinés (c.-à-d. le nombre de périodes de collecte, 4). Les poids initiaux (S0) sont multipliés par ce facteur d'ajustement (1/4) pour l'obtention du poids S1. Suivant cet ajustement, l'échantillon complet correspond à la population moyenne dans la base de l'ECSA pour l'ensemble des périodes de collecte.

S5 – Retrait des unités hors champ

Parmi tous les logements échantillonnés, une certaine proportion de ceux-ci est, lors de la collecte, identifiée comme unités hors du champ de l'enquête. Les établissements et les logements qui sont saisonniers ou secondaires constituent des exemples de cas hors du champ de l'ECSA. Les logements dans lesquels tous les résidents sont hors du champ de l'ECSA, par exemple, les logements ne comptant pas d'aîné ou dans lesquels tous les aînés sont des membres à temps plein des Forces armées ou des personnes en visite au Canada, sont également des logements hors champ. Comme pour la base aréolaire, les logements hors champ et leurs poids sont retirés de l'échantillon; les poids des logements dont l'admissibilité est inconnue sont réduits du taux provincial d'unités hors champ; et les logements faisant partie du champ de l'enquête conservent le même poids qu'à l'étape précédente. Suivant cette étape, le poids résultant, S5, est représentatif des logements ou des ménages faisant partie du champ de l'enquête.

S6 – Non-réponse au niveau du ménage

L'ajustement appliqué ici pour compenser l'effet de la non-réponse au niveau du ménage est identique à celui appliqué dans le cas des unités de la base aréolaire (ajustement A6). Le facteur d'ajustement calculé à l'intérieur de chaque groupe de réponses homogènes est obtenu comme suit :

$$\frac{\text{Somme des poids S5 relatifs à l'ensemble des ménages}}{\text{Sommes des poids S5 relatifs à tous les ménages répondants}}$$

Le poids S5 des ménages répondants est multiplié par ce facteur d'ajustement pour l'obtention du poids S6 des ménages répondants. Les ménages non répondants sont éliminés du processus à ce stade.

Pour les périodes de collecte de septembre à décembre 2020, les périodes de collecte ont été plus courtes que les périodes régulières de trois mois. Les taux de réponses ont également été plus bas que ce qui était initialement prévu. Moins de GRH ont pu être créés et la correction du biais dû à la non-réponse n'est pas probablement pas aussi efficace que pour des cycles précédents de l'ESCC.

8.4 Ajustement des poids en fonction de la combinaison de la base aréolaire et de la base de l'ECSA

I1 – Intégration

Cette étape consiste à intégrer les poids des ménages âgés communs à la base aréolaire et à la base de l'ECSA en appliquant une méthode d'intégration¹¹. Le poids de deux types de ménages figurant dans la base intégrée n'est pas ajusté : les ménages qui ne sont pas communs aux deux bases de sondage et les ménages dans la base aréolaire ne comptant pas d'ainé. Pour toutes les autres unités, un facteur d'ajustement provincial, α_p , se situant entre 0 et 1 est appliqué aux poids. Pour chaque province (p), α_p est déterminé de manière à représenter l'importance relative de chaque échantillon (échantillon de la base aréolaire ou suréchantillon de l'ECSA) dans l'échantillon provincial total intégré. Le terme « importance » se rapporte dans ce cas-ci à la taille d'échantillon effective estimée de chaque base de sondage, soit le ratio de la taille d'échantillon et de l'effet de plan estimé. Le poids des unités de la base aréolaire est multiplié par ce facteur α_p , tandis que le poids des unités de la base de l'ECSA est multiplié par $1 - \alpha_p$. Il importe de souligner que, dans les provinces uniquement couvertes par la base aréolaire (Ontario et Québec), l'ajustement est égal à 1 pour tous les ménages. Le produit entre l'ajustement obtenu à cette étape et le poids final au niveau du ménage calculé auparavant (A6 ou S6, selon la base de sondage dont fait partie l'unité) donne le poids au niveau du ménage intégré I1.

¹¹ Skinner, C.J. and Rao, J.N.K. 1996. Estimation in Dual Frame Surveys with Complex Designs. *Journal of the American Statistical Association*. 91, 433, 349-356.

I3 – Sélection des personnes

Puisque l'unité d'échantillonnage finale est la personne, les poids au niveau du ménage calculés jusqu'ici doivent être convertis en poids au niveau de la personne. Le poids au niveau de la personne est obtenu en multipliant le poids au niveau du ménage intégré par l'inverse de la probabilité de sélection de la personne sélectionnée dans le ménage. Nous obtenons ainsi le poids I3. La probabilité de sélection d'une personne vivant dans un logement sélectionné à partir de la base aréolaire varie selon le nombre de personnes au sein du ménage et selon leur âge (voir la sous-section 5.6 pour plus de détails). Dans la base de l'ECSA, tous les aînés faisant partie du champ de l'enquête au sein d'un ménage ont une probabilité égale de sélection; cette probabilité est égale à l'inverse du nombre d'aînés faisant partie du champ de l'enquête au sein dudit ménage (voir la sous-section 5.8.3).

Pour l'ESCC 2020 pour les périodes de collecte de septembre à décembre 2020, il a été observé que la proportion de répondants dans des ménages d'une seule personne était plus grande comparativement aux années antérieures. Ceci est probablement causé par une période de collecte plus courte. Dans les cas de ménages avec plus d'une personne, si la personne sélectionnée n'était pas la personne avec qui le contact a été initialement fait, les chances de pouvoir recontacter la personne sélectionnée étaient plus faibles. Veuillez noter que cette surreprésentation des ménages avec une seule personne pourrait introduire un biais potentiel dans les estimations de variables d'enquêtes corrélées avec la taille du ménage.

I4 – Non-réponse au niveau de la personne

L'interview de l'ESCC auprès des logements de la base intégrée des adultes peut être considérée comme un processus en deux volets. Dans un premier temps, l'intervieweur obtient la liste complète des membres du ménage, puis il procède à l'interview de la personne sélectionnée dans le ménage. Dans certains cas, les intervieweurs ne réussissent qu'à réaliser la première étape, soit parce qu'ils ne peuvent pas entrer en contact avec la personne sélectionnée, ou encore parce que la personne sélectionnée refuse d'être interviewée. De tels cas sont définis comme des non-réponses au niveau de la personne, et un facteur d'ajustement doit être appliqué aux poids des personnes répondantes pour compenser cette non-réponse. Comme lors du traitement de la non-réponse au niveau du ménage, l'ajustement est appliqué à l'intérieur de groupes de réponses homogènes. Dans le cadre de ce processus, la méthode de cotation permet de définir une probabilité de réponse à partir des caractéristiques des répondants et des non-répondants accessibles. Toutes les caractéristiques recueillies lors du listage des membres du ménage, en plus de l'information géographique et de parodontées, sont accessibles aux fins d'estimation des probabilités de réponse. Ces probabilités sont groupées sous forme de groupes de réponses homogènes et le facteur d'ajustement suivant est calculé à l'intérieur de chaque groupe :

$$\frac{\text{Somme des poids I3 relatifs à toutes les personnes sélectionnées}}{\text{Somme des poids I3 relatifs à toutes les personnes sélectionnées répondantes}}$$

Le poids I3 des personnes répondantes est multiplié par le facteur d'ajustement ci-dessus pour l'obtention du poids I4. Les personnes non répondantes sont éliminées du processus de pondération à partir de ce point.

Pour les périodes de collecte de septembre à décembre 2020, les périodes de collecte ont été plus courtes que les périodes régulières de trois mois. Les taux de réponses ont également été plus bas que ce qui était initialement prévu. Moins de GRH ont pu être créés et la correction du biais due à la non-réponse n'est pas probablement pas aussi efficace que pour des cycles précédents de l'ESCC.

8.5 Pondération de l'échantillon de la base de l'ACE

C0 – Poids initial

Le poids initial correspond à l'inverse de la probabilité de sélection. Dans le cas de la base de l'ACE, la sélection des enfants est effectuée à la partir de la base liste de l'ACE à l'intérieur de chaque strate de l'ESCC.

Dans le cas de la base de l'ACE, les enfants sont sélectionnés au hasard parmi ceux rattachés à la RS visée. La probabilité de sélection correspond au ratio entre le nombre d'unités échantillonnées et le nombre d'enfants sur la liste de la RS. Le ratio est basé sur la base disponible et le nombre d'unités sélectionnées pour la période de collecte donnée. La probabilité de sélection peut par conséquent varier selon la répartition de l'échantillon et les mises à jour de la base. L'inverse de ces probabilités représente le poids initial C0.

C1 – Nombre de périodes de collecte

Des échantillons sont sélectionnés à chaque périodes de collecte de la base de l'ACE. À chacune des unités échantillonnées correspond un poids initial faisant en sorte que chaque échantillon est représentatif de la population à l'échelle des RS. Toutefois, pour que l'échantillon total ne représente qu'une seule fois la population, il faut appliquer un facteur d'ajustement qui réduira les poids des unités dans chaque échantillon. Le facteur d'ajustement appliqué à chaque échantillon est égal à l'inverse du nombre d'échantillons combinés (c.-à-d. le nombre de périodes de collecte). À partir de ce moment, l'échantillon complet provenant de la base liste correspond à la population moyenne prise sur l'ensemble des périodes de collecte. Les poids initiaux (C0) sont multipliés par ce facteur d'ajustement pour l'obtention du poids C1.

C2 – Retrait des unités hors champ

Les enfants à temps plein dans les Forces armées canadiennes, ceux faisant partie d'une réserve des Premières Nations ou ceux qui sont décédés sont des exemples d'unités hors champ dans le cas de la base de l'ACE. Comme pour la base aréolaire, ces cas sont simplement retirés de l'échantillon, ce qui ne laisse dans l'échantillon que les enfants admissibles. Ces derniers conservent les mêmes poids qu'au cours de l'étape précédente, désormais appelés poids C2.

C3 – Ajustement lié à l’admissibilité inconnue en raison de l’âge

Certains des enfants sélectionnés ont leur anniversaire au cours d’une période de collecte. La date d’anniversaire d’une certaine proportion de ces enfants peut modifier leur admissibilité en raison de leur âge. Certains enfants sélectionnés pourraient être âgés de 11 ans au début de la période de collecte et en conséquence représenter des unités hors champ, mais devenir des unités admissibles une fois âgés de 12 ans. D’autres enfants pourraient être admissibles à 17 ans, mais devenir des unités hors champ une fois qu’ils atteignent l’âge de 18 ans. L’âge de ces enfants, et en conséquence leur admissibilité, peut être rectifié (et traité en conséquence) quand un contact est établi et que leur âge est déterminé au moment de l’établissement du contact. On ignore toutefois s’ils sont admissibles lorsqu’il est impossible de communiquer avec eux ou de leur demander leur âge. On rajuste leur poids en fonction de la probabilité modélisée de leur admissibilité si on avait communiqué avec eux et on leur avait demandé leur âge à ce moment. Le taux d’enfants admissibles qui deviennent âgés de 12 et de 18 ans parmi les cas admissibles est utilisé comme facteur d’ajustement des poids des cas dont l’admissibilité est inconnue. Les poids des enfants hors champ en raison de leur âge sont établis à zéro, tout comme ceux des autres unités hors champ; les poids rattachés aux enfants admissibles demeurent inchangés et ceux dont l’admissibilité est inconnue et dont l’anniversaire pourrait modifier l’admissibilité sont ajustés. Les poids obtenus sont les poids C3.

C6 – Non-réponse ménage

L’ajustement effectué dans ce cas pour compenser l’effet de la non-réponse ménage est identique à celui effectué dans le cas des unités de la base aréolaire jointes par téléphone (ajustement A6). Dans le cas de la base de l’ACE, la sélection a été effectuée au niveau des personnes, mais une réponse au niveau du ménage est définie lorsque le listage des membres du ménage est obtenu, comme dans le cas de la base aréolaire. Le facteur d’ajustement calculé à l’intérieur de chaque GRH a été obtenu de la façon suivante :

$$\frac{\text{Somme des poids C3 relatifs à tous les ménages}}{\text{Somme des poids C3 relatifs à tous les ménages répondants}}$$

Le poids C3 des ménages répondants est multiplié par ce facteur d’ajustement pour l’obtention du poids C6. Les ménages non répondants sont éliminés du processus à partir de ce point.

Pour les périodes de collecte de septembre à décembre 2020, les périodes de collecte ont été plus courtes que les périodes régulières de trois mois. Les taux de réponses ont également été plus bas que ce qui était initialement prévu. Moins de GRH ont pu être créés et la correction du biais dû à la non-réponse n’est pas probablement pas aussi efficace que pour des cycles précédents de l’ESCC.

C9 – Non-réponse personne

L’ajustement effectué ici pour compenser l’effet de la non-réponse au niveau de la personne est identique à celui effectué pour la base intégrée des adultes (ajustement I4). Le facteur d’ajustement calculé à l’intérieur de chaque GRH a été obtenu de la façon suivante :

Somme des poids C6 relatifs à toutes les personnes

Somme des poids C6 relatifs à toutes les personnes répondantes

Le poids C6 des personnes répondantes est multiplié par ce facteur d'ajustement pour produire le poids C9. Les personnes non répondantes sont éliminées du processus à partir de ce point.

Pour les périodes de collecte de septembre à décembre 2020, les périodes de collecte ont été plus courtes que les périodes régulières de trois mois. Les taux de réponses ont également été plus bas que ce qui était initialement prévu. Moins de GRH ont pu être créés et la correction du biais dû à la non-réponse n'est pas probablement pas aussi efficace que pour des cycles précédents de l'ESCC.

8.6 Ajustement des poids finals en fonction de la combinaison de la base aréolaire et de la base de l'ACE

À ce stade du processus, l'échantillon des deux bases peut être combiné avant d'appliquer des ajustements finaux par la winsorisation et le calage. Il n'y a pas d'ajustement supplémentaire des poids lors de la combinaison de la base intégrée des adultes et de la base de l'ACE parce qu'elles couvrent des populations mutuellement exclusives. La probabilité de sélection et la pondération subséquente d'une unité d'une base donnée ne sont aucunement affectées par la structure ou la sélection des unités de l'autre base.

A10 – Winsorisation

À la suite de la série d'ajustements appliqués aux répondants, il est possible que certaines unités se retrouvent avec des poids extrêmes comparativement aux autres unités du même domaine d'intérêt. Ces unités peuvent influencer fortement sur la variance. Pour prévenir une telle situation, on rajuste à la baisse le poids des unités aberrantes au moyen de l'approche de réduction par winsorisation.

A11 – Calage

La dernière étape nécessaire pour l'obtention du poids finaux de l'ESCC est le calage (A11). Le calage est effectué au moyen du système d'estimation généralisé de Statistique Canada, G-EST, afin que la somme des poids finaux corresponde aux estimations de la population définies à l'échelle des RS pour l'ensemble des douze groupes d'âge et de sexe d'intérêt. Les six groupes d'âge comprennent les groupes de 12 à 17 ans, de 18 à 34 ans, de 35 à 49 ans, de 50 à 64 ans, de 65 à 74 ans et de 75 ans et plus, chez les hommes et les chez femmes. À partir de 2009, des contrôles additionnels à des niveaux géographiques plus détaillés ont été introduits pour les RS pour lesquelles de l'information supplémentaire est disponible. Une taille minimale de 20 répondants est nécessaire aux fins du calage au niveau RS-âge-sexe. Lorsque moins de 20 répondants sont présents, des regroupements sont faits à l'échelle de la province ou du sexe. Les poids sont également ajustés afin que chaque période de collecte soit également représentée par l'échantillon. Il est à noter que le calage est effectué au moyen des données géographiques les plus récentes, qui peuvent différer des données géographiques utilisées lors de l'échantillonnage.

