



**ENQUÊTE SUR LA SANTÉ DANS LES COLLECTIVITÉS
CANADIENNES (ESCC)
CYCLE 3.1 (2005)**

**FICHER DE MICRODONNÉES À GRANDE DIFFUSION
(FMGD)**

GUIDE DE L'UTILISATEUR

Juin 2006



Statistique
Canada

Statistics
Canada

Canada

Table des matières

1.	INTRODUCTION	1
2.	CONTEXTE	2
3.	OBJECTIFS	3
4.	CONTENU DE L'ENQUÊTE	4
4.1	PROCESSUS DE CONSULTATION	4
4.2	CONTENU COMMUN.....	4
4.3	CONTENU OPTIONNEL	5
4.4	CONTENU DES SOUS-ÉCHANTILLONS	5
5.	PLAN D'ÉCHANTILLONNAGE	8
5.1	POPULATION CIBLE	8
5.2	RÉGIONS SOCIO-SANITAIRES.....	8
5.3	TAILLE ET RÉPARTITION DE L'ÉCHANTILLON	8
5.4	BASES DE SONDAGE ET STRATÉGIES D'ÉCHANTILLONNAGE DES MÉNAGES	9
5.4.1	ÉCHANTILLONNAGE DES MÉNAGES À PARTIR DE LA BASE ARÉOLAIRE	9
5.4.2	ÉCHANTILLONNAGE DES MÉNAGES À PARTIR DE LA BASE LISTE DE NUMÉROS DE TÉLÉPHONE	12
5.4.3	ÉCHANTILLONNAGE DES MÉNAGES À PARTIR DE LA BASE DE SONDAGE À CA DE NUMÉROS DE TÉLÉPHONE	12
5.5	ÉCHANTILLONNAGE DES PERSONNES INTERVIEWÉES	13
5.6	RÉPARTITION DE L'ÉCHANTILLON SUR LA PÉRIODE DE COLLECTE DES DONNÉES	13
5.7	ACHAT D'UNITÉS D'ÉCHANTILLONNAGE SUPPLÉMENTAIRES DANS TROIS RSS DE LA PROVINCE DE QUÉBEC.....	14
6.	COLLECTE DES DONNÉES	16
6.1	INTERVIEWS ASSISTÉES PAR ORDINATEUR.....	16
6.2	DÉVELOPPEMENT DES APPLICATIONS DE L'ESCC.....	16
6.3	FORMATION DES INTERVIEWEURS	18
6.4	L'INTERVIEW	18
6.5	OPÉRATIONS SUR LE TERRAIN	21
6.6	CONTRÔLE DE LA QUALITÉ ET GESTION DE LA COLLECTE.....	22
7.	TRAITEMENT DES DONNÉES	24
7.1	VÉRIFICATION	24
7.2	CODAGE	24
7.3	CRÉATION DE VARIABLES DÉRIVÉES ET GROUPEES.....	24
7.4	PONDÉRATION.....	25
7.5	CONVERSION DU FICHIER MAÎTRE DU CYCLE 3.1 DE L'ESCC EN FICHIER DE MICRODONNÉES À GRANDE DIFFUSION (FMGD)	25
8.	PONDÉRATION	28
8.1	PONDÉRATION DE L'ÉCHANTILLON	28
8.1.1	PONDÉRATION DE L'ÉCHANTILLON PROVENANT DE LA BASE ARÉOLAIRE.....	29
8.1.2	PONDÉRATION DE L'ÉCHANTILLON PROVENANT DE LA BASE TÉLÉPHONIQUE.....	32
8.1.3	INTÉGRATION DES BASES ARÉOLAIRE ET TÉLÉPHONIQUE (I1).....	36
8.1.4	EFFET SAISONNIER ET WINSORIZATION (I2).....	36
8.1.5	POSTSTRATIFICATION (I3).....	37
8.1.6	PARTICULARITÉS DE LA PONDÉRATION POUR LES TROIS TERRITOIRES.....	37