Les estimations des populations sont basées sur des projections des comptes du Recensement de 2011, de même que sur les chiffres des naissances, des décès, de l'immigration et de l'émigration depuis ce temps. La moyenne des estimations mensuelles relatives à chaque post-strate RS-âge-sexe par période de collecte a été retenue pour le calage. Le poids A10 est rajusté à l'aide de G-EST pour l'obtention du poids final A11. Ce dernier correspond au ***poids final au niveau de la personne de l'ESCC*** que l'on retrouve dans le fichier maître portant le nom de variable WTS_M.

Pour l'ESCC 2020, la collecte a été interrompue d'avril à août. Exceptionnellement, plutôt que d'effectuer un calage par période de collecte, le calage a été fait sur les périodes d'avant la pandémie (janvier à mars) et pour la période couvrant le reste de l'année (avril à décembre).

Les estimations démographiques sont des projections fondées sur les chiffres du recensement, de même que sur les chiffres des naissances, des décès, de l'immigration et de l'émigration depuis ce temps. En 2020, l'année de base du recensement utilisée pour établir ces projections démographiques a changé; les projections de 2019 sont fondées sur le Recensement de 2011, et les projections de 2020 sont fondées sur le Recensement de 2016. La moyenne des estimations mensuelles pour chacune des poststrates RS-âge-sexe par période de collecte a été retenue pour réaliser le calage. Le poids A10 est ajusté au moyen de G-EST afin d'obtenir le poids final A11. Le poids A11 correspond au ***poids-personne final de l'ESCC*** que l'on retrouve dans le fichier maître portant le nom de variable WTS_M.

8.7 Création d'un poids partagé

En plus du fichier maître et du FMGD qui contiennent tous les répondants de l'ESCC, un fichier partagé renfermant seulement une partie (habituellement 90 %) des répondants originaux de l'ESCC est créé. Les individus dans le fichier partagé ont consenti à partager leurs données à certains partenaires. Pour compenser la perte de certains répondants du fichier, les poids des répondants qui ont consenti à partager leurs données doivent être ajustés à l'aide du facteur :

$$\frac{\text{Somme des poids I4 (ou C9) relatifs à tous les répondants}}{\text{Somme des poids I4 (ou C9) relatifs à tous les répondants ayant accepté de partager leurs données}}$$

À l'instar des ajustements liés à la non-réponse, ce facteur est calculé à l'intérieur de groupes de réponses homogènes regroupant, dans ce cas, les ménages ayant une propension à partager jugée semblable. Cet ajustement de partage est calculé séparément pour les répondants de la base intégrée des adultes et ceux de la base de l'ACE. Le poids I4 (ou C9 dans le cas des répondants de la base de l'ACE) des répondants qui ont consenti à partager leurs données est multiplié par le facteur d'ajustement ci-dessus pour l'obtention d'un poids partagé. Comme on le fait dans le cadre du processus de pondération du fichier maître, la base intégrée des adultes et la base de l'ACE sont combinées à ce stade et on applique la winsorisation et le calage (de façon semblable aux ajustements A10 et A11) aux unités partagées. Le poids final après ces ajustements est appelé WTS_S.

8.8 Particularités de la pondération pour les trois territoires

À partir de 2015-2016, les territoires ont été échantillonnés tous les deux ans. Avant 2015, chaque strate était échantillonnée sur une base annuelle, mais afin de réduire les coûts de collecte, les strates comme les capitales ont été échantillonnées chaque année tandis que d'autres strates ont été échantillonnées seulement une fois tous les deux ans. Étant donné que les caractéristiques de la population sont assez variables d'une collectivité à l'autre, l'échantillon est maintenant seulement représentatif sur une période de deux ans.

En plus du plan de sondage sur deux ans, la base de sondage utilisée pour les trois territoires est quelque peu différente de celle utilisée pour les provinces. La stratégie de pondération est donc adaptée pour répondre à ces différences. Cette sous-section résume les changements apportés à la stratégie décrite aux sous-sections 8.1 à 8.7. Pour la base aréolaire, une étape additionnelle de sélection est ajoutée pour les territoires. Chaque territoire est stratifié selon des regroupements de collectivités à l'intérieur desquels on a sélectionné une collectivité. La capitale de chaque territoire forme une strate à elle seule et est sélectionnée automatiquement lors de cette première étape. Cette particularité a un effet dans le calcul de la probabilité de sélection, et donc dans la valeur du poids initial (A0). Le poids initial est calculé chaque année. Une fois ce poids initial calculé, la même série d'ajustements (A1 à A5) est appliquée aux unités de la base aréolaire. L'ajustement de stabilisation (A3) est généralement fait à l'échelle de la collectivité. L'ajustement lié aux unités hors champ (A5) est réalisé à l'échelle de la collectivité et par mode de collecte. De plus, étant donné que certaines strates sont échantillonnées chaque année et d'autres tous les deux ans, un ajustement visant à combiner les strates en un échantillon unique sur deux ans est effectué après l'ajustement lié aux unités hors champ. Des classes d'ajustement liées à la non-réponse à l'échelle du ménage et à l'échelle de la personne sont établies de la même façon que pour les provinces, à l'aide du même ensemble de variables en plus de la variable année (année de collecte).

Depuis la diffusion des produits pour les périodes de référence de 2008 et de 2007-2008, des contrôles ont été mis en place pour assurer que la proportion d'Autochtones et la proportion de personnes vivant dans les régions des capitales sont contrôlées dans les Territoires du Nord-Ouest et au Yukon. Un contrôle similaire basé sur le statut Inuit a été introduit pour le Nunavut à partir de 2009. Ces contrôles assurent que la proportion des estimations représentée par ces groupes est compatible avec les proportions issues du Recensement de 2011. Depuis le cycle de 2019-2020, ces proportions ont été calculées au moyen du Recensement de 2016.

Avant 2013, l'Enquête sur la santé dans les collectivités canadiennes (ESCC) couvrait seulement les 10 plus grandes collectivités du Nunavut. Les chiffres de population utilisés pour le calage ont été ajustés pour tenir en compte ce sous-dénombrement. À partir de 2013, l'ESCC a augmenté sa couverture afin qu'elle corresponde à celle de l'Enquête sur la population active, dans laquelle 93 % de la population est couverte. Par conséquent, les chiffres de population utilisés pour le calage sont fondés sur la population totale et ne sont plus ajustés pour ce sous-dénombrement.

8.9 Modification de la pondération dans les trois territoires en raison de la COVID-19

La pandémie de COVID-19 a entraîné l'interruption de la collecte de données par des intervieweurs sur place (IPAO) en mars 2020. Dans les régions éloignées du Nord, l'IPAO est le seul mode de

collecte disponible et, en fait, la seule collecte réalisée dans les territoires au cours des trois derniers trimestres de 2020 a été effectuée dans les capitales territoriales (Whitehorse, Yellowknife et Iqaluit). Puisque chaque communauté, à l'exception des capitales, n'est visitée qu'au cours d'une période du cycle de deux ans, chaque territoire n'a pas été entièrement couvert. Les communautés qui devaient faire l'objet d'une collecte entre avril et décembre 2020 représentaient 8 % de la population du Yukon, 25 % de la population des Territoires du Nord-Ouest (T.N.-O.) et 34 % de la population du Nunavut, ce qui représente 22 % de l'ensemble des territoires.

À la suite de ces arrêts de collecte, plusieurs changements ont été apportés au processus de pondération pour les territoires, comme il est précisé à la section 8.8.

- a) Aucune paradedonnée n'était disponible pour les unités recueillies par interview en personne dans les territoires en 2020, y compris celles recueillies de janvier à mars 2020. Par conséquent, aucune modélisation de la non-réponse au niveau des ménages ou des personnes n'a été effectuée pour ces unités, et l'ajustement pour la non-réponse a été calculé au moyen d'un simple ratio de pondération dans chaque territoire.
- b) Les poids bootstrap n'ont pas pu être coordonnés avec les cycles précédents. Étant donné que certaines communautés étaient entièrement absentes du dossier, il a fallu recréer les poids bootstrap pour les villes restantes. Cela a une incidence sur les calculs de la variance pour les estimations regroupées obtenues en combinant ce fichier aux cycles précédents. L'écart tel que calculé sera plus élevé qu'il ne l'aurait été si les poids bootstrap étaient restés coordonnés.
- c) Pour tenir compte de la différence entre la région échantillonnée et la région recueillie, le calage a été effectué au moyen de projections de population obtenues uniquement pour la région collectée. Les projections initiales de population ont été ajustées à la baisse en fonction de la proportion de la population vivant dans les communautés recueillies. Par conséquent, le poids des répondants du Nord s'élève à seulement 78 616 au lieu des 100 519 de la population cible prévue du Nord. Il n'est donc pas recommandé d'extrapoler à partir d'analyses effectuées sur les unités recueillies en vue de tirer des conclusions sur chaque territoire dans son ensemble.
- d) Il a également été nécessaire de modifier certaines des exigences de calage. Comme il est mentionné à l'article 8.8, l'étape de calage contrôle la proportion d'Autochtones vivant au Yukon et dans les Territoires du Nord-Ouest. Ce processus exige des projections fiables de la population autochtone âgée de 12 ans et plus vivant dans ces territoires. Toutefois, comme la collecte de 2019-2020 ne couvrait pas chaque territoire, ces projections n'ont pas pu être utilisées. Au lieu de cela, les projections pour les régions recueillies ont dû être dérivées à partir d'autres sources de données, qui n'étaient pas disponibles, ventilées selon l'âge. Par conséquent, le calage de la population autochtone de ces territoires a été effectué en utilisant des projections pour l'ensemble de la population, au lieu de la population de 12 ans et plus. Cela aurait également été le cas au Nunavut. Toutefois, le Nunavut n'a pas pu faire l'objet de calage en fonction de la population inuite, en raison du faible nombre de répondants en 2020. Il faut faire preuve d'une extrême prudence lorsqu'on analyse ces groupes en raison d'un biais potentiel. Enfin, il convient de noter que le calage du Nord a été effectué en utilisant cinq groupes d'âge au lieu de six, comme pour les provinces. En

raison du faible nombre de répondants, il a fallu regrouper les 65 à 74 ans avec les 75 ans et plus.

8.10 Pondération d'un fichier de deux ans

Les territoires sont pondérés, comme il est décrit à la section 8.8, avec les modifications précisées à la section 8.9. Ils sont pondérés pour être représentatifs de la période de deux ans. Les provinces ont été pondérées sur une période d'une année. Quand deux années de données sont combinées pour créer un fichier de deux ans, on calcule de nouveaux poids en divisant directement les poids annuels en deux. L'opération nous assure que la somme totale des poids équivaut à la taille moyenne de la population pour la période de deux ans. Dans certains cas, l'ajustement est un peu plus complexe lorsque les groupes avaient été agrégés dans le calage pour les fichiers couvrant une période d'une année. Pour plus de renseignements sur la façon de combiner plusieurs années, consulter l'article « Combiner les cycles de l'Enquête sur la santé dans les collectivités canadiennes » paru dans les Rapports sur la santé (82-003) de Statistique Canada au lien suivant :

<https://www150.statcan.gc.ca/n1/pub/82-003-x/2009001/article/10795-fra.pdf>

Il convient également de noter que le calage pendant la période de collecte a été effectué différemment selon l'année et la région. Le calage pour 2019 a été effectué pour quatre périodes de collecte de 3 mois, tandis que pour 2020, il a été effectué en deux groupes : avant la COVID-19 et pendant la COVID-19. Pour le Nord, le calage a été effectué au moyen de quatre périodes de collecte de 3 mois, qui regroupent les unités recueillies au cours de la même période, mais au cours d'années différentes. Ainsi, les unités recueillies au cours de la première période de collecte de 2019 ont été calées ensemble avec les unités de la première période de collecte de 2020.

Il faut aussi noter que l'Enquête sur la santé dans les collectivités canadiennes (ESCC) n'avait pas d'échantillon dans les territoires. De plus, aucun rajustement supplémentaire n'a été nécessaire pour tenir compte de l'ESCC pour les provinces.

Il est à noter qu'il n'est pas recommandé de combiner des données des cycles de l'ESCC d'avant et d'après le remaniement (p. ex. combiner les fichiers annuels de 2014 et de 2015) en raison des nombreux changements apportés au contenu et à la méthodologie.

9. QUALITÉ DES DONNÉES

9.1 Taux de réponse

Au total, 264 677 des unités sélectionnées à l'ESCC 2019-2020 et à l'ECSA 2019-2020 faisaient partie du champ de l'enquête¹². Parmi ces unités, 108 604 ont accepté de participer à l'enquête, ce qui résulte en un taux de réponse de 41,0 %. L'annexe E présente les dénombrements des unités comprises dans le champ de l'enquête, les dénombrements des répondants et les taux de réponse correspondants, par province/territoire. Dans le tableau 9.1 (Annexe E), ces chiffres sont présentés séparément pour les adultes et pour les jeunes, ainsi que pour l'échantillon complet de l'ESCC. Les chiffres pour l'ECSA sont présentés dans le tableau 9.2 (Annexe E).

9.2 Erreurs dans les enquêtes

L'enquête permet de produire des estimations fondées sur l'information recueillie à partir d'un échantillon de personnes. On aurait pu obtenir des estimations quelque peu différentes si on avait effectué un recensement complet en utilisant le même questionnaire, les mêmes intervieweurs, les mêmes superviseurs, les mêmes méthodes de traitement, etc. que ceux utilisés pour l'enquête. La différence entre les estimations tirées de l'échantillon et celles qui découlent d'un dénombrement complet effectué dans des conditions semblables s'appelle l'erreur due à l'échantillonnage des estimations.

Les erreurs qui ne sont pas liées à l'échantillonnage peuvent être commises à presque toutes les étapes d'une enquête. Il est possible que les intervieweurs comprennent mal les instructions, que les répondants fassent des erreurs en complétant le questionnaire, que les réponses soient mal saisies et que des erreurs se produisent au moment du traitement et de la totalisation des données. Tous ces exemples représentent des erreurs non dues à l'échantillonnage.

9.2.1 Erreurs non dues à l'échantillonnage

Sur un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations tirées de l'enquête. Toutefois, les erreurs qui se produisent systématiquement contribueront à des biais dans les estimations de l'enquête. On a consacré beaucoup de temps et d'efforts à réduire les erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été appliquées à chaque étape du cycle de collecte et de traitement des données afin de contrôler la qualité des données. On a notamment fait appel à des intervieweurs hautement qualifiés, une formation poussée sur les méthodes d'enquête et le questionnaire et l'observation des intervieweurs afin de déceler les problèmes. La mise à l'essai de l'application IAO et les essais sur le terrain ont également été au nombre des procédures essentielles pour réduire au maximum les erreurs de collecte de données.

¹² Parmi les unités sélectionnées, certaines n'entrent pas dans le champ de l'enquête. Il s'agit, par exemple, de logements sélectionnés parmi la base aréolaire, mais qui se sont révélés inoccupés, ont été démolis ou ne sont pas des logements résidentiels, ou de jeunes qui ont quitté le Canada. Ces unités ont été déterminées pendant la collecte des données, autrement, elles auraient été exclues avant la sélection de l'échantillon. Ces unités ne sont pas prises en compte dans le calcul des taux de réponse. Ce taux de réponse comprend les trois territoires.

L'effet de la *non-réponse* sur les résultats de l'enquête constitue une source importante d'erreurs non dues à l'échantillonnage dans les enquêtes. L'ampleur de la non-réponse varie de la non-réponse partielle (le fait de ne pas répondre à une ou plusieurs questions) à la non-réponse totale. Dans l'ESCC, il y a peu de non-réponse partielle car une fois le questionnaire débuté les répondants avaient tendance à le terminer. Il y a eu non-réponse totale lorsque la personne sélectionnée pour participer à l'enquête a refusé de le faire ou que l'intervieweur a été incapable d'entrer en contact avec elle. Les taux de réponse ont été plus bas pour l'ESCC 2020 que pour un cycle régulier de l'ESCC, étant donné l'impact de la pandémie de COVID-19 sur les activités de collecte. Pour les unités qui n'ont pas de numéro de téléphone disponible sur la base de sondage, les taux de contact étaient très bas puisque les visites en personne par un interviewer n'étaient pas permises. Pour les unités qui pouvaient être contactées par téléphone, avoir des périodes de collecte plus courtes et une capacité réduite pour les activités de collecte ont été les principales causes de la baisse des taux de réponse. Les cas de non-réponse totale ont été considérés lors de la pondération en corrigeant les poids des personnes qui ont répondu à l'enquête afin de compenser pour ceux qui n'ont pas répondu. Voir la section 8 pour avoir de plus amples détails sur la correction de la pondération pour la non-réponse.

Les utilisateurs sont priés de porter une attention particulière à la note sur la qualité des données de l'ESCC 2019-2020 dans l'encadré de la page 4 de ce document.

9.2.2 Erreurs dues à l'échantillonnage

Étant donné que les estimations d'une enquête par sondage comportent inévitablement des erreurs dues à l'échantillonnage, de bonnes méthodes statistiques exigent que les chercheurs fournissent aux utilisateurs une certaine indication de l'ampleur de cette erreur. La mesure de l'importance éventuelle des erreurs dues à l'échantillonnage est fondée sur l'écart type des estimations tirées des résultats de l'enquête. Cependant, en raison de la grande diversité des estimations que l'on peut tirer d'une enquête, l'écart type d'une estimation est habituellement exprimé en fonction de l'estimation à laquelle il se rapporte. Cette mesure, appelée coefficient de variation (CV), s'obtient en divisant l'écart type de l'estimation par l'estimation elle-même et on l'exprime en pourcentage de l'estimation.

Par exemple, supposons qu'une personne estime que 25 % des Canadiens âgés de 12 ans et plus sont des fumeurs réguliers et que cette estimation comporte un écart type de 0,003. On calcule alors le CV de cette estimation de la façon suivante :

$$(0,003/0,25) \times 100 \% = 1,20 \%$$

Statistique Canada utilise fréquemment les résultats du CV pour l'analyse des données et conseille vivement aux utilisateurs produisant des estimations à partir des fichiers de données de l'ESCC de faire de même. Pour plus d'information sur le calcul des CV, voir la section 11. Pour consulter les lignes directrices sur la façon d'interpréter les résultats du CV, se référer au tableau à la fin de la sous-section 10.4.

10. LIGNES DIRECTRICES POUR LA TOTALISATION, L'ANALYSE ET LA DIFFUSION

La présente section de la documentation comprend des lignes directrices destinées aux utilisateurs, pour la totalisation, l'analyse, la publication ou d'autres formes de diffusion des estimations tirées des fichiers d'enquête. Grâce à ces lignes directrices, les utilisateurs de microdonnées devraient pouvoir produire des chiffres semblables à ceux produits par Statistique Canada. Ils pourront aussi élaborer des chiffres non publiés actuellement en conformité avec les lignes directrices établies. Des méthodes visant à mesurer la précision et la qualité sont aussi décrites en parallèle avec les lignes directrices de diffusion afin de permettre de décider quand une estimation devrait être utilisée dans une publication.

10.1 Lignes directrices pour l'arrondissement

Afin que les estimations calculées d'après ces fichiers de microdonnées (fichier maître, fichier de partage, ou FMGD) correspondent à celles produites par Statistique Canada, il est vivement conseillé à l'utilisateur de les arrondir en se conformant aux lignes directrices suivantes.

- a) Les estimations qui figurent dans le corps d'un tableau statistique doivent être arrondies à la centaine près par la méthode d'arrondissement classique. Selon cette méthode, si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre retenu ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente d'une unité (1) la valeur du dernier chiffre retenu. Par exemple, si l'on veut arrondir à la centaine près de la façon classique une estimation dont les deux derniers chiffres sont compris entre 00 et 49, il faut les remplacer par 00 et ne pas modifier le chiffre précédent (le chiffre des centaines). Si les deux derniers chiffres sont compris entre 50 et 99, il faut les remplacer par 00 et augmenter d'une unité (1) le chiffre précédent.
- b) Les totaux partiels de marge et les totaux de marge des tableaux statistiques doivent être calculés à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la centaine près selon la méthode d'arrondissement classique¹³.
- c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir d'éléments non arrondis (c'est-à-dire les numérateurs et (ou) dénominateurs), puis arrondis à une décimale par la méthode d'arrondissement classique. Si l'on veut arrondir une estimation à un seul chiffre décimal par cette méthode et que le dernier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à retenir ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente d'une unité (1) le dernier chiffre à retenir.

¹³ Les tableaux de l'ESCC depuis 2015 et des années suivantes utilisent un algorithme qui arrondit les totaux, puis qui résout les incohérences à l'intérieur du tableau pour que les totaux correspondent entre eux. Cette méthode d'arrondissement fait en sorte que les tableaux produits au moyen des microdonnées de l'ESCC pourraient ne pas fournir les mêmes totaux que ceux figurant dans les tableaux de données.

- d) Les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs éléments correspondants non arrondis, puis arrondies à leur tour à la centaine près (ou à la décimale près) selon la méthode d'arrondissement classique.
- e) Si, en raison de contraintes d'ordre technique ou autre, on applique une autre méthode que l'arrondissement classique, si bien que les estimations qui seront publiées ou diffusées de toute autre façon diffèrent des estimations correspondantes publiées par Statistique Canada, il est vivement conseillé à l'utilisateur d'indiquer la raison de ces divergences dans le ou les documents à publier ou à diffuser.
- f) Des estimations non arrondies ne doivent être publiées ou diffusées de toute autre façon en aucune circonstance. Des estimations non arrondies donnent l'impression d'être beaucoup plus précises qu'elles ne le sont en réalité.

10.2 Lignes directrices pour la pondération de l'échantillon en vue de la totalisation

Le plan d'échantillonnage utilisé pour cette enquête n'est pas autopondéré. Autrement dit, le poids d'échantillonnage n'est pas le même pour toutes les personnes qui font partie de l'échantillon. Même pour produire des estimations simples, y compris des tableaux statistiques ordinaires, l'utilisateur doit employer le poids d'échantillonnage approprié. Sinon, les estimations calculées à partir du fichier de microdonnées ne pourront être considérées comme représentatives de la population observée et ne correspondront pas à celles de Statistique Canada.

L'utilisateur ne doit pas non plus perdre de vue qu'en raison du traitement réservé au champ du poids, certains logiciels ne permettent pas d'obtenir des estimations qui coïncident exactement avec celles de Statistique Canada. Si des options sont disponibles, l'utilisateur devrait veiller à préciser que le poids est un poids d'échantillon plutôt qu'un poids de fréquence.

10.2.1 Définitions des catégories d'estimations : de type nominal par opposition à quantitatives

Avant d'exposer la façon de totaliser et d'analyser les données de l'enquête, il est bon de décrire les deux grandes catégories d'estimations ponctuelles des caractéristiques de la population qui peuvent être produites à partir du fichier de microdonnées.

Estimations de type nominal :

Les estimations de type nominal sont des estimations du nombre ou du pourcentage de personnes qui, dans la population visée par l'enquête, possèdent certaines caractéristiques ou rentrent dans une catégorie particulière. Le nombre de personnes qui fument tous les jours est un exemple d'estimation de ce genre. L'estimation du nombre de personnes qui possèdent une caractéristique particulière peut aussi être appelée « estimation d'un agrégat ».

Exemple de question de type nominal :

Actuellement, est-ce que ... fume(z) des cigarettes tous les jours, à l'occasion ou jamais?
(SMK_005)

- ___ Tous les jours
- ___ À l'occasion
- ___ Jamais

Estimations quantitatives :

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes ou d'autres mesures de tendance centrale de quantités qui ont trait à tous les membres de la population observée ou à certains d'entre eux.

Un exemple d'estimation quantitative est le nombre moyen de cigarettes que fument par jour les personnes qui fument tous les jours. Le numérateur correspond à l'estimation du nombre total de cigarettes que fument par jour les personnes qui fument tous les jours et le dénominateur, à l'estimation du nombre de personnes qui fument tous les jours.

Exemple de question quantitative :

Actuellement, combien de cigarettes fume(z)-vous/t-il/t-elle chaque jour? (SMK_045)

□□□ Nombre de cigarettes

10.2.2 Totalisation d'estimations de type nominal

On peut obtenir, à partir du fichier de microdonnées, des estimations du nombre de personnes qui possèdent une caractéristique donnée en additionnant les poids finaux de tous les enregistrements contenant des données sur la caractéristique étudiée.

Pour obtenir les proportions et les rapports de la forme \hat{X} / \hat{Y} on doit :

- a) additionner les poids finaux des enregistrements contenant la caractéristique voulue pour le numérateur (\hat{X});
- b) additionner les poids finaux des enregistrements contenant la caractéristique voulue pour le dénominateur (\hat{Y});
- c) diviser l'estimation du numérateur par celle du dénominateur.

10.2.3 Totalisation d'estimations quantitatives

Pour obtenir l'estimation d'une somme ou d'une moyenne pour une variable quantitative, on procède aux étapes suivantes (seule l'étape a) est nécessaire pour obtenir l'estimation pour une somme) :

- a) multiplier la valeur de la variable étudiée par le poids finaux puis faire la somme de cette quantité pour tous les enregistrements visés pour obtenir le numérateur (\hat{X});
- b) faire la somme des poids finaux des enregistrements contenant la variable étudiée pour obtenir le dénominateur (\hat{Y});
- c) diviser l'estimation du numérateur par l'estimation du dénominateur.

Par exemple, pour estimer le nombre moyen de cigarettes que fument chaque jour les personnes qui fument tous les jours, on calcule d'abord le numérateur (\hat{X}) en additionnant le produit entre la

valeur de la variable **SMK_045** et le poids **WTS_M**. Ensuite additionnez cette valeur pour les enregistrements pour lesquels la valeur de la variable **SMK_005** est « tous les jours ». On obtient ensuite le dénominateur (\hat{Y}) en additionnant le poids final de tous les enregistrements pour lesquels la valeur de la variable **SMK_005** est « tous les jours ». Le nombre moyen de cigarettes fumées chaque jour par les personnes qui fument tous les jours est finalement obtenu en divisant (\hat{X}) par (\hat{Y}).

10.3 Lignes directrices pour l'analyse statistique

L'ESCC se fonde sur un plan de sondage complexe qui prévoit une stratification et un échantillonnage à plusieurs degrés, ainsi que la sélection des répondants avec probabilités inégales. L'utilisation des données provenant d'une enquête aussi complexe pose des difficultés aux analystes, car le choix des méthodes d'estimation et de calcul de la variance dépend du plan de sondage et des probabilités de sélection. L'utilisation des poids de sondage représente une première étape en vue de l'obtention de résultats appropriés.

Nombre de méthodes d'analyse intégrées aux progiciels statistiques permettent d'utiliser des poids, mais la signification et la définition de ces poids peuvent différer de celles applicables dans le contexte d'une enquête par sondage. Cela a pour résultat final que, même si dans plusieurs cas les estimations calculées au moyen des progiciels sont exactes, les variances et les statistiques fondées sur ces variances n'ont pratiquement aucune signification. Pour obtenir les résultats appropriés, les poids doivent être interprétés correctement par le progiciel, et on doit avoir recours à des techniques d'estimation de la variance appropriées.

Il existe plusieurs options pour assurer l'utilité des résultats obtenus au moyen de ce plan de sondage complexe. La première option et la plus pertinente consiste à utiliser les options de calcul de la variance énoncées au chapitre 11, fondées sur la méthode *bootstrap*. En deuxième lieu, il existe dans de nombreux progiciels des procédures propres aux enquêtes qui permettent d'interpréter le poids correctement. Par exemple, dans le SAS, PROC SURVEYMEANS fournit des résultats plus pertinents que les résultats produits à partir de PROC MEANS, car les poids *bootstrap* sont utilisés pour les estimations de la variance. PROC MEANS peut utiliser les poids de sondage pour redresser correctement les estimations, en tenant compte du plan de pondération inégal, mais tend à sous-estimer les variances en n'utilisant pas les poids *bootstrap* et en supposant que le plan de sondage prévoit un échantillon aléatoire simple. Veuillez consulter le document sur l'estimation de la variance *bootstrap* dans la documentation sur la diffusion de l'ESCC pour de plus amples renseignements. Il est recommandé aux utilisateurs de données d'utiliser des progiciels et des fonctions pouvant incorporer à la fois les poids d'échantillonnage et les poids *bootstrap* durant la réalisation de leurs analyses.

Toutefois, seules les options énoncées au chapitre 11 fondées sur la méthode *bootstrap* tiendront compte de la stratification, de la mise en grappes et du plan à bases de sondage multiples au moment du calcul de la variance.

10.4 Lignes directrices pour la diffusion

Étant donné que les estimations obtenues à partir de l'enquête sont fondées sur un échantillon, il existe une variabilité dans les valeurs obtenues, dans le sens où un échantillon différent pourrait donner lieu à des résultats différents. Pour tenir compte de cela, les utilisateurs doivent d'abord s'assurer qu'il y a suffisamment d'observations pour estimer de façon appropriée la statistique, ainsi que la variance. Une fois la variance obtenue, les utilisateurs devraient veiller à ce qu'elle soit suffisamment raisonnable pour que l'estimation puisse être interprétée de façon appropriée comme représentant à peu près la véritable valeur de population.

Fichier maître ou de partage

Dans le cas des utilisateurs de fichiers maîtres ou de partage, il est recommandé d'avoir au moins 10 observations comportant la caractéristique d'intérêt et 20 dans le domaine pour le calcul d'une proportion. Lorsqu'il y a suffisamment d'observations, l'utilisateur peut calculer la variance et les coefficients de variation au moyen des poids bootstrap fournis avec les données, ainsi que du logiciel approprié pour effectuer l'analyse. Les variances peuvent servir au calcul des C.V., ce qui aide à évaluer la fiabilité d'une estimation en fonction de la variabilité de l'échantillonnage. Le C.V. fournit une mesure relative de l'erreur d'échantillonnage proportionnellement à l'estimation. Les estimations devraient être vérifiées suivant les lignes directrices du tableau 10.1. Les C.V. augmenteront au fur et à mesure que croît la variabilité d'une estimation et ils diminueront au fur et à mesure qu'une estimation devient plus précise. Une estimation peut toutefois être précise tout en présentant un C.V. important, en particulier si l'estimation a une valeur proche de zéro. L'examen de l'intervalle de confiance de l'estimation fournira une autre indication de la qualité de l'estimation du point de vue de la variabilité. Les longs intervalles de confiance signalent une précision inférieure de l'estimation, tandis que les intervalles de confiance plus courts signalent une précision accrue. Il faut tenir compte des intervalles de confiance des estimations lors de l'évaluation de la fiabilité des proportions d'un échantillon.

Tableau 10.1 Lignes directrices relatives à la variabilité d'échantillonnage

Type d'estimation	CV (en %)	Lignes directrices
Acceptable	$CV \leq 15,0$	On peut envisager une diffusion générale non restreinte des estimations. Aucune annotation particulière n'est nécessaire.
Marginale	$15,0 < CV \leq 35,0$	On peut envisager une diffusion générale non restreinte des estimations, en y joignant une mise en garde aux utilisateurs quant à la variabilité d'échantillonnage élevée liée aux estimations. Les estimations de ce genre doivent être identifiées par la lettre E (ou d'une autre manière similaire).
Inacceptable	$CV > 35,0$	Statistique Canada recommande de ne pas publier des estimations dont la qualité est inacceptable. Toutefois, si l'utilisateur choisit de le faire, il doit alors adjoindre la lettre F (ou un autre identificateur semblable) et les diffuser avec l'avertissement suivant : « Nous avisons l'utilisateur que ... (précisez les données)... ne répondent pas aux normes de qualité de Statistique Canada pour ce programme statistique. Les conclusions tirées de ces données ne sauraient être fiables et seront fort probablement erronées. Ces données et toute conclusion qu'on pourrait en tirer ne doivent pas être publiées. Si l'utilisateur choisit de les publier, il est alors tenu de publier également le présent avertissement. »

11. TABLEAUX DE LA VARIABILITÉ D'ÉCHANTILLONNAGE APPROXIMATIVE

Afin de fournir une approximation rapide des coefficients de variation applicables à un large éventail d'estimations catégoriques tirées du fichier partagé, un ensemble de tables de variabilité d'échantillonnage approximative a été produit. Ces tableaux permettent aux utilisateurs d'obtenir un coefficient de variation approximatif selon la taille de l'estimation calculée à partir des données de l'enquête. Le tableau 11.1 présente un tableau de consultation de l'ESCC à l'échelle du Canada en exemple.