9.	QUALITÉ DES DONNÉES	39
9.1	TAUX DE RÉPONSE	39
9.2	ERREURS DANS LES ENQUÊTES.....	44
9.2.1	ERREURS NON DUES À L'ÉCHANTILLONNAGE.....	45
9.2.2	ERREURS DUES À L'ÉCHANTILLONNAGE	45
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	52
11.2	EXEMPLES D'UTILISATION DES TABLEAUX DE CV POUR DES ESTIMATIONS DE TYPE NOMINAL	54
11.3	COMMENT UTILISER LES TABLEAUX DE CV POUR CALCULER LES LIMITES DE CONFIANCE.....	57
11.4	EXEMPLE D'UTILISATION DE TABLEAUX DE CV POUR OBTENIR DES LIMITES DE CONFIANCE	58
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	59
11.8	SEUILS POUR LA DIFFUSION DES ESTIMATIONS RELATIVES À L'ESCC	61
12.	UTILISATION DU FICHIER.....	62
12.1	UTILISATION DE LA VARIABLE DE PONDÉRATION	62
12.2	CONVENTION APPLIQUÉE POUR NOMMER LES VARIABLES	62
12.2.1	STRUCTURE ÉLÉMENTAIRE DES NOMS DES VARIABLES DE L'ESCC	62
12.2.2	POSITIONS 1 À 3 : NOM DE LA VARIABLE/SECTION DU QUESTIONNAIRE	63
12.2.3	POSITION 4 : CYCLE	64
12.2.4	POSITION 5 : TYPE DE VARIABLE.....	65
12.2.5	POSITIONS 6 À 8 : NOM DE LA VARIABLE.....	65
12.3	ACCÈS AU FICHIER MAÎTRE	66
ANNEXE A	68
ANNEXE B	75

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. Le cycle de collecte des données de l'ESCC s'étend présentement sur deux années. La première année du cycle, indiquée par la notation « .1 », correspond à une enquête générale sur la santé de la population réalisée auprès d'un grand échantillon et conçue pour fournir des estimations fiables à l'échelle de la région sociosanitaire. La deuxième année du cycle, représentée par la notation « .2 », a un plus petit échantillon et est conçue pour fournir des données à l'échelle provinciale sur des sujets particuliers ayant trait à la santé.

Le fichier de microdonnées à grande diffusion (FMGD) contient les données recueillies entre janvier 2005 et décembre 2005 pour le cycle 3.1 de l'ESCC. Ces données sont recueillies auprès de personnes âgées de 12 ans et plus vivant dans des logements privés dans 122 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 des établissements, les membres à temps plein des Forces canadiennes 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 du fichier de microdonnées à grande diffusion du cycle 3.1 de l'ESCC qui est décrit en détails dans le texte et les annexes qui suivent.

Pour toute question concernant les ensembles de données ou leur utilisation, s'adresser 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, Division de la statistique de la santé : (613) 951-1746
Courriel : hd-ds@statcan.ca

Renseignements sur le télé-accès : (613) 951-1653
Courriel : cchs-escc@statcan.ca
Télécopieur : (613) 951-4198

2. Contexte

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 canadienne¹.

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 cette 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.

L'objet de cette publication, le FMGD, est produit en réponse à l'exigence du mandat de recueillir des données fiables et pertinentes sur les services de santé, l'état de santé et les questions relatives à la santé revêtant une importance pour la population canadienne — à l'échelle régionale, provinciale et nationale — et de diffuser cette information au public.

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

² 1999. Initiative du carnet de route ... Lancer le processus. Institut canadien d'information sur la santé/Statistique Canada. ISBN 1-895581-70-2. p. 19.

3. Objectifs

Les objectifs principaux des cycles « .1 » de l'ESCC sont les suivants :

- fournir des estimations transversales à jour et fiables des déterminants de la santé, de l'état de santé et de l'utilisation des services de santé à travers le Canada,
- recueillir des données à l'échelle infraprovinciale,

En tant que composante importante du Programme des enquêtes sur la santé de la population de Statistique Canada, les cycles « .1 » de l'ESCC contribuent à combler des besoins d'information accrus en matière de santé au Canada. Il s'agit de :

- faciliter l'élaboration de politiques gouvernementales,
- fournir des données permettant de réaliser des études analytiques qui aideront à comprendre les déterminants de la santé,
- recueillir des données sur les corrélations entre la santé et les facteurs économiques, sociaux, démographiques, professionnels et environnementaux,
- permettre de mieux comprendre la relation entre l'état de santé et l'utilisation des services de santé.