Table 11.1, ESCC Tableaux de la variabilité d'échantillonnage approximative, 2017

Partagé															
NUMÉRATEUR DU POURCENTAGE (''000)	POURCENTAGE ESTIMÉ														
	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%	
1	126.8	126.2	125.5	123.6	120.3	116.9	113.4	109.8	106.1	102.2	98.2	89.7	69.5	40.1	
2	89.6	89.2	88.8	87.4	85.1	82.7	80.2	77.7	75.0	72.3	69.5	63.4	49.1	28.4	
3	73.2	72.9	72.5	71.4	69.5	67.5	65.5	63.4	61.3	59.0	56.7	51.8	40.1	23.2	
4	63.4	63.1	62.8	61.8	60.2	58.5	56.7	54.9	53.1	51.1	49.1	44.8	34.7	20.1	
5	56.7	56.4	56.1	55.3	53.8	52.3	50.7	49.1	47.5	45.7	43.9	40.1	31.1	17.9	
6	51.7	51.5	51.3	50.5	49.1	47.7	46.3	44.8	43.3	41.7	40.1	36.6	28.4	16.4	
7	47.9	47.7	47.5	46.7	45.5	44.2	42.9	41.5	40.1	38.6	37.1	33.9	26.3	15.2	
8	44.8	44.6	44.4	43.7	42.5	41.3	40.1	38.8	37.5	36.1	34.7	31.7	24.6	14.2	
9	42.3	42.1	41.8	41.2	40.1	39.0	37.8	36.6	35.4	34.1	32.7	29.9	23.2	13.4	
10	40.1	39.9	39.7	39.1	38.0	37.0	35.9	34.7	33.6	32.3	31.1	28.4	22.0	12.7	
11	38.2	38.0	37.9	37.3	36.3	35.3	34.2	33.1	32.0	30.8	29.6	27.0	20.9	12.1	
12	36.6	36.4	36.2	35.7	34.7	33.8	32.7	31.7	30.6	29.5	28.4	25.9	20.1	11.6	
13	35.2	35.0	34.8	34.3	33.4	32.4	31.5	30.5	29.4	28.4	27.2	24.9	19.3	11.1	
14	33.9	33.7	33.6	33.0	32.2	31.2	30.3	29.4	28.4	27.3	26.3	24.0	18.6	10.7	
15	32.7	32.6	32.4	31.9	31.1	30.2	29.3	28.4	27.4	26.4	25.4	23.2	17.9	10.4	
16	31.7	31.5	31.4	30.9	30.1	29.2	28.4	27.5	26.5	25.6	24.6	22.4	17.4	10.0	
17	30.7	30.6	30.4	30.0	29.2	28.4	27.5	26.6	25.7	24.8	23.8	21.7	16.8	9.7	
18	29.9	29.7	29.6	29.1	28.4	27.6	26.7	25.9	25.0	24.1	23.2	21.1	16.4	9.5	
19	29.1	28.9	28.8	28.4	27.6	26.8	26.0	25.2	24.3	23.5	22.5	20.6	15.9	9.2	
20	28.3	28.2	28.1	27.6	26.9	26.1	25.4	24.6	23.7	22.9	22.0	20.1	15.5	9.0	
21	27.7	27.5	27.4	27.0	26.3	25.5	24.8	24.0	23.2	22.3	21.4	19.6	15.2	8.8	
22	27.0	26.9	26.8	26.4	25.7	24.9	24.2	23.4	22.6	21.8	20.9	19.1	14.8	8.6	
23	26.4	26.3	26.2	25.8	25.1	24.4	23.7	22.9	22.1	21.3	20.5	18.7	14.5	8.4	
24	25.9	25.8	25.6	25.2	24.6	23.9	23.2	22.4	21.7	20.9	20.1	18.3	14.2	8.2	
25	25.4	25.2	25.1	24.7	24.1	23.4	22.7	22.0	21.2	20.4	19.6	17.9	13.9	8.0	
30	23.1	23.0	22.9	22.6	22.0	21.3	20.7	20.1	19.4	18.7	17.9	16.4	12.7	7.3	
35	21.4	21.3	21.2	20.9	20.3	19.8	19.2	18.6	17.9	17.3	16.6	15.2	11.7	6.8	
40	20.0	20.0	19.9	19.5	19.0	18.5	17.9	17.4	16.8	16.2	15.5	14.2	11.0	6.3	
45	18.9	18.8	18.7	18.4	17.9	17.4	16.9	16.4	15.8	15.2	14.6	13.4	10.4	6.0	
50	17.9	17.8	17.8	17.5	17.0	16.5	16.0	15.5	15.0	14.5	13.9	12.7	9.8	5.7	
55	17.1	17.0	16.9	16.7	16.2	15.8	15.3	14.8	14.3	13.8	13.2	12.1	9.4	5.4	
60	16.4	16.3	16.2	16.0	15.5	15.1	14.6	14.2	13.7	13.2	12.7	11.6	9.0	5.2	
65	15.7	15.7	15.6	15.3	14.9	14.5	14.1	13.6	13.2	12.7	12.2	11.1	8.6	5.0	
70	15.2	15.1	15.0	14.8	14.4	14.0	13.6	13.1	12.7	12.2	11.7	10.7	8.3	4.8	
75	14.6	14.6	14.5	14.3	13.9	13.5	13.1	12.7	12.3	11.8	11.3	10.4	8.0	4.6	
80	14.2	14.1	14.0	13.8	13.5	13.1	12.7	12.3	11.9	11.4	11.0	10.0	7.8	4.5	
85	13.7	13.7	13.6	13.4	13.0	12.7	12.3	11.9	11.5	11.1	10.7	9.7	7.5	4.3	
90	13.4	13.3	13.2	13.0	12.7	12.3	12.0	11.6	11.2	10.8	10.4	9.5	7.3	4.2	
95	13.0	12.9	12.9	12.7	12.3	12.0	11.6	11.3	10.9	10.5	10.1	9.2	7.1	4.1	
100	12.7	12.6	12.6	12.4	12.0	11.7	11.3	11.0	10.6	10.2	9.8	9.0	6.9	4.0	
125	11.3	11.3	11.2	11.1	10.8	10.5	10.1	9.8	9.5	9.1	8.8	8.0	6.2	3.6	
150	10.3	10.3	10.3	10.1	9.8	9.5	9.3	9.0	8.7	8.3	8.0	7.3	5.7	3.3	
200	9.0	8.9	8.9	8.7	8.5	8.3	8.0	7.8	7.5	7.2	6.9	6.3	4.9	2.8	
250	8.0	8.0	7.9	7.8	7.6	7.4	7.2	6.9	6.7	6.5	6.2	5.7	4.4	2.5	
300	7.3	7.3	7.2	7.1	6.9	6.8	6.5	6.3	6.1	5.9	5.7	5.2	4.0	2.3	
350	*****	6.7	6.7	6.6	6.4	6.2	6.1	5.9	5.7	5.5	5.3	4.8	3.7	2.1	
400	*****	6.3	6.3	6.2	6.0	5.8	5.7	5.5	5.3	5.1	4.9	4.5	3.5	2.0	
450	*****	5.9	5.9	5.8	5.7	5.5	5.3	5.2	5.0	4.8	4.6	4.2	3.3	1.9	
500	*****	5.6	5.6	5.5	5.4	5.2	5.1	4.9	4.7	4.6	4.4	4.0	3.1	1.8	
750	*****	*****	4.6	4.5	4.4	4.3	4.1	4.0	3.9	3.7	3.6	3.3	2.5	1.5	
1000	*****	*****	4.0	3.9	3.8	3.7	3.6	3.5	3.4	3.2	3.1	2.8	2.2	1.3	
1500	*****	*****	3.2	3.2	3.1	3.0	2.9	2.8	2.7	2.6	2.5	2.3	1.8	1.0	
2000	*****	*****	*****	2.8	2.7	2.6	2.5	2.5	2.4	2.3	2.2	2.0	1.6	0.9	
3000	*****	*****	*****	2.3	2.2	2.1	2.1	2.0	1.9	1.9	1.8	1.6	1.3	0.7	
4000	*****	*****	*****	1.9	1.8	1.8	1.7	1.7	1.6	1.6	1.6	1.4	1.1	0.6	
5000	*****	*****	*****	*****	1.7	1.6	1.6	1.5	1.4	1.4	1.4	1.3	1.0	0.6	
6000	*****	*****	*****	*****	1.5	1.5	1.4	1.4	1.3	1.3	1.3	1.2	0.9	0.5	
7000	*****	*****	*****	*****	*****	1.4	1.3	1.3	1.2	1.2	1.2	1.1	0.8	0.5	
8000	*****	*****	*****	*****	*****	*****	1.2	1.2	1.1	1.1	1.1	1.0	0.8	0.4	
9000	*****	*****	*****	*****	*****	*****	1.2	1.1	1.1	1.1	1.0	0.9	0.7	0.4	
10000	*****	*****	*****	*****	*****	*****	*****	1.1	1.0	1.0	1.0	0.9	0.7	0.4	
12500	*****	*****	*****	*****	*****	*****	*****	*****	0.9	0.8	0.8	0.6	0.4	0.4	
15000	*****	*****	*****	*****	*****	*****	*****	*****	*****	0.8	0.7	0.6	0.3	0.3	

Les CV sont calculés en employant la formule de la variance utilisée pour l'échantillonnage aléatoire simple et en y incorporant un facteur qui reflète la structure en grappes à plusieurs degrés du plan d'échantillonnage. Pour obtenir ce facteur, appelé effet de plan, on a d'abord calculé les effets de plan pour une vaste gamme de caractéristiques, puis pour chaque tableau, choisi une valeur conservatrice parmi tous les effets de plan relatifs à ce tableau. Cette valeur choisie a ensuite été utilisée pour générer le tableau qui peut alors s'appliquer à l'ensemble complet des caractéristiques.

Les tableaux de la variabilité d'échantillonnage approximative, de même que les effets de plan, les tailles d'échantillon et les comptes de population qui ont servi à les produire, sont disponibles pour les utilisateurs du fichier partagé dans le document Tableaux de la variabilité d'échantillonnage approximative. Tous les coefficients de variation sont approximatifs dans les tableaux de la variabilité d'échantillonnage approximative et ils ne doivent donc pas être considérés comme des valeurs exactes. Les possibilités concernant le calcul d'un coefficient de variation exact sont discutées dans la sous-section 11.7. Une analyse fondée sur l'utilisation de poids bootstrap est la méthode qui devrait être utilisée pour obtenir des résultats plus précis.

Rappel : Tel qu'indiqué dans « Les lignes directrices relatives à la variabilité d'échantillonnage » à la section 10.4, si le nombre d'observations sur lesquelles une estimation est basée est inférieur à 10, l'estimation pondérée ne doit pas être diffusée, quelle que soit la valeur de son coefficient de variation. Les coefficients de variation basés sur des échantillons de petite taille sont trop imprévisibles pour être adéquatement représentés dans les tableaux. L'utilisation des tableaux pour estimer les coefficients de variation produit des résultats encore moins précis que l'utilisation de poids bootstrap de sorte que, dans ce cas, un minimum de 30 observations est requis.

11.1 Comment utiliser les tableaux de CV pour les estimations de type nominal

Les règles suivantes devraient permettre à l'utilisateur de calculer, à partir des tableaux de la variabilité d'échantillonnage, les coefficients de variation approximatifs d'estimations relatives au nombre, à la proportion ou au pourcentage de personnes dans la population observée qui possèdent une caractéristique donnée ainsi que des rapports et des écarts entre ces estimations.

Règle 1 : Estimations du nombre de personnes possédant une caractéristique donnée (agrégats)

Le coefficient de variation dépend uniquement de la taille de l'estimation elle-même. Dans le tableau de coefficients de variation approximatifs correspondant à la région appropriée, il faut repérer l'estimation calculée dans la colonne d'extrême gauche (intitulée « Numérateur du pourcentage ») et suivre les astérisques (s'il y en a) de gauche à droite jusqu'au premier nombre. Puisque toutes les valeurs possibles de l'estimation ne sont pas disponibles, il faut prendre la valeur la plus petite qui s'en rapproche le plus (par exemple, si l'estimation vaut 1 700 et que les deux valeurs disponibles sont 1 000 et 2 000, il faut choisir 1 000). Ce nombre constitue le coefficient de variation approximatif pour l'estimation en question.

Règle 2 : Estimations de proportions ou de pourcentages de personnes possédant une caractéristique donnée

Le coefficient de variation d'une proportion (ou d'un pourcentage) estimée dépend à la fois de l'ordre de grandeur de cette proportion et de l'ordre de grandeur du numérateur utilisé dans le calcul de la proportion. Les proportions estimées sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion lorsque celle-ci est fondée sur un sous-ensemble de la population. Cela est dû au fait que les coefficients de variation des estimations du dernier type sont basés sur le chiffre le plus élevé dans une rangée d'un tableau particulier, tandis que les coefficients de variation des estimations du premier type sont basés sur un chiffre quelconque de cette même rangée (pas nécessairement le plus élevé). (Il convient de noter que dans les tableaux, la valeur des coefficients de variation décroît de gauche à droite sur une même ligne.) Par exemple, la proportion estimée de personnes qui fument tous les jours parmi les fumeurs est plus fiable que le nombre estimé de personnes qui fument tous les jours.

Lorsque la proportion (ou le pourcentage) est fondée sur la population totale de la région géographique à laquelle le tableau s'applique, le coefficient de variation de la proportion est égal à celui du numérateur de la proportion. Dans ce cas-ci, cela équivaut à appliquer la règle 1.

Lorsque la proportion (ou le pourcentage) est fondée sur un sous-ensemble de la population totale (par exemple, les personnes qui fument), il faut se reporter à la proportion (haut du tableau) et au numérateur de la proportion ou du pourcentage (côté gauche du tableau). Puisque toutes les valeurs possibles de la proportion et du numérateur ne sont pas disponibles, il faut, dans les deux cas, prendre la valeur la plus petite qui s'en rapproche le plus (par exemple, si la proportion est de 23 % et que les deux valeurs disponibles dans la colonne s'en rapprochant le plus sont 20 % et 25 %, il faut choisir 20 %). Le coefficient de variation se trouve à l'intersection de la ligne et de la colonne appropriée.

Règle 3 : Estimations des différences entre des agrégats ou des pourcentages

L'erreur-type d'une différence entre deux estimations est à peu près égale à la racine carrée de la somme des carrés de chaque erreur-type considérée séparément. L'erreur-type d'une différence ($\hat{d} = \hat{X}_2 - \hat{X}_1$) est donc :

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1 \alpha_1)^2 + (\hat{X}_2 \alpha_2)^2}$$

où \hat{X}_1 représente l'estimation 1, \hat{X}_2 l'estimation 2, et α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement. Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d}$. Cette formule donne un résultat exact pour ce qui est de la différence entre des sous-populations indépendantes mais n'est autrement qu'approximative. Cette formule mènera à une surestimation de l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélés positivement et à une sous-estimation de l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélés négativement.