4. Contenu de l'enquête

La présente section décrit le processus général de consultation suivi pour élaborer le contenu de l'enquête et résume le contenu final choisi pour le cycle 3.1 de l'ESCC. La deuxième sous-section décrit en détail le contenu commun, suivie d'une sous-section expliquant le contenu optionnel de l'ESCC (cycle 3.1).

4.1 Processus de consultation

L'un des objectifs principaux de l'ESCC est de combler les lacunes statistiques en ce qui concerne les déterminants de la santé, l'état de santé et l'utilisation des services de santé, à l'échelle de la région sociosanitaire.

Pour identifier ces lacunes, des consultations ont été menées à l'automne 2001 auprès de plus de 200 représentants d'organismes gouvernementaux régionaux, provinciaux et fédéraux, ainsi qu'auprès de chercheurs spécialisés dans l'étude de la santé de la population.

Les consultations menées avant le cycle 1.1 de l'ESCC s'appuyaient sur une combinaison de méthodes qualitatives et quantitatives en vue de déterminer la priorité relative de grands domaines thématiques. L'objectif principal des consultations menées antérieurement au cycle 2.1 était de cerner les nouveaux domaines pour lesquels existent des lacunes statistiques. En ce qui concerne le cycle 3.1, les consultations étaient axées sur l'identification des améliorations minimales des modules de questionnaires existants.

À la suite de ces consultations, une liste de sujets à inclure dans le cycle 3.1 de l'enquête a été dressée par Statistique Canada et approuvée par un comité consultatif formé de représentants des régions sociosanitaires, des ministères provinciaux et territoriaux de la santé et de Santé Canada.

Le questionnaire final du cycle 3.1 de l'ESCC comprend les éléments suivants : environ 25 minutes de contenu commun, s'adressant à tous les répondants, environ 5 minutes de contenu pour les sous-échantillons, c'est-à-dire certains modules du questionnaire que l'on a posé uniquement à un nombre de répondants suffisant pour produire des estimations fiables aux échelles nationale et provinciale, et environ 10 minutes de contenu optionnel.

Chaque région sociosanitaire a eu droit à dix minutes de contenu optionnel. Les représentants régionaux ont choisi les modules du questionnaire d'après une liste pré-établie, en se fondant sur les besoins et les priorités à l'échelle locale. Les modules de contenu optionnel n'ont été posés qu'aux répondants vivant dans les régions sociosanitaires les ayant sélectionnés.

4.2 Contenu commun

Les sujets formant le contenu commun sont très variés : ils vont de la consommation d'alcool à l'exposition à la fumée secondaire, en passant par l'activité physique et l'incapacité au cours des deux dernières semaines. Le tableau 4.1 présente le contenu commun du cycle 3.1 de l'ESCC. Ces questions de l'enquête portant sur les sujets du contenu commun ont été posées à tous les répondants dans toutes les régions sociosanitaires.

Tableau 4.1 Modules de contenu commun du cycle 3.1 de l'ESCC

<ul style="list-style-type: none"> • Consommation d'alcool • Expériences maternelles • Problèmes de santé chroniques • Exposition à la fumée des autres • Vaccin contre la grippe • État de santé général • Utilisation des soins de santé • Taille/poids • Blessures • Mammographie • Comportement sexuel 	<ul style="list-style-type: none"> • Test PAP • Activités physiques • Limitation des activités • Usage du tabac • Incapacité des deux dernières semaines • Usage du tabac chez les jeunes • Revenu • Renseignements socio-démographiques • Administration • Soins de santé à domicile • Population active – Questionnaire abrégé
---	---

4.3 Contenu optionnel

Certains modules de questionnaires ont été considérés comme optionnels de sorte que les régions puissent choisir les modules pertinents à leurs besoins et priorités particuliers (voir le tableau 4.2). Il convient de mentionner que contrairement aux modules formant le volet du contenu commun, les données provenant des modules de contenu optionnel ne peuvent être généralisées facilement à l'échelle du Canada.

4.4 Contenu des sous-échantillons

Trois ensembles de modules de questionnaires ont été posés à seulement un sous-ensemble des répondants. Le but de ces modules était de permettre d'effectuer les calculs des estimations provinciales et nationales tout en minimisant le fardeau de réponse. Dans la plupart des cas, on a posé les questions des sous-échantillons à trois différents groupes de répondants. Le tableau 4.3 dresse le contenu de chaque sous-échantillon.