Règle 4 : Estimations de rapports

Si le numérateur est un sous-ensemble du dénominateur, il faut convertir le rapport en pourcentage et appliquer la règle 2. Ce serait le cas, par exemple, si le dénominateur est le nombre de personnes

qui fument et le numérateur est le nombre de personnes qui fument tous les jours parmi celles qui fument.

Si le numérateur n'est pas un sous-ensemble du dénominateur (par exemple, le rapport du nombre de personnes qui fument tous les jours ou à l'occasion au nombre de personnes qui ne fument pas du tout), l'erreur-type du rapport entre les estimations est à peu près égal à la racine carrée de la somme des carrés de chaque coefficient de variation pris séparément multipliée par \hat{R} , où \hat{R} est le rapport des estimations ($\hat{R} = \hat{X}_1 / \hat{X}_2$). L'erreur-type d'un rapport est donc :

$$\sigma_{\hat{R}} = \hat{R} \sqrt{\alpha_1^2 + \alpha_2^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement.

Le coefficient de variation de \hat{R} est donné par $\sigma_{\hat{R}} / \hat{R} = \sqrt{\alpha_1^2 + \alpha_2^2}$. La formule tend à surestimer l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélés positivement et à sous-estimer l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélés négativement.

Règle 5 : Estimations des différences entre des rapports

Dans ce cas-ci, les règles 3 et 4 sont combinées. On commence par calculer les coefficients de variation des deux rapports au moyen de la règle 4, puis le coefficient de variation de leur différence au moyen de la règle 3.

11.2 Exemples d'utilisation des tableaux de CV pour des estimations de type nominal

Les exemples réels suivants ont pour but d'aider les utilisateurs à appliquer les règles décrites ci-dessus. Les chiffres non-arrondis sont utilisés dans les exemples pour faciliter les directives étape par étape. Ces exemples utilisent les tableaux de la variabilité d'échantillonnage approximative de l'ESCC 2017 (Tableau 11.1). Les utilisateurs devraient toujours suivre les lignes directrices sur l'arrondissement tel que mentionné dans la section 10.1 avant de publier des chiffres.

Exemple 1 : Estimations du nombre de personnes possédant une caractéristique donnée (agrégats)

Supposons qu'un utilisateur estime à 3 665 449 le nombre de personnes qui fument tous les jours au Canada. Comment l'utilisateur fait-il pour déterminer le coefficient de variation de cette estimation?

- 1) Se reporter au tableau de CV pour le CANADA.
- 2) L'agrégat estimé (3 665 449) ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le nombre le plus petit qui s'en rapproche le plus, soit 3 000 000.
- 3) Le coefficient de variation d'un agrégat estimé (exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 2,3 %.

- 4) Le coefficient de variation approximatif de l'estimation est donc 2,3 %. Par conséquent, selon les « Lignes directrices relatives à la variabilité d'échantillonnage » présentées à la section 10.4, l'estimation selon laquelle 3 665 449 personnes fument tous les jours peut être diffusée sans réserve après avoir mis-en-œuvre les règles d'arrondissement.

Exemple 2 : Estimations de proportions ou de pourcentages de personnes possédant une caractéristique donnée

Supposons qu'un utilisateur estime à $3\,665\,449/5\,151\,237=71,2\%$ le pourcentage de personnes, parmi les fumeurs, qui fument tous les jours au Canada. Comment l'utilisateur fait-il pour déterminer le coefficient de variation de cette estimation?

- 1) Se reporter au tableau de CV pour le CANADA.
- 2) Parce que l'estimation est un pourcentage basé sur un sous-ensemble de la population totale (c.-à-d. les personnes qui fument tous les jours ou à l'occasion), il faut utiliser à la fois le pourcentage (71,2%) et la partie numérateur du pourcentage (3 665 449) pour déterminer le coefficient de variation.
- 3) Le numérateur (3 665 449) ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le nombre le plus petit qui s'en rapproche le plus, soit 3 000 000. De même, l'estimation du pourcentage ne figure pas parmi les en-têtes de colonnes; il faut donc utiliser le nombre le plus petit qui s'en rapproche le plus, soit 70,0 %.
- 4) Le nombre qui se trouve à l'intersection de la ligne et de la colonne utilisées, soit 1,3 %, est le coefficient de variation (exprimé en pourcentage) à utiliser.
- 5) Le coefficient de variation de l'estimation est donc 1,3 %. Par conséquent, selon les « Lignes directrices relatives à la variabilité d'échantillonnage » présentées à la section 10.4, l'estimation selon laquelle 71,2 % des gens qui fument le font tous les jours peut être diffusée sans réserve.

Exemple 3 : Estimations des différences entre des agrégats ou des pourcentages

Supposons qu'un utilisateur estime que, parmi les hommes, $2\,067\,955/15\,115\,880 = 13,7\%$ fument tous les jours (estimation 1), alors que chez les femmes, ce pourcentage est estimé à $1\,597\,494/15\,528\,483 = 10,3\%$ (estimation 2). Comment l'utilisateur fait-il pour déterminer le coefficient de variation de la différence entre ces deux estimations?

- 1) À l'aide du tableau de CV pour le CANADA, utilisé de la même façon que dans l'exemple 2, vous établissez à 2,7 % le CV de l'estimation 1 (exprimé en pourcentage) et à 3,1 % le CV de l'estimation 2 (exprimé en pourcentage).
- 2) Selon la règle 3, l'erreur-type pour une différence ($\hat{d} = \hat{X}_2 - \hat{X}_1$) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1 \alpha_1)^2 + (\hat{X}_2 \alpha_2)^2}$$

où \hat{X}_1 représente l'estimation 1, \hat{X}_2 l'estimation 2, et α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement. L'erreur-type de la différence $\hat{d} = (0,137 - 0,103) = 0,034$ est donc :

$$\sigma_{\hat{d}} = \sqrt{[(0,137)(0,027)]^2 + [(0,103)(0,031)]^2} \\ = 0,00489$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,00489 / 0,034 = 0,144$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 14,4 % (exprimé en pourcentage). Par conséquent, toujours selon les « Lignes directrices relatives à la variabilité d'échantillonnage » présentées à la section 10.4, cette estimation peut être publiée sans réserve.

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime à 3 665 449 le nombre de personnes qui fument tous les jours et à 1 485 788 le nombre de celles qui fument à l'occasion. L'utilisateur veut comparer ces deux estimations sous la forme d'un rapport. Comment fait-il pour déterminer le coefficient de variation de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation de rapport, où le numérateur de l'estimation ($= \hat{X}_1$) est le nombre de personnes qui fument à l'occasion. Le dénominateur de l'estimation ($= \hat{X}_2$) est le nombre de personnes qui fument tous les jours.
- 2) Se reporter au tableau de CV pour le CANADA.
- 3) Le numérateur de cette estimation de rapport est 1 485 788. Le nombre le plus petit qui se rapproche le plus de ce nombre est 1 000 000. Le coefficient de variation de cette estimation (exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 4,0 %.
- 4) Le dénominateur de cette estimation de rapport 3 665 449. Le nombre le plus petit qui se rapproche le plus de ce nombre est 3 000 000. Le coefficient de variation de cette estimation (exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 2,3 %.
- 5) Le coefficient de variation approximatif de l'estimation du rapport est donc donné par la règle 4 :

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^2 + \alpha_2^2},$$

c'est-à-dire,

$$\alpha_{\hat{R}} = \sqrt{(0,040)^2 + (0,023)^2} \\ = 0,046$$

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement. Le rapport des personnes qui fument occasionnellement à celles qui fument tous les jours est 1 485 788/3 665 449, soit 0,41:1. Le coefficient de variation de cette estimation est 4,6 % (exprimé en

pourcentage); selon les « Lignes directrices relatives à la variabilité d'échantillonnage » présentées à la section 10.4, l'estimation peut donc être diffusée sans réserve.

11.3 Comment utiliser les tableaux de CV pour calculer les limites de confiance

Bien que les coefficients de variation soient largement utilisés, l'intervalle de confiance d'une estimation représente une mesure plus intuitive de l'erreur d'échantillonnage. Un intervalle de confiance est une façon d'énoncer la probabilité que la valeur vraie de la population se situe dans une plage de valeurs données. Par exemple, un intervalle de confiance de 95 % peut être décrit comme suit : si l'échantillonnage de la population se répète à l'infini, chacun des échantillons donnant un nouvel intervalle de confiance pour une estimation, l'intervalle contiendra la valeur vraie de la population dans 95 % des cas.

Une fois déterminée l'erreur-type d'une estimation, on peut calculer des intervalles de confiance pour les estimations en partant de l'hypothèse qu'en procédant à un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique de la population sont réparties selon une distribution normale autour de la valeur vraie de la population. Selon cette hypothèse, il y a environ 68 chances sur 100 que l'écart entre une estimation de l'échantillon et la valeur vraie de la population soit inférieur à une erreur-type, environ 95 chances sur 100 que l'écart soit inférieur à deux erreurs-types et environ 99 chances sur 100 que l'écart soit inférieur à trois erreurs-types. On appelle ces différents degrés de confiance des niveaux de confiance.

L'intervalle de confiance (CI) d'une estimation \hat{X} est généralement exprimé sous la forme de deux nombres, l'un étant inférieur à l'estimation et l'autre supérieur à celle-ci, sous la forme $(\hat{X} - k, \hat{X} + k)$, où k varie selon le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

On peut calculer directement les intervalles de confiance d'une estimation à partir des tableaux de la variabilité d'échantillonnage approximative, en trouvant d'abord dans le tableau approprié le coefficient de variation de l'estimation \hat{x} , puis en utilisant la formule suivante pour obtenir l'intervalle de confiance CI correspondant :

$$CI_X = [\hat{X} - z \hat{X} \alpha_{\hat{X}}, \hat{X} + z \hat{X} \alpha_{\hat{X}}]$$

où $\alpha_{\hat{X}}$ est le coefficient de variation trouvé pour \hat{X} , et

- $z = 1$ si l'on désire un intervalle de confiance de 68 %
- $z = 1,6$ si l'on désire un intervalle de confiance de 90 %
- $z = 2$ si l'on désire un intervalle de confiance de 95 %
- $z = 3$ si l'on désire un intervalle de confiance de 99 %

Note : Les lignes directrices concernant la diffusion des estimations de la section 10.4 s'appliquent aussi aux intervalles de confiance. Par conséquent, si l'estimation ne peut être diffusée, alors l'intervalle de confiance ne peut l'être lui non plus.

11.4 Exemple d'utilisation de tableaux de CV pour obtenir des limites de confiance

Voici la marche à suivre pour calculer un intervalle de confiance de 95 % pour la proportion estimée de personnes qui fument tous les jours parmi celles qui fument (d'après l'exemple 2 de la sous-section 11.2).

$$\hat{X} = 0.712$$

$$z = 2$$

$$\alpha_{\hat{X}} = 0,013 \text{ est le coefficient de variation de cette estimation selon les tableaux.}$$

$$CI_{\hat{X}} = \{0.712 - (2) (0.712) (0.013), 0.712 + (2) (0.712) (0.013)\}$$

$$CI_{\hat{X}} = \{0.693, 0.731\}$$

11.5 Comment utiliser les tableaux de CV pour effectuer un test Z

On peut aussi utiliser les erreurs-types pour effectuer des tests d'hypothèses, une technique qui permet de faire la distinction entre les paramètres d'une population à l'aide d'estimations basées sur un échantillon. Ces estimations peuvent être des nombres, des moyennes, des pourcentages, des rapports, etc. Les tests peuvent être effectués à divers niveaux de signification; un niveau de signification est la probabilité de conclure que les caractéristiques sont différentes quand, en fait, elles sont identiques.

Supposons que \hat{X}_1 et \hat{X}_2 sont des estimations basées sur un échantillon pour deux caractéristiques voulues. Supposons aussi que l'erreur-type de la différence $\hat{X}_1 - \hat{X}_2$ est $\sigma_{\hat{d}}$. Si le ration de $\hat{X}_1 - \hat{X}_2$ sur $\sigma_{\hat{d}}$ est compris entre -2 et 2, alors on ne peut tirer aucune conclusion à propos de la différence entre les caractéristiques au niveau de signification de 5 %. Toutefois, si ce rapport est inférieur à -2 ou supérieur à +2, la différence observée est significative au niveau de 5%.

11.6 Exemple d'utilisation des tableaux de CV pour effectuer un test Z

Supposons que nous voulons tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y a pas de différence entre la proportion d'hommes qui fument tous les jours et cette même proportion chez les femmes. Dans l'exemple 3 de la sous-section 11.2, nous avons déterminé que l'erreur-type de la différence entre ces deux estimations est égale à 0,00495. Par conséquent,

$$z = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0,137 - 0,103}{0,0048995} = \frac{0,034}{0,0048995} = 6,87$$

Puisque $z = 6,95$ est supérieur à 2, on doit conclure qu'il existe une différence significative entre les deux estimations au niveau de signification de 5%. À noter que les deux sous-groupes comparés sont considérés comme étant indépendants, ce qui fait en sorte que le résultat du test est valide.

11.7 Variances ou coefficients de variation exacts

Tous les coefficients de variation qui figurent dans les tableaux de la variabilité d'échantillonnage approximative (tableaux de CV) sont effectivement approximatifs, donc, non officiels.

Le calcul de variance ou coefficient de variation exact n'est pas chose évidente puisqu'il n'existe pas de formule mathématique simple pouvant prendre en compte tous les aspects du plan d'échantillonnage et de la pondération de l'ESCC. On doit donc avoir recours à d'autres méthodes pour estimer ces mesures de précision, telles que des méthodes par rééchantillonnage. Parmi celles-ci, la méthode du bootstrap est celle recommandée pour l'analyse des données de l'ESCC.

De nombreux progiciels statistiques permettent de réaliser des analyses à l'aide de poids bootstrap. De plus amples renseignements sont présentés à la section 10.3 et dans le document sur l'estimation de la variance bootstrap.

Les raisons pour lesquelles un utilisateur pourrait souhaiter connaître la précision exacte de ses estimations sont diverses. En voici quelques-unes.

Premièrement, si un utilisateur désire obtenir des estimations à un niveau géographique autre que ceux présentés dans les tableaux (par exemple, au niveau urbain ou rural), l'utilisation des tableaux de CV publiés ne convient pas parfaitement. Néanmoins, on peut obtenir les coefficients de variation de ce type d'estimations en appliquant la méthode d'estimation par domaine, au moyen du programme de calcul de la variance exacte.

Deuxièmement, si un utilisateur demande des analyses plus complexes, telles que des estimations de paramètres de modèles de régression linéaire ou logistique, les tableaux de CV ne pourront pas fournir les coefficients de variation pour ceux-ci. Certains progiciels statistiques courants permettent d'incorporer les poids d'échantillonnage aux analyses, mais, souvent, les variances produites ne tiennent pas bien compte de la stratification et de la mise en grappe de l'échantillon, contrairement à celles obtenues grâce au programme de calcul de la variance exacte.

Troisièmement, dans le cas de l'estimation de variables quantitatives, il est nécessaire d'utiliser des tableaux distincts pour déterminer l'erreur d'échantillonnage. Or, la plupart des variables de l'ESCC étant de type nominal, de tels tableaux n'ont pas été produits. Les utilisateurs qui souhaitent connaître les coefficients de variation de variables quantitatives peuvent néanmoins obtenir ces derniers grâce au programme de calcul de la variance réelle. À noter, toutefois, que le coefficient de variation d'un total quantitatif est généralement plus grand que celui de l'estimation de type nominal correspondante (c'est-à-dire, l'estimation du nombre de personnes qui contribuent à l'estimation quantitative). Si l'estimation de type nominal correspondante ne peut être diffusée, il en sera de même pour l'estimation quantitative. Par exemple, le coefficient de variation de l'estimation du nombre total de cigarettes que fument chaque jour les personnes qui fument tous les jours serait supérieur à celui de l'estimation correspondante du nombre de personnes qui fument tous les jours. Par conséquent, si on ne peut diffuser le coefficient de variation de cette dernière estimation, on ne pourra non plus diffuser celui de l'estimation quantitative correspondante.