L'Annexe A présente les lignes directrices sur l'utilisation et l'interprétation du contenu des sous-échantillons.

Tableau 4.2 Modules de contenu optionnel du cycle 3.1 de l'ESCC

<ul style="list-style-type: none"> • Accès aux services de soins de santé * • Dépendance à l'alcool ** • Tension artérielle • Examen des seins • Auto-examen des seins • Changements pour améliorer la santé • Facteurs de stress durant l'enfance et la vie adulte • Dépistage du cancer du côlon et du rectum • Utilisation des services – santé mentale • Visites chez le dentiste • Dépression • Les soins pour le diabète • Suppléments de vitamines et de minéraux** • Détresse • Conduite automobile et sécurité ** • Évaluation des troubles alimentaires ** • Exames de la vue • Choix alimentaires • Sécurité alimentaire • Consommation de fruits et de légumes * • Satisfaction du système des soins de santé • Indice de l'état de santé * • Mesure de sécurité à la maison ** • Hormonothérapie de substitution** • Drogues illicites • Couverture d'assurance • Population active* • Activités de loisirs ** • Maîtrise de soi ** • Consommation de médicaments • Humeur ** • Usage du tabac – étapes du changement 	<ul style="list-style-type: none"> • Dépendance à la nicotine • Outils pour cesser de fumer • Usage du tabac – consultation d'un médecin • Problèmes actuels ** • Santé bucco-dentaire 1 ** • Santé bucco-dentaire 2 • Satisfaction des patients * • Examen général • Jeu pathologique • Dépistage du cancer de la prostate • Échelle de bien-être** • Évènements récents de la vie ** • Satisfaction de la vie • Activités sédentaires • Estime de soi • État de santé – SF-36 • Sentiment de cohérence ** • Sommeil • Soutien social – disponibilité • Soutien social – utilisation • Valeurs spirituelles ** • Faire face au stress • Sources du stress • Pensées suicidaires et tentative de suicide • Protection contre le soleil • Variantes du tabagisme ** • Utilisation de l'équipement protecteur • Organismes bénévoles • Temps d'attente * • Stress au travail
--	--

* Prévu pour le contenu pour les sous-échantillons, mais pouvait aussi être choisi par les régions sociosanitaires à titre de contenu optionnel.

** Offert comme contenu optionnel mais non choisi.

Tableau 4.3 Modules des sous-échantillons du cycle 3.1 de l'ESCC

<p>Sous-échantillon 1</p> <ul style="list-style-type: none">• L'indice de l'état de santé (HUI)• Consommation de fruits et de légumes• Population active (questionnaire détaillé). <p>Sous-échantillon 2</p> <ul style="list-style-type: none">• Mesures de la taille et du poids	<p>Sous-échantillon 3</p> <ul style="list-style-type: none">• Accès aux services de santé• Temps d'attente• Satisfaction des patients
---	--

5. Plan d'échantillonnage

5.1 Population cible

L'ESCC vise la population de 12 ans et plus vivant à domicile des dix provinces et des trois territoires. Sont exclues du champ de l'enquête les personnes vivant sur les réserves indiennes et les terres de la Couronne, les résidents des établissements, les membres à temps plein des Forces canadiennes et les personnes vivant dans certaines régions éloignées. L'ESCC couvre environ 98 % de la population canadienne de 12 ans et plus.

5.2 Régions sociosanitaires

À des fins administratives, chaque province est divisée en plusieurs régions sociosanitaires (RSS) et chaque territoire est considéré comme formant une RSS unique (tableau 5.1). En collaboration avec les provinces, Statistique Canada a modifié légèrement les limites de certaines RSS afin qu'elles correspondent aux données géographiques du Recensement de 2001. Durant le cycle 3.1 de l'ESCC, des données ont été recueillies pour 122 RSS dans les 10 provinces, ainsi que pour une RSS par territoire, soit, en tout, pour 125 RSS.