Cette situation tient au fait que l'estimation des coefficients de variation grâce aux tableaux de la variabilité d'échantillonnage approximative est basée sur une vaste gamme de variables et est donc jugée grossière, tandis que le programme de calcul de la variance réelle produit le coefficient de variation précis associé à la variable en question.

12. FICHIERS DE MICRODONNÉES : DESCRIPTION, ACCÈS ET UTILISATION

L'ESCC produit trois types de fichiers de microdonnées : les fichiers maîtres, les fichiers de partage et les fichiers de microdonnées à grande diffusion (FMGD). Le tableau 12.1 contient la liste des fichiers disponibles pour les données de 2019-2020, ainsi que ceux pour les périodes de collecte de 2020.

12.1 Fichiers maîtres

Les fichiers maîtres contiennent toutes les variables et tous les enregistrements de l'enquête associés à une période spécifique. Ces fichiers sont accessibles à Statistique Canada pour usage interne, dans les Centres de données de recherche (CDR) de Statistique Canada et peuvent aussi faire l'objet de demandes de totalisations personnalisées.

Dans le cas de 2020, comme la collecte a seulement eu lieu dans la moitié des collectivités de chaque territoire, une seule année de données n'est pas représentative des territoires. C'est pourquoi le fichier maître de 2020 ne renfermera que les données des répondants des dix provinces. Les données recueillies en 2020 des répondants des territoires seront diffusées pour la première fois dans le fichier maître de 2019-2020.

12.1.1 Centre de données de recherche

Le Programme des CDR permet aux chercheurs d'utiliser les données d'enquête contenues dans les fichiers maîtres dans un environnement sécuritaire situé dans plusieurs universités à travers le Canada. Les chercheurs doivent soumettre des propositions de recherche qui une fois acceptées leur donneront accès aux CDR. Pour plus de renseignements, consultez la page web suivante : <https://www.statcan.gc.ca/fra/cdr/index>

12.1.2 Totalisations personnalisées

Une autre méthode d'accès aux fichiers maîtres consiste à offrir à tous les utilisateurs de faire appel au personnel du Service à la clientèle du Centre de données sur la santé de la population pour produire des totalisations personnalisées. Ce service est offert moyennant le recouvrement des coûts. Il permet aux utilisateurs qui ne savent pas se servir de logiciels de totalisation d'obtenir des résultats personnalisés. Les résultats sont filtrés pour s'assurer qu'ils sont conformes aux normes de confidentialité et de fiabilité avant d'être diffusés. Pour plus de renseignements, communiquez avec le Service à la clientèle (613) 951-1746 ou par courriel à hd-ds@statcan.gc.ca.

12.1.3 Télé-accès

Le service de télé-accès aux fichiers maîtres de l'enquête représente un autre moyen d'accéder à ces données si l'utilisateur ne peut avoir accès à un centre de données de recherche alors que l'un de ses projets de recherche est approuvé. On peut fournir au chercheur un fichier maître de données synthétiques ou fichier « fictif » et le plan des enregistrements correspondants. Il pourra grâce à ces outils mettre au point son propre ensemble de programmes informatisés d'analyse. Le code pour les totalisations personnalisées est ensuite transmis par courrier électronique à hd-ds@statcan.gc.ca. Le code sera alors transmis au réseau interne sécurisé de Statistique Canada et traité en regard du fichier maître de données pertinent de l'ESCC. Les estimations générées seront communiquées à l'utilisateur, sous réserve des directives sur l'analyse et la communication des données exposées à la section 10 du présent document. Les résultats sont filtrés pour vérifier qu'ils sont conformes aux normes de confidentialité et de fiabilité, puis les données de sortie sont renvoyées au client. Ce service est gratuit.

12.1.4 Système d'accès à distance en temps réel

Le système d'accès à distance en temps réel (ADTR) offre un accès rapide aux microdonnées de Statistique Canada. Le système d'accès à distance en temps réel l'ADTR est une installation d'accès à distance en ligne permettant aux utilisateurs d'exécuter en temps réel des programmes SAS portant sur des ensembles de microdonnées situées dans un emplacement central et sécurisé. Les chercheurs qui utilisent le système l'ADTR ne bénéficient pas d'un accès direct aux microdonnées et ne sont pas en mesure d'afficher le contenu du fichier de microdonnées. Au lieu de cela, les utilisateurs soumettent des programmes SAS pour extraire les résultats sous la forme de tableaux statistiques. Étant donné que les chercheurs qui utilisent l'ADTR ne peuvent pas visualiser les microdonnées, il n'est plus nécessaire d'obtenir le statut de personne réputée être employée de Statistique Canada. Ce lien est l'élément clé qui permet à l'ADTR d'offrir rapidement des services à ses clients. Pour plus de renseignements, regardez : <https://www.statcan.gc.ca/fra/adtr/adtr>

12.2 Fichiers de partage

Les fichiers de partage contiennent toutes les variables et tous les enregistrements des répondants de l'ESCC qui ont accepté de partager leurs données avec les partenaires de partage de Statistique Canada, soit Santé Canada et l'Agence de santé publique du Canada. Les identificateurs personnels sont retirés des fichiers de partage pour préserver la confidentialité des répondants. Les utilisateurs de ces fichiers doivent au préalable avoir porté serment qu'ils ne divulgueront en aucun temps toute information susceptible d'identifier un répondant à l'enquête.

12.3 Fichiers de microdonnées à grande diffusion

Les fichiers de microdonnées à grande diffusion (FMGD) sont élaborés à partir des fichiers maîtres suivant une technique qui vise à concilier l'impératif d'assurer la confidentialité des répondants et la nécessité de produire des données d'utilité maximale à l'échelle de la région sociosanitaire. Les FMGD doivent répondre à des normes sévères de sécurité et de confidentialité, conformément à la *Loi sur la statistique* avant qu'ils ne soient diffusés pour l'accès public. Pour s'assurer du respect de ces normes, chaque FMGD est soumis à un processus officiel d'examen et d'approbation par

un comité formé de hauts gestionnaires de Statistique Canada.

Les variables les plus susceptibles de permettre l'identification d'une personne sont supprimées du fichier ou agrégées en catégories moins détaillées. Dû aux risques de divulgation, les poids des ménages ne sont pas inclus dans le FMGD.

Le FMGD contient les données des questions posées sur une période de deux ans. À moins d'indication contraire, ces questions sont habituellement celles comprises dans le contenu de base annuel et dans le contenu thématique de deux ans, ainsi que dans le contenu optionnel, choisi pour deux ans par les provinces et les territoires.

L'accès au FMGD est gratuit dans un établissement d'enseignement postsecondaire participant à l'Initiative de démocratisation des données, un partenariat entre les établissements postsecondaire et Statistique Canada qui vise à améliorer l'accès aux sources de données canadiennes. Pour de plus amples renseignements sur l'Initiative de démocratisation des données, consulter le <https://www.statcan.gc.ca/fra/idd/idd>

Les fichiers FMGD de l'ESCC sont également gratuits pour les chercheurs communiquant avec les Services à la clientèle, au 613-951-1746 ou à l'adresse électronique hd-ds@statcan.gc.ca.

Tableau 12.1 Fichiers de données de l'ESCC 2019-2020 et 2020

Période de référence	Fichiers	Type de fichier	Nom du fichier	Poids d'échantillonnage	Fichiers des poids bootstrap	Variables incluses	Enregistrements inclus
2019-2020	Fichier principal	Maître	HS.txt	WTS_M	bsw.txt	Tous les modules de base et tous les modules optionnels / thématiques qui ont été demandés sur toute la période de deux ans	Enregistrements de tous les répondants dans les 10 provinces et 3 territoires
		Partage		WTS_S			Enregistrements de tous les répondants qui ont accepté de partager leurs données.
2020	Fichier principal	Maître	HS.txt	WTS_M	bsw.txt	Tous les modules de base et tous les modules optionnels / thématiques, incluant le nouveau module COV	Enregistrements de tous les répondants dans les 10 provinces
		Partage		WTS_S			Enregistrements de tous les répondants qui ont accepté de partager leurs données.
	Fichier de réponse rapide (SCH2/TAV)	Maître	HS_SCH2.txt	WTS_M	bsw.txt	Tous les modules de base et questions de la réponse rapide SCH2/TAV	Enregistrements de tous les répondants dans les 10 provinces (Excluant le suréchantillon)
		Partage		WTS_S			Enregistrements de tous les répondants qui ont accepté de partager leurs données (Excluant le suréchantillon)

Janvier à Mars 2020	Main file (Pre-COVID) Excluant les cas CHSS	Maître	HS.txt	WTS_M	bsw.txt	Tous les modules de base et tous les modules optionnels / thématiques	Enregistrements de tous les répondants dans les 10 provinces (Excluant les cas CHSS)
		Partage		WTS_S			Enregistrements de tous les répondants qui ont accepté de partager leurs données (Excluant les cas CHSS)
	Main file (Pre-COVID) Incluant les cas CHSS	Maître		WTS_M			Enregistrements de tous les répondants dans les 10 provinces (Incluant les cas CHSS)
		Partage		WTS_S			Enregistrements de tous les répondants qui ont accepté de partager leurs données (Incluant les cas CHSS)
	Fichier de réponse rapide (SCH2/TAV)	Maître	HS_SCH2.txt	WTS_M	bsw.txt	Tous les modules de base et questions de la réponse rapide SCH2/TAV	Enregistrements de tous les répondants dans les 10 provinces (Excluant le suréchantillon)
		Partage		WTS_S			Enregistrements de tous les répondants qui ont accepté de partager leurs données (Excluant le suréchantillon)
	Fichier de réponse rapide (HLV)	Maître	HS_HLV.txt	WTS_M	bsw.txt	Tous les modules de base et questions de la réponse rapide HLV	Enregistrements de tous les répondants dans les 10 provinces (Excluant le suréchantillon)
		Partage		WTS_S			Enregistrements de tous les répondants qui ont accepté de partager leurs données (Excluant le suréchantillon)
	Main file (During- COVID) Excluant les cas CHSS	Maître	HS.txt	WTS_M	bsw.txt	Tous les modules de base et tous les modules optionnels / thématiques, incluant le nouveau module COV	Enregistrements de tous les répondants dans les 10 provinces (Excluant les cas CHSS)
		Partage		WTS_S			Enregistrements de tous les répondants qui ont accepté de partager leurs données (Excluant les cas CHSS)
Septembre à Décembre 2020	Main file (During- COVID) Incluant les cas CHSS	Maître		WTS_M			Enregistrements de tous les répondants dans les 10 provinces (Incluant les cas CHSS)
		Partage		WTS_S			Enregistrements de tous les répondants qui ont accepté de partager leurs données (Incluant les cas CHSS)
	Fichier de réponse rapide (SCH2/TAV)	Maître	HS_SCH2.txt	WTS_M	bsw.txt	Tous les modules de base et questions de la réponse rapide SCH2/TAV	Enregistrements de tous les répondants dans les 10 provinces (Excluant le suréchantillon)
		Partage		WTS_S			Enregistrements de tous les répondants qui ont accepté de partager leurs données (Excluant le suréchantillon)

	Fichier de réponse rapide (TBI)	Maître	HS_TBI.txt	WTS_M	bsw.txt	Tous les modules de base et questions de la réponse rapide TBI	Enregistrements de tous les répondants dans les 10 provinces (Excluant le suréchantillon)
		Partage		WTS_S			Enregistrements de tous les répondants qui ont accepté de partager leurs données (Excluant le suréchantillon)

12.4 Utilisation des fichiers de l'ESCC : fichier annuel ou fichier deux ans ?

Les utilisateurs qui ont accès aux fichiers communs ou aux fichiers maîtres ont le choix d'utiliser les fichiers de données d'un an ou de deux ans. Les décisions concernant la période à utiliser pour une analyse donnée devraient être guidées par le niveau de détail et de qualité requis. Avec un fichier d'un an, il n'y aura pas toujours d'estimations en raison de la qualité associée à des échantillons de taille limitée. Comme les territoires ne figurent pas dans le fichier d'un an, les utilisateurs souhaitant de véritables estimations nationales devraient utiliser le fichier de deux ans.

Dans le cadre de l'ESCC, avant d'interpréter et d'utiliser une estimation, il est recommandé de s'assurer que cette estimation rencontre les règles suivantes :

- coefficient de variation 35,0 % ou moins ;
- minimum de 10 répondants ayant la caractéristique dans ce domaine (numérateur);
- total minimum de 20 répondants pour le domaine d'intérêt (dénominateur).

Cela ne sera pas possible dans le cas des caractéristiques rares et des domaines détaillés pour les fichiers d'un an. Les utilisateurs devront plutôt se fier aux fichiers de deux ans ou aux fichiers pluriannuels. Pour de plus amples renseignements sur la combinaison des cycles de l'ESCC à l'intérieur d'un fichier de plusieurs années, veuillez consulter le <https://www150.statcan.gc.ca/n1/pub/82-003-x/2009001/article/10795-fra.pdf>

Lorsque l'utilisation d'un fichier d'un an ou de deux ans est possible, l'utilisateur devrait envisager un compromis entre l'exactitude et l'utilisation de données en temps opportun. S'il est important de rendre compte des caractéristiques courantes d'une population le plus précisément possible, les fichiers d'un an sont préférables. Toutefois, du fait de l'augmentation de la taille de l'échantillon, des estimations et des analyses plus détaillées peuvent être effectuées avec les fichiers deux ans.

12.5 Utilisation de la variable de pondération

La variable de pondération **WTS_M** représente le poids d'échantillonnage pour les fichiers principaux de l'enquête. Pour un répondant donné, ce poids d'échantillonnage peut être interprété comme étant le nombre de personnes que le répondant représente dans la population canadienne. Ce poids doit être utilisé en tout temps dans les calculs d'estimations statistiques, afin de permettre l'inférence à l'échelle de la population. La production de résultats non pondérés n'est pas recommandée. La répartition de l'échantillon, de même que les détails du plan de sondage, peuvent entraîner des résultats biaisés qui ne représentent pas correctement la population. Pour une description plus détaillée du calcul de ce poids, consulter la section 8 sur la pondération. La variable de pondération **WTS_M** doit être utilisée pour des analyses régionales.

Le module Sécurité alimentaire (FSC) inclut dans les fichiers de certaines périodes de référence, mesure des concepts qui s'appliquent non seulement à la situation du répondant, mais à celle de l'ensemble de son ménage. L'analyse des variables peut, selon le niveau d'analyse, nécessiter le recours à un poids calculé de manière à représenter le nombre de ménages au Canada, plutôt que

le nombre de personnes. Cette variable de pondération **WTS_HH** se trouve dans un fichier distinct (HS_HHWT.txt). Elle peut être utilisée en remplacement de la variable **WTS_M** pour des analyses au niveau des ménages à l'échelle nationale et provinciale.

12.6 Convention appliquée pour nommer les variables

Les conventions appliquées pour nommer des variables permettent aux utilisateurs des données de repérer et d'utiliser facilement les variables en fonction du module et du type de variable. La convention appliquée pour nommer les variables de l'ESCC respectent deux exigences: limiter les noms des variables à huit caractères au plus pour qu'il soit facile de les utiliser avec les logiciels d'analyse et permettre de repérer facilement les variables conceptuellement identiques d'une période de collecte à l'autre de l'enquête. Les questions auxquelles on a apporté des changements entre deux périodes de collecte, qui modifient le concept mesuré par la question, sont entièrement renommées pour éviter toute confusion dans l'analyse.