Tableau 5.1 Nombre de régions sociosanitaires et tailles visées d'échantillon selon la province/territoire

Province	Nombre de RSS	Taille totale de l'échantillon (visée)
Terre-Neuve-et-Labrador	6	4 010
Île-du-Prince-Édouard	4	2 000
Nouvelle-Écosse	6	5 040
Nouveau-Brunswick	7	5 150
Québec	16	24 280
Ontario	37	42 260
Manitoba	10	7 500
Saskatchewan	11	7 720
Alberta	9	12 200
Colombie-Britannique	16	16 090
Yukon	1	850
Territoires du Nord-Ouest	1	900
Nunavut	1	700
Canada	125	128 700

5.3 Taille et répartition de l'échantillon

Afin de produire des estimations fiables pour les 125 RSS et compte tenu du budget accordé pour le cycle 3.1 de l'ESCC, il a été établi que cette composante de l'enquête devrait être réalisée auprès d'un échantillon de 128 700 personnes. La production d'estimations fiables pour la RSS

était l'objectif primordial, 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 RSS et aux provinces. Lors des deux premières étapes, l'échantillon a été réparti entre les provinces en fonction de leur compte de population et du nombre de RSS qu'elles contiennent (tableau 5.1). À la troisième étape, chaque échantillon provincial a été réparti entre les RSS proportionnellement à la racine carrée de la population estimée de la RSS.

Cette stratégie en trois étapes permet d'obtenir un échantillon suffisant pour chaque RSS, sans perturber considérablement la répartition interprovinciale de l'échantillon. L'effectif des échantillons a été augmenté avant la collecte des données afin de tenir compte des logements hors du champ de l'enquête ou vacants, et du taux prévu de non-réponse. Pour la liste complète des RSS et des tailles finales d'échantillons, consulter la section 9 sur la qualité des données.

Il convient de souligner que les trois territoires, qui ont été traités séparément, n'étaient pas visés par la stratégie susmentionnée de répartition de l'échantillon. En tout, 850 unités d'échantillonnage ont été attribuées au Yukon, 900 aux Territoires du Nord-Ouest et 700 au Nunavut.

5.4 Bases de sondage et stratégies d'échantillonnage des ménages

L'échantillon de ménages du cycle 3.1 de l'ESCC a été sélectionné à partir de trois bases de sondage : 49 % de l'échantillon de ménages provenait d'une base de sondage aréolaire, 50 % provenait d'une base liste de numéros de téléphone et 1 % provenait d'une base de sondage à composition aléatoire (CA).

5.4.1 Échantillonnage des ménages à partir de la base aréolaire

La base aréolaire conçue pour l'Enquête sur la population active (EPA) du Canada a servi de base de sondage principale pour l'ESCC. Le plan d'échantillonnage de l'EPA est un plan d'échantillonnage en grappes stratifié à plusieurs degrés où le logement représente l'unité finale d'échantillonnage³. À la première étape, on a formé des strates homogènes et sélectionné un échantillon indépendant de grappes dans chaque strate. À la deuxième étape, on a dressé la liste des logements pour chaque grappe, puis on a sélectionné les logements, ou les ménages, d'après les listes.

Pour les besoins du plan d'échantillonnage de l'EPA, chaque province est répartie en trois catégories de région, à savoir les grands centres urbains, les villes et les régions rurales. Des strates géographiques ou socioéconomiques sont formées à l'intérieur de chaque grand centre urbain. Dans les strates, des grappes sont formées par regroupement de 150 à 250 logements. Dans certains centres urbains, des strates distinctes sont créées pour les appartements ou pour les secteurs de dénombrement (SD) du recensement pour cibler les ménages à haut revenu, les immigrants et les autochtones. Dans chaque strate, on sélectionne six grappes ou des bâtiments d'habitation (parfois 12 ou 18 appartements) par une méthode d'échantillonnage aléatoire avec

³ Statistique Canada (1998). *Méthodologie de l'enquête sur la population active du Canada*. Statistique Canada numéro 71-526-XPB au catalogue.

probabilité proportionnelle à la taille (PPT), cette dernière correspondant au nombre de ménages. Le nombre 6 est utilisé pour l'ensemble du plan d'échantillonnage afin de permettre le renouvellement mensuel d'un sixième de l'échantillon de l'EPA.

Les autres villes et régions rurales de chaque province sont stratifiées, en premier lieu, en fonction de données géographiques, puis selon les caractéristiques socioéconomiques. Dans la plupart des strates, on sélectionne six grappes (habituellement des SD du recensement) par la méthode PPT. Certains centres urbains isolés géographiquement sont couverts par un plan de sondage à trois degrés. Ce type de plan de sondage est utilisé au Québec, en Ontario, en Alberta et en Colombie-Britannique.

Une fois que la liste des nouvelles grappes est établie, on obtient l'échantillon par échantillonnage systématique des logements. Le tableau 5.2 donne un aperçu des catégories d'UPE utilisées pour l'échantillon complet de l'EPA. Le *rendement* correspond au nombre de ménages sélectionnés dans le cadre de l'EPA pour un mois donné. Comme les taux d'échantillonnage sont prédéterminés, il existe souvent un écart entre la taille prévue d'échantillon et les chiffres obtenus. Ainsi, le rendement de l'échantillon est parfois excessif. Cette situation peut, par exemple, se présenter dans des secteurs où le nombre de logements a augmenté à la suite de nouveaux projets de construction. Pour réduire le coût de la collecte des données, on corrige la production excessive par élimination, dès le départ, d'une partie des unités sélectionnées et modifications du coefficient de pondération appliqué dans le plan d'échantillonnage. Cette opération, habituellement réalisée au niveau agrégé, porte le nom de *stabilisation de l'échantillon*. En plus, on augmente la taille requise de l'échantillon de ménages pour tenir compte du fait qu'en général, environ 12 % de logements ne sont pas occupés par des ménages faisant partie du champ d'observation (certains logements sont vacants ou occupés de façon saisonnière, et d'autres sont occupés par des ménages non visés par l'enquête).

Tableau 5.2 Unité primaire d'échantillonnage, taille et rendement

Région	Unité primaire d'échantillonnage (UPE)	Taille (ménages par UPE)	Rendement (ménages échantillonnés)
Toronto, Montréal, Vancouver	Grappe	150 à 250	6
Autres villes	Grappe	150 à 250	8
La plupart des régions rurales/petits centres urbains	Grappe	100 à 250	10

Afin de répondre aux exigences particulières à l'ESCC, certaines modifications ont dû être apportées à cette stratégie d'échantillonnage⁴. Pour obtenir un échantillon de base de 62 000 ménages pour l'ESCC, il faut sélectionner 86 000 logements d'après la base aréolaire (pour tenir compte des logements vacants et des ménages non-répondants). Chaque mois, le plan d'échantillonnage de l'EPA fournit environ 60 000 logements répartis entre les diverses régions

⁴ Morano, M., Lessard, S. et Béland, Y. (2000). Creation of a dual frame for the Canadian Community Health Survey, 2000 *Proceedings of the Survey Methods Section*, Ottawa: La Société statistique du Canada, 249-254

économiques du Canada, alors que, pour l'ESCC, il faut obtenir un échantillon total de 86 000 logements répartis entre les RSS, dont les limites géographiques diffèrent de celles des régions économiques de l'EPA. Globalement, l'ESCC nécessite environ 43 % plus de logements que le nombre produit par le mécanisme de sélection de l'EPA, ce qui correspond à un facteur de redressement de 1,43 (86 000/60 000). Cependant, au niveau de la RSS, ce facteur de redressement varie de 0,6 à 6,0.

Les modifications apportées au processus de sélection dans une RSS varient selon la grandeur du facteur de redressement. Pour les RSS pour lesquelles le facteur est égal ou inférieur à 1, on procède à une simple stabilisation de l'échantillon de logements, telle que décrite plus haut. Pour celles pour lesquelles le facteur est supérieur à 1, mais inférieur ou égal à 2, on répète le processus d'échantillonnage des logements à l'intérieur d'une UPE pour toutes les UPE sélectionnées appartenant à la RSS en question. Pour les RSS pour lesquelles le facteur est supérieur à 2, mais inférieur ou égal à 4, on répète le processus d'échantillonnage des UPE ainsi que celui des logements dans les UPE. Pour les RSS pour lesquelles le facteur est compris entre 4 et 6, on répète le processus d'échantillonnage des UPE deux fois au lieu d'une, et celui de sélection des logements une fois uniquement. Dans les cas où la méthode choisie produit un excédent inutile de logements, on procède à la stabilisation de l'échantillon.