Une variable (REFPER, format = AAAAMM-AAAAMM) a été ajoutée aux fichiers de microdonnées afin d'identifier le début et la fin de la période de référence au cours de laquelle les données ont été recueillies. Celle-ci s'avérera utile en particulier pour les utilisateurs désireux d'utiliser les données de plusieurs périodes de collecte à la fois. Ainsi, les noms des variables correspondant à des modules ou à des questions identiques entre une année de collecte et une autre (par ex., 2015 et 2017) seront les mêmes. Cependant, quelques changements mineurs ont été apportés aux noms des variables depuis le remaniement de 2015.

La convention appliquée pour nommer les variables à compter de l'ESCC 2015 utilise jusqu'à huit caractères. Les noms de variables sont structurés de la manière suivante :

Positions 1 à 3:	Nom du module/de la section du questionnaire
Position 4:	Type de variable (sous-tiret, C, D, F ou G)
Positions 5 à 8:	Numéro de la question et option de réponse par incréments de cinq (p. ex. 005, 010, 015)

L'exemple 1 présente la structure du nom de la variable correspondant à la question 005 du module Usage du tabac, c'est-à-dire SMK_005 :

Positions 1 à 3:	SMK	Module sur l'usage du tabac
Position 4 :	_	(sous-tiret = données recueillies)
Position 5 à 8:	005	numéro de la question

L'exemple 2 présente la structure du nom de la variable correspondant à la question 2 du module Contacts avec des professionnels de la santé - partie 1 (CHP_015) qui constitue une question à réponse multiple :

Positions 1 à 3:	CHP	Module sur l'utilisation des soins de santé
Position 4 :	_	(sous-tiret = données recueillies)

Position 5 à 8: 015 numéro de la question correspondant et option de réponse

Dans les positions 1 à 3, on retrouve l'acronyme de chacun des modules. Ces acronymes apparaissent à côté des noms de modules qui sont tous présentés dans la figure de l'annexe A. Il est à noter que certains sigles de modules sont composés de quatre caractères. Les noms des variables de ces modules sont toujours limités à un maximum de huit caractères.

La position 4 désigne le type de variable selon qu'il s'agit d'une variable collectée directement à partir d'une question du questionnaire (« _ »), d'une variable codée (« C »), dérivée (« D »), d'une variable indicatrice (« F ») ou groupée (« G »).

En général, les quatre dernières positions (5 à 8) correspondent à la numérotation de la variable qui figure sur le questionnaire. On supprime la lettre « Q » utilisée pour représenter le mot «question» et on présente tous les numéros de question au moyen d'un groupe de deux ou trois chiffres. Par exemple, la question Q005A du questionnaire devient simplement 005A et la question Q010, simplement 010.

Tableau 12.2 Désignation des codes utilisés à la 4e position du nom des variables de l'ESCC

_	Variable collectée	Variable qui figure directement sur le questionnaire
C	Variable codée	Variable codée à partir d'une ou de plusieurs variables collectées (par exemple, code de la Classification type des industries (CTI))
D	Variable dérivée	Variable calculée d'après une ou plusieurs variables collectées ou codées, ordinairement pendant le traitement au bureau central (p. ex., indice de l'état de santé)
F	Variable indicatrice	Variable calculée à partir d'une ou de plusieurs variables collectées (comme variable dérivée), mais ordinairement par l'application informatique de collecte des données, aux fins de son utilisation ultérieure durant l'interview (p. ex., indicateur de travail)
G	Variable groupée	Variables collectées, codées, supprimées ou dérivées, agrégées en un groupe (p. ex., groupes d'âge)

Parfois, certaines questions comportent plusieurs réponses alors la position finale dans la séquence du nom de la variable est représentée par une lettre. Pour ce genre de questions, de nouvelles variables sont créées dans le but de différencier un « oui » d'un « non » pour chaque possibilité de réponse. Par exemple, si la question Q010 a 4 réponses possibles, les nouvelles questions seraient Q010A pour la première possibilité, Q010B pour la deuxième, Q010C pour la troisième et ainsi de suite. Si seulement les options 2 et 3 sont choisies, alors Q010A = Non, Q010B = Oui, Q010C = Oui et Q010D = Non.

12.7 Dictionnaires de données

Des rapports distincts de dictionnaires de données, notamment les noms des variables, les concepts, les énoncés d'univers et les fréquences, sont fournis dans le fichier maître principal, les fichiers de partage et chacun des fichiers de réponse rapide (le cas échéant). Toutes les variables dans ces fichiers figureront dans le rapport du dictionnaire de données, mais certaines pourraient ne pas inclure une fréquence (SampleID par exemple).

Quand une variable inclut les fréquences des réponses, le rapport comportera quatre colonnes :

1. Contenu – Étiquette de texte de chacune des catégories de réponses.
2. Code – Valeur numérique associée à chaque catégorie.
3. Échantillon – Fréquence non pondérée du fichier des répondants avec chaque réponse.
4. Population – Fréquence pondérée de la réponse.

Dans le cas de chaque module des rapports de dictionnaire de données, un indicateur est utilisé afin d'indiquer la disponibilité d'un module pour chaque répondant. Lorsque l'indicateur est 2 (non), toutes les variables du module ont des valeurs « Sans objet ». Par exemple, la variable DOWST indique si le module Stress au travail s'applique à un répondant donné.

12.8 Différences dans le calcul des variables du contenu de base et du contenu thématique utilisant différents fichiers.

Les variables des modules de contenu de base peuvent être estimées à partir de l'un ou l'autre des deux fichiers de données lors qu'un fichier un an et un fichier deux ans sont disponibles. Selon le fichier utilisé, des différences très faibles seront observées.

Toutes les estimations officielles faites par Statistique Canada des variables des modules de contenu commun sont fondées sur le poids d'échantillonnage du fichier maître principal.

**ANNEXE A - L'ENQUETE SUR LA SANTE DANS LES COLLECTIVITES CANADIENNES (ESCC) - APERÇU
DU CONTENU – 2019-2020**

Aperçu du contenu pour les provinces et territoires – 2019-2020

Acronyme	Nom du module	2019-2020												
		TNL (10)	IPE (11)	NÉ (12)	NB (13)	QC (24)	ON (35)	MB (46)	SK (47)	AB (48)	CB (59)	YK ¹ (60)	TNO (61)	NU ¹ (62)
ADM	Renseignements administratifs	C	C	C	C	C	C	C	C	C	C	C	C	C
ALC	Consommation d'alcool	C	C	C	C	C	C	C	C	C	C	C	C	C
ALW	Consommation d'alcool au cours de la dernière semaine					O	O							O
AMU	Utilisation d'antibiotiques	O												
ANC1	Âge du répondant	C	C	C	C	C	C	C	C	C	C	C	C	C
BPC	Tension artérielle - vérification			O										
CAN	Utilisation du cannabis	C	C	C	C	C	C	C	C	C	C	C	C	C
CCC²	Problèmes de santé chroniques	C/T	C/T	C/T	C/T	C/T	C/T	C/T	C/T	C/T	C/T	C/T	C/T	C/T
CCT	Tests du cancer du côlon et du rectum									O				O
CEX	Expériences vécues dans l'enfance		O			O	O			O		T	T	T
CIH	Changements faits pour améliorer la santé		O			O						O		
CMH	Consultations au sujet de la santé mentale	O		O		O	O	O		O	O	T	T	T
CP2	Contacts avec des professionnels de la santé - bloc rallongé					O						O		O
CP3	Contacts avec des professionnels de la santé dentaire			O								C-S	C-S	C-S
DEP	Dépression							O					O	
DRG	Usage de substances							O	O	O	O			
DRV	Conduite et sécurité										O	O		
DWI	Conduite sous l'influence						O	O			O			
ETS	Exposition à la fumée secondaire				O	O	O		O		O			
FGU	Utilisation du Guide alimentaire canadien	O		O		O								
FLU	Vaccins contre la grippe	C	C	C	C	C	C	C	C	C	C	C	C	C
FSC	Sécurité alimentaire	O	O	O	O	O	O	O	O	O		O	O	O
FVC	Consommation de fruits et légumes		O	O					O	O			O	
GAM	Jeu de hasard	O												
GEN	État de santé général	C	C	C	C	C	C	C	C	C	C	C	C	C
GDR	Sexe et genre	C	C	C	C	C	C	C	C	C	C	C	C	C
GR	Interview par personne interposée	C	C	C	C	C	C	C	C	C	C	C	C	C
HMC	Services de soins de santé à domicile - avec soins palliatifs	T	T	T	T	T	T	T	T	T	T	T	T	T
HUI	Indice de l'état de santé	T	T	T	T	T	T	T	T	T	T	T	T	T
HWT	Taille et poids - autodéclarés	C	C	C	C	C	C	C	C	C	C	C	C	C

		2019-2020												
Acronyme	Nom du module	TNL (10)	IPE (11)	NÉ (12)	NB (13)	QC (24)	ON (35)	MB (46)	SK (47)	AB (48)	CB (59)	YK ¹ (60)	TNO (61)	NU ¹ (62)
INC	Revenu (sous-bloc de ADM)	C	C	C	C	C	C	C	C	C	C	C	C	C
INJ	Blessures							O						
INS	Couverture d'assurance de santé			O				O	O			T	T	T
LBF	Population active	C	C	C	C	C	C	C	C	C	C	C	C	C
MAC	Activité principale	C	C	C	C	C	C	C	C	C	C	C	C	C
MAM	Mammographie					O	O							O
MDA	Attachement au médecin	O		O	O			O						
MEX	Expériences maternelles				O						O			O
MXA	Expériences maternelles - consommation d'alcool										O			
NDE	Dépendance à la nicotine													O
OHT	Santé buccodentaire											O	O	
PAA	Activités physiques - adultes 18 ans et plus		O						O	O				
PAP	Test pap													O
PAY	Activités physiques pour les jeunes		O						O	O				
PCN2	Ordonnances - non-respect lié aux coûts - version courte											O		
PEX	Expériences comme patient			O						O		T	T	T
PHC	Soins de santé primaires	C	C	C	C	C	C	C	C	C	C	C	C	C
PMK	Personne la mieux renseignée au sujet du ménage	C	C	C	C	C	C	C	C	C	C	C	C	C
PNC	Besoins perçus de soins	T	T	T	T	T	T	T	T	T	T	T	T	T
PSC	Satisfaction des patients - soins de santé communautaires			O										
REL	Matrice des relations	C	C	C	C	C	C	C	C	C	C	C	C	C
SBE	Comportements sédentaires										O			
SCA	Méthodes pour cesser de fumer								O					O
SDC	Renseignements sociodémographiques	C	C	C	C	C	C	C	C	C	C	C	C	C
SDS	L'échelle de gravité de dépendance au cannabis	T	T	T	T	T	T	T	T	T	T	T	T	T
SMK	Usage du tabac	C	C	C	C	C	C	C	C	C	C	C	C	C
SPC	Consultations avec un professionnel de la santé au sujet du tabagisme											O		
SPI	Spirométrie											O		
SPS	Provisions sociales (réduit)	O			O		O					T	T	T

Acronyme	Nom du module	2019-2020												
		TNL (10)	IPE (11)	NÉ (12)	NB (13)	QC (24)	ON (35)	MB (46)	SK (47)	AB (48)	CB (59)	YK ¹ (60)	TNO (61)	NU ¹ (62)
STS	Sources de stress						O				O	T	T	T
SUI	Pensées suicidaires et tentatives de suicide	T	T	T	T	T	T	T	T	T	T	T	T	T
SWL	Satisfaction à l'égard de la vie	O												
SXB	Comportements sexuels	O	O				O	O			O			
TAL	Produits alternatifs du tabac					O					O			O
UCN	Besoins de santé non comblés	O		O	O	O	O	O			O			O
YAC	Accès aux cigarettes chez les jeunes												O	

Légende - Type de contenu	
C	Contenu de base
C-S	Contenu de base - sélectionné
T	Thématique
O	Optionnel
(vide)	Optionnel - Non sélectionné

- 1 Les données des répondants vivant dans les territoires sont recueillies tous les ans mais sont diffusées seulement dans les fichiers de deux ans.
- 2 En 2020, le contenu de base de CCC a été demandé dans les provinces et les territoires. Une version plus courte du contenu thématique 1 a été demandée dans les provinces et les territoires; le contenu thématique 2 n'a pas été demandé. Pour des détails, veuillez consulter le tableau CCC à l'annexe B.

**ANNEXE B - L'ENQUETE SUR LA SANTE DANS LES COLLECTIVITES
CANADIENNES (ESCC) - PROBLEMES DE SANTE CHRONIQUES – 2019-2020**

Problèmes de santé chroniques (CCC) – 2019-2020

En 2019 et 2020, le contenu de base de CCC a été demandé dans les provinces et les territoires. Une version plus courte du contenu thématique 1 a été demandée dans les provinces et les territoires; le contenu thématique 2 n'a pas été demandé.

En 2020, une nouvelle question (CCC_190) a été demandée dans les provinces et les territoires. Elle n'est pas diffusée dans les fichiers de deux ans.

Problèmes de santé chronique (CCC)		Provinces et Territoires			
Question	Concept	Type de contenu	2019	2020	2019-2020
CCC_005	Douleur articulaire – 30 j	T1	N	N	N
CCC_010	Symptômes douleurs articulaires – 3 mo	T1	N	N	N
CCC_015	Fait de l'asthme	C	Y	Y	Y
CCC_020	Asthme – a eu des symptômes ou des crises – 12 mo	C	Y	Y	Y
CCC_025	Asthme – a pris des médicaments – 12 mo	C	Y	Y	Y
CCC_030	Atteint d'une MPOC	C	Y	Y	Y
CCC_035	Fait de l'apnée du sommeil	T1	Y	Y	Y
CCC_040	Souffre d'une scoliose	T1	N	N	N
CCC_045	Souffre de fibromyalgie	T1	Y	Y	Y
CCC_050	Souffre d'arthrite (ex. arthrose, arthrite rhumatoïde, goutte)	C	Y	Y	Y
CCC_055	A des maux de dos (excluant scoliose, fibromyalgie et arthrite)	T1	N	N	N
CCC_060	Souffre d'ostéoporose	T1	Y	Y	Y
CCC_065	Fait de l'hypertension	C	Y	Y	Y
CCC_070	Hypertension – a pris des médicaments – 1 mo	C	Y	Y	Y
CCC_075	A un taux élevé de cholestérol / lipides sanguins	C	Y	Y	Y
CCC_080	Taux élevé de cholestérol / lipides – a pris des médicaments – 1 mo	C	Y	Y	Y
CCC_085	A une maladie cardiaque	C	Y	Y	Y
CCC_087	A reçu un diagnostic de maladie cardiaque - vie	C	Y	Y	Y
CCC_090	Atteint de troubles dus à un accident vasculaire cérébral	C	Y	Y	Y
CCC_095	Atteint du diabète	C	Y	Y	Y
CCC_100	Diabète – âge au premier diagnostic	C	Y	Y	Y
CCC_105	Diabète – diagnostic – durant grossesse	C	Y	Y	Y
CCC_110	Diabète – diagnostic – autre que durant la grossesse	C	Y	Y	Y
CCC_115	Diabète – temps entre diagnostic et début insuline	C	Y	Y	Y
CCC_120	Diabète – prend actuellement de l'insuline	C	Y	Y	Y
CCC_125	Diabète – pris pilules pour contrôler niveau de sucre – 1 mo	C	Y	Y	Y
CCC_130	A un cancer	C	Y	Y	Y
CCC_135	A reçu un diagnostic de cancer – vie	C	Y	Y	Y
CCC_140	A des migraines	T1	N	N	N
CCC_145	Atteint d'Alzheimer ou autre démence	C	Y	Y	Y
CCC_150	A ulcères à l'intestin ou à l'estomac	T2	N	N	N
CCC_155	A une maladie intestinale (Crohn, colite ulcéreuse, SCI, incontinence)	T2	N	N	N
CCC_160	Genre de trouble intestinaux	T2	N	N	N
CCC_165	Fait de l'incontinence urinaire	T2	N	N	N
CCC_170	A des allergies – a été informé par un professionnel de la santé	T2	N	N	N
CCC_175	Source(s) d'allergies	T2	N	N	N
CCC_180	A souffert de rhume des foins ou d'allergie nasale – vie	T2	N	N	N
CCC_185	Souffre du syndrome de fatigue chronique	T1	Y	Y	Y
CCC_190	Souffre de sensibilités aux agresseurs chimiques	T1	N	Y	N
CCC_195	Atteint trouble de l'humeur (dépression, bipolaire, manie, dysthymie)	C	Y	Y	Y
CCC_200	Atteint d'un trouble d'anxiété (phobie, TOC, panique)	C	Y	Y	Y