Il convient de souligner que les modifications apportées au processus d'échantillonnage de l'EPA aboutissent, au plus, au triplement du nombre d'UPE sélectionnées et, au plus, au doublement du nombre de logements sélectionnés dans les UPE, ce qui explique la valeur maximale de 6,0 du facteur de redressement. Pour les RSS, on a plafonné délibérément la valeur du facteur de redressement à 6,0 pour deux raisons : pour limiter le nombre de listes de grappes à produire (chaque nouvelle UPE sélectionnée nécessite une liste) et pour éviter les effets éventuels de grappes dus à la sélection d'un trop grand nombre de logements dans une même UPE. Cette limite du facteur de redressement appliqué pour certaines RSS a conséquemment dicté le nombre de ménages qu'il fallait sélectionner à partir des bases de sondage téléphoniques.

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 utilisé pour les trois territoires comporte une étape supplémentaire. Pour chaque territoire, on a commencé par stratifier les collectivités (municipalités) faisant partie du champ de l'enquête en fonction de diverses caractéristiques (population, données géographiques, proportion d'Inuit et/ou d'Autochtones et revenu médian du ménage). On a défini de la sorte 5 strates pour le Yukon, 10 pour les Territoires du Nord-Ouest et 6 pour le Nunavut. Puis, le premier degré d'échantillonnage a consisté à sélectionner aléatoirement une collectivité avec probabilité proportionnelle à la taille de la population dans chaque strate définie. Puis, de là, on a appliqué, à l'intérieur de chaque collectivité, une stratégie d'échantillonnage des ménages à partir de la base aréolaire identique à celle décrite plus haut.

Il convient de mentionner que la base de sondage de l'ESCC couvre 90 % des ménages privés du Yukon, 97 % de ceux des Territoires du Nord-Ouest et 71 % de ceux du Nunavut.

5.4.2 Échantillonnage des ménages à partir de la base liste de numéros de téléphone

À l'exception de 5 RSS, la base liste de numéros de téléphone a été utilisée dans toutes les régions pour compléter la base aréolaire; seules les 2 RSS qui ont eu recours à la base CA et les trois territoires n'ont pas utilisé la base liste. À cette fin, on a apparié l'annuaire téléphonique du Canada, un disque compact disponible sur le marché contenant les noms, les adresses et les numéros de téléphone répertoriés dans les annuaires de téléphone du Canada, à des fichiers internes administratifs de conversion afin d'obtenir les codes postaux que l'on a fait correspondre aux RSS afin de créer des listes stratifiées de numéros de téléphone, à raison d'une liste par RSS. Dans chaque strate, on a sélectionné le nombre requis de numéros de téléphone d'après la base liste par échantillonnage aléatoire simple. Comme pour la base de sondage à CA, on a sélectionné des numéros de téléphone supplémentaires pour tenir compte des numéros hors service ou hors du champ d'observation. Le taux de réussite enregistré pour la sélection d'après la liste de numéros de téléphone variant de 75 % à 88 % selon la province, qui est nettement plus élevé que celui observé pour la liste établie par CA.

Il importe de souligner que la couverture de la base liste de numéros de téléphone est moins importante que celle de la base de sondage à CA, car les numéros non publiés n'ont aucune chance d'être sélectionnés. Néanmoins, comme la liste des numéros de téléphone est toujours utilisée en complément de la base aréolaire, l'effet du sous-dénombrement dû à l'utilisation de la base liste de numéros de téléphone est minimal et peut être corrigé par pondération.

5.4.3 Échantillonnage des ménages à partir de la base de sondage à CA de numéros de téléphone

Dans 4 RSS, un échantillon de numéros de téléphone provenant de la base de sondage à composition aléatoire (CA) a été utilisé pour sélectionner un échantillon de ménages. L'échantillonnage de ménages à partir de la base à CA a été réalisé selon la méthode d'élimination des banques non valides (EBNV) adoptée par l'Enquête sociale générale⁵. Une banque de cent numéros (c'est-à-dire les huit premiers chiffres d'un numéro de téléphone à 10 chiffres) est considérée comme non valide si elle ne contient aucun numéro de téléphone résidentiel. Au départ, la base de sondage comprend la liste de toutes les banques valides de cent numéros et celles qui ne sont pas valides sont éliminées de la base de sondage à mesure qu'on les repère. Il convient de souligner que ces banques de cent numéros ne sont éliminées de la base de sondage que lorsque l'on possède des preuves qu'elles ne sont pas valides provenant de sources diverses multiples. En l'absence de renseignements, la banque est retenue dans la base de sondage. Pour éliminer les banques non valides, on s'est servi de l'annuaire du téléphone, ainsi que de divers fichiers administratifs internes.