**ANNEXE C - GÉOGRAPHIE DISPONIBLE SUR LE FICHIER MAÎTRE ET DE
PARTAGE ET CODES CORRESPONDANTS: CANADA, PROVINCES,
RÉGIONS SOCIO-SANITAIRES ET GROUPES HOMOLOGUES**

Annexe C - Géographie disponible sur le fichier maître et de partage et codes correspondants: Canada, provinces, régions sociosanitaires et groupes homologues

0	Canada	
10	Terre-Neuve-et-Labrador	
1011-C		Eastern Regional Integrated Health Authority
1012-E		Central Regional Integrated Health Authority
1013-E		Western Regional Integrated Health Authority
1014-E		Labrador-Grenfell Regional Integrated Health Authority
11	Île-du-Prince-Édouard	
1100-C		Île-du-Prince Édouard
12	Nouvelle-Écosse	
1201-E		Zone 1 - Western
1202-C		Zone 2 - Northern
1203-E		Zone 3 - Eastern
1204-A		Zone 4 - Central
13	Nouveau-Brunswick	
1301-C		Zone 1 (région de Moncton)
1302-A		Zone 2 (région de Saint John)
1303-C		Zone 3 (région de Fredericton)
1304-E		Zone 4 (région de Edmundston)
1305-E		Zone 5 (région de Campbellton)
1306-E		Zone 6 (région de Bathurst)
1307-E		Zone 7 (région de Miramichi)
24	Québec	
2401-C		Région du Bas-Saint-Laurent
2402-C		Région du Saguenay — Lac-Saint-Jean
2403-A		Région de la Capitale-Nationale

2404-C	Région de la Mauricie et du Centre-du-Québec
2405-C	Région de l'Estrie
2406-G	Région de Montréal
2407-B	Région de l'Outaouais
2408-C	Région de l'Abitibi-Témiscamingue
2409-C	Région de la Côte-Nord
2410-C	Région du Nord-du-Québec
2411-E	Région de la Gaspésie — Îles-de-la-Madeleine
2412-D	Région de la Chaudière-Appalaches
2413-A	Région de Laval
2414-B	Région de Lanaudière
2415-B	Région des Laurentides
2416-B	Région de la Montérégie
35	Ontario par Réseau local d'intégration des services de santé
3501	Réseau d'intégration des services de santé de Erie St. Clair
3502	Réseau d'intégration des services de santé du Sud-Ouest
3503	Réseau d'intégration des services de santé de Waterloo Wellington
3504	Réseau d'intégration des services de santé de Hamilton Niagara Haldimand Brant
3505	Réseau d'intégration des services de santé du Centre-Ouest
3506	Réseau d'intégration des services de santé de Mississauga Halton
3507	Réseau d'intégration des services de santé de Toronto-Centre
3508	Réseau d'intégration des services de santé du Centre
3509	Réseau d'intégration des services de santé du Centre-Est
3510	Réseau d'intégration des services de santé du Sud-Est

3511	Réseau d'intégration des services de santé de Champlain
3512	Réseau d'intégration des services de santé de Simcoe-Nord Muskoka
3513	Réseau d'intégration des services de santé du Nord-Est
3514	Réseau d'intégration des services de santé du Nord-Ouest
35	Ontario par circonscription sanitaire
3526-E	Circonscription sanitaire du district d'Algoma
3527-A	Circonscription sanitaire du comté de Brant
3530-B	Circonscription sanitaire régionale de Durham
3533-D	Circonscription sanitaire de Grey Bruce
3534-D	Circonscription sanitaire de Haldimand-Norfolk
3535-D	Circonscription sanitaire du district de Haliburton, Kawartha et Pine Ridge
3536-B	Circonscription sanitaire régionale de Halton
3537-A	Circonscription sanitaire de la cité de Hamilton
3538-C	Circonscription sanitaire des comtés de Hastings et Prince Edward
3539-D	Circonscription sanitaire du comté de Huron
3540-E	Circonscription sanitaire de Chatham-Kent
3541-C	Circonscription sanitaire de Kingston, Frontenac et Lennox et Addington
3542-C	Circonscription sanitaire de Lambton
3543-D	Circonscription sanitaire de Leeds, Grenville et Lanark
3544-A	Circonscription sanitaire de Middlesex-London
3546-A	Circonscription sanitaire régionale de Niagara
3547-C	Circonscription sanitaire du district de North Bay Parry Sound
3549-C	Circonscription sanitaire du Nord-Ouest
3551-B	Circonscription sanitaire de la cité d'Ottawa

3553-H		Circonscription sanitaire régionale de Peel
3554-D		Circonscription sanitaire du district de Perth
3555-C		Circonscription sanitaire du comté et de la cité de Peterborough
3556-C		Circonscription sanitaire de Porcupine
3557-D		Circonscription sanitaire du comté et du district de Renfrew
3558-D		Circonscription sanitaire de l'Est de l'Ontario
3560-B		Circonscription sanitaire du district de Simcoe Muskoka
3561-C		Circonscription sanitaire de Sudbury et son district
3562-C		Circonscription sanitaire du district de Thunder Bay
3563-C		Circonscription sanitaire de Timiskaming
3565-B		Circonscription sanitaire de Waterloo
3566-B		Circonscription sanitaire de Wellington-Dufferin-Guelph
3568-A		Circonscription sanitaire de Windsor-Comté d'Essex
3570-H		Circonscription sanitaire régionale de York
3575-D		Circonscription sanitaire d'Oxford, Elgin et St. Thomas
3595-G		Circonscription sanitaire de la cité de Toronto
46	Manitoba	
4601-A		Winnipeg Regional Health Authority
4602-D		Prairie Mountain Health
4603-D		Interlake-Eastern Regional Health
4604-F		Northern Regional Health Authority
4605-B		Santé Sud
47	Saskatchewan	
4701-D		Sun Country Regional Health Authority
4702-D		Five Hills Regional Health Authority
4703-D		Cypress Regional Health Authority

4704-B		Regina Qu'Appelle Regional Health Authority
4705-D		Sunrise Regional Health Authority
4706-B		Saskatoon Regional Health Authority
4707-D		Heartland Regional Health Authority
4708-C		Kelsey Trail Regional Health Authority
4709-C		Prince Albert Parkland Regional Health Authority
4710-B		Prairie North Regional Health Authority
4714-F		Mamawetan/Keewatin/Athabasca ¹⁴
48	Alberta	
4831-B		South Zone
4832-B		Calgary Zone
4833-B		Central Zone
4834-B		Edmonton Zone
4835-I		North Zone
59	Colombie-Britannique	
5911-D		East Kootenay Health Service Delivery Area
5912-C		Kootenay-Boundary Health Service Delivery Area
5913-A		Okanagan Health Service Delivery Area
5914-C		Thompson/Cariboo Health Service Delivery Area
5921-B		Fraser East Health Service Delivery Area
5922-H		Fraser North Health Service Delivery Area
5923-H		Fraser South Health Service Delivery Area
5931-H		Richmond Health Service Delivery Area

¹⁴ Dans le cas de la plupart des sources de données, on ne dispose pas de données au niveau de la région socio-sanitaire pour certaines régions du Nord de la Saskatchewan dont la population est faible. Pour ne pas devoir supprimer de données, on a fusionné avec des régions voisines les régions pour lesquelles la petite taille de l'échantillon ou les petits nombres observés risquaient de nuire à la qualité des données. Ainsi : L'Athabasca Health Authority (4713), de la Saskatchewan, a été fusionnée avec la Mamawetan Churchill River Regional Health Authority (4711) et avec la Keewatin Yatthe Regional Health Authority (4712) pour former une région appelée « Mamawetan/Keewatin/Athabasca Regional Health Authorities (4714) »

5932-G	Vancouver Health Service Delivery Area
5933-A	North Shore/Coast Garibaldi Health Service Delivery Area
5941-A	South Vancouver Island Health Service Delivery Area
5942-C	Central Vancouver Island Health Service Delivery Area
5943-C	North Vancouver Island Health Service Delivery Area
5951-C	Northwest Health Service Delivery Area
5952-C	Northern Interior Health Service Delivery Area
5953-I	Northeast Health Service Delivery Area
A	Groupe de régions homologues A
B	Groupe de régions homologues B
C	Groupe de régions homologues C
D	Groupe de régions homologues D
E	Groupe de régions homologues E
F	Groupe de régions homologues F
G	Groupe de régions homologues G
H	Groupe de régions homologues H
I	Groupe de régions homologues I

**ANNEXE D - REPARTITION DE L'ECHANTILLON PAR
PROVINCE/TERRITOIRE (ESCC 2019-2020 ET ECSA 2019-2020)**

Annexe D – Répartition de l'échantillon par province/territoire (ESCC 2019-2020 et ECSA 2019-2020)¹⁵

Province/Territoire	ESCC		ECSA	Total
	Adultes	Jeunes	Aînés	
Canada	228 739	16 900	49 068	294 707
T.-N.-L.	5 949	553	9 741	16 243
Î.-P.-É.	3 348	471	7 688	11 507
N.-É.	8 877	649	4 698	14 224
N.-B.	7 409	587	5 818	13 814
Qc	40 942	2 712		43 654
Ont.	83 341	5 012		88 353
Man.	10 215	897	5 648	16 760
Sask.	9 222	830	7 104	17 156
Alb.	25 802	1 999	5 760	33 561
C.-B.	30 443	2 418	2 611	35 472
Yn	1 182	288		1 470
T.N.-O.	1 234	274		1 508
Nt	775	210		985

¹⁵ La taille d'échantillon correspond ici au nombre d'unités qui ont été envoyées à la collecte

**ANNEXE E - TAUX DE REPONSE PAR PROVINCE/TERRITOIRE
(ESCC 2019-2020 ET ECSA 2019-2020)**

Annexe E – Tableau 9.1 Taux de réponse par province/territoire (ESCC 2020)

Province/ Territoire	Adultes			Jeunes			Combiné
	# champ de l'enquête	# répondants	Taux de réponse	# champ de l'enquête	# répondants	Taux de réponse	Taux de réponse
Canada	201 834	83 112	41,2	16 343	6 526	39,9	41,1
T.-N.-L.	5 046	2 118	42,0	533	156	29,3	40,8
Î.-P.-É.	2 553	1 094	42,9	463	147	31,7	41,1
N.-É.	7 603	3 265	42,9	644	242	37,6	42,5
N.-B.	6 232	2 701	43,3	577	228	39,5	43,0
Qc	35 134	17 972	51,2	2 658	1 495	56,2	51,5
Ont.	76 632	28 783	37,6	4 899	1 811	37,0	37,5
Man.	8 750	3 644	41,6	812	343	42,2	41,7
Sask.	7 128	3 062	43,0	772	297	38,5	42,5
Alb.	22 886	8 704	38,0	1 940	766	39,5	38,1
C.B.	27 175	10 488	38,6	2 360	794	33,6	38,2
Yn	1 066	471	44,2	266	114	42,9	43,9
T.N.-O.	1 011	491	48,6	240	89	37,1	46,4
Nt	618	319	51,6	179	44	24,6	45,5

Annexe E – Tableau 9.2 Taux de réponse par province (ECSA 2019-2020)

Base	Géo	Contenu principal ESCC			Contenu supplémentaire ECSA				
		# personnes champ enquête	# répondants ¹⁶	Taux de réponse	# aînés champ enquête ¹⁷	# resp ¹⁸	Taux de réponse	# répondants ayant accepté de partager leurs données ¹⁹	Taux de partage
Les deux	Canada	245,639	100,797	41.0%	45,863	41,635	90.8%	38,522	92.5%
	T.-N.-L.	14,387	5,182	36.0%	3,879	3,425	88.3%	3,157	92.2%
	Î.-P.-É.	9,873	4,008	40.6%	3,319	2,996	90.3%	2,766	92.3%
	N.-É.	12,054	5,405	44.8%	3,474	3,092	89.0%	2,904	93.9%
	N.-B.	11,713	5,154	44.0%	3,545	3,120	88.0%	2,906	93.1%
	Qc.	35,134	17,972	51.2%	6,368	5,808	91.2%	5,370	92.5%
	Ont.	76,632	28,783	37.6%	8,538	7,461	87.4%	6,983	93.6%
	Man.	14,074	6,123	43.5%	3,676	3,502	95.3%	3,213	91.7%
	Sask.	13,835	5,932	42.9%	3,818	3,635	95.2%	3,319	91.3%
	Alb.	28,301	10,856	38.4%	4,528	4,303	95.0%	3,968	92.2%
	C.-B.	29,636	11,382	38.4%	4,718	4,293	91.0%	3,936	91.7%
Aréolaire	Canada	199,139	81,831	41.1%	26,897	24,357	90.6%	22,642	93.0%
	T.-N.-L.	5,046	2,118	42.0%	815	753	92.4%	703	93.4%
	Î.-P.-É.	2,553	1,094	42.9%	405	375	92.6%	347	92.5%
	N.-É.	7,603	3,265	42.9%	1,334	1,194	89.5%	1,119	93.7%
	N.-B.	6,232	2,701	43.3%	1,092	1,006	92.1%	949	94.3%
	Qc.	35,134	17,972	51.2%	6,368	5,808	91.2%	5,370	92.5%
	Ont.	76,632	28,783	37.6%	8,538	7,461	87.4%	6,983	93.6%
	Man.	8,750	3,644	41.6%	1,197	1,125	94.0%	1,042	92.6%
	Sask.	7,128	3,062	43.0%	948	908	95.8%	843	92.8%
	Alb.	22,886	8,704	38.0%	2,376	2,251	94.7%	2,093	93.0%
	C.-B.	27,175	10,488	38.6%	3,824	3,476	90.9%	3,193	91.9%
ECSA	Canada	46,500	18,966	40.8%	18,966	17,278	91.1%	15,880	91.9%
	T.-N.-L.	9,341	3,064	32.8%	3,064	2,672	87.2%	2,454	91.8%
	Î.-P.-É.	7,320	2,914	39.8%	2,914	2,621	89.9%	2,419	92.3%
	N.-É.	4,451	2,140	48.1%	2,140	1,898	88.7%	1,785	94.0%
	N.-B.	5,481	2,453	44.8%	2,453	2,114	86.2%	1,957	92.6%
	Man.	5,324	2,479	46.6%	2,479	2,377	95.9%	2,171	91.3%
	Sask.	6,707	2,870	42.8%	2,870	2,727	95.0%	2,476	90.8%
	Alb.	5,415	2,152	39.7%	2,152	2,052	95.4%	1,875	91.4%
	C.-B.	2,461	894	36.3%	894	817	91.4%	743	90.9%

¹⁶ Comprend les répondants adultes de l'ESCC

¹⁷ Comprend tous les aînés de l'ESCC à l'exception de 1 980 répondants provenant du suréchantillon de l'Ontario (les unités du suréchantillon de l'Ontario n'étaient pas éligibles pour l'ECSA)

¹⁸ Comprend les aînés de l'ESCC ayant complété le questionnaire de l'ECSA et qui ont accepté que leurs réponses à l'ESCC soient appariées à celles de l'ECSA

¹⁹ Répondants à l'ECSA ayant accepté de partager leurs données avec certains partenaires