D'après les renseignements géographiques disponibles (codes postaux), les banques de cent numéros retenues dans la base de sondage ont été regroupées pour créer des strates CA englobant, de façon aussi exacte que possible, les régions sociosanitaires. À l'intérieur de chaque strate CA,

⁵ Norris, D.A., Paton, D.G. (1991), L'Enquête sociale générale canadienne: bilan des cinq premières années. *Techniques d'enquête* (Statistique Canada, Catalogue 12-001); 17, pp. 245-260.

on a choisi au hasard une banque de cent numéros et généré aléatoirement un numéro compris entre 00 et 99 afin de créer un numéro de téléphone complet à 10 chiffres. Cette méthode a été répétée jusqu'à ce que l'on ait atteint le nombre requis de numéros de téléphone pour la strate CA. Comme, fréquemment, le numéro obtenu n'est pas en service ou est hors du champ d'observation, il faut générer un grand nombre de numéros supplémentaires pour atteindre la taille visée d'échantillon. Ce taux de réussite diffère selon la région. Dans le cas de l'ESCC, il variait de 27 % à 49 % parmi les 4 RSS qui y ont eu recours.

5.5 Échantillonnage des personnes interviewées

Comme pour les cycles précédents, la sélection des répondants a été conçue de façon à ce que les jeunes (de 12 à 19 ans) soient surreprésentés dans l'échantillon. La stratégie d'échantillonnage adoptée tient compte des besoins des utilisateurs de données, du coût, de l'efficacité du plan d'échantillonnage, du fardeau de réponse et des contraintes opérationnelles. Pour le cycle 3.1 de l'ESCC, il a été décidé de sélectionner une personne par ménage en utilisant diverses probabilités de sélection variant selon l'âge et selon la composition du ménage. On a simulé plusieurs scénarios reposant sur divers paramètres dans le but de déterminer l'approche optimale sans générer de poids d'échantillonnage extrêmes en bout de ligne.

Le Tableau 5.3 donne les facteurs multiplicatifs de poids de sélection utilisés pour déterminer les probabilités de sélection des personnes dans les ménages échantillonnés, par âge. Par exemple, pour un ménage de trois personnes (deux adultes âgés entre 45 et 64 et un jeune de 15 ans), le jeune de 15 ans avait 5 fois plus de chances d'être sélectionné que les adultes. 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 ou si le nombre de personnes âgées de 12 à 19 ans est plus grand ou égal à 3, alors le facteur multiplicatif du poids de sélection 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 5.3 Facteurs multiplicatifs du poids de sélection pour la stratégie d'échantillonnage au niveau de la personne, par âge

Facteurs multiplicatifs du poids de sélection					
Age	12-19	20-29	30-44	45-64	65+
Facteur	5	2	2	1	1

5.6 Répartition de l'échantillon sur la période de collecte des données

Afin d'équilibrer la charge de travail des intervieweurs et de réduire au minimum les effets saisonniers éventuels sur des caractéristiques importantes telle que l'activité physique, dans chaque RSS, l'échantillon initial de logements/numéros de téléphone a été réparti au hasard, de façon égale, sur les 11 mois de la collecte des données (le douzième mois sert habituellement de mois de « nettoyage »). Pour commencer, chaque UPE sélectionnée au premier degré de l'échantillonnage à partir de la base aréolaire a été affectée au hasard à un trimestre de collecte des données (Q1 : janvier à mars 2005, Q2 : avril et mai 2005, Q3 : juin à août 2005 et Q4 : septembre

à novembre 2005). Pour chaque trimestre de collecte, les logements sélectionnés ont alors été attribués au hasard à un mois de collecte. Pour les listes des numéros de téléphone, des échantillons indépendants ont été sélectionnés chaque mois. Cette stratégie a permis d'assurer que chaque échantillon trimestriel soit représentatif de la population canadienne faisant partie du champ d'observation de l'enquête.

5.7 Achat d'unités d'échantillonnage supplémentaires dans trois RSS de la province de Québec

Afin d'obtenir des estimations fiables à l'échelle infrarégionale, trois RSS de la province de Québec ont fourni des fonds supplémentaires afin que l'on puisse sélectionner un échantillon plus important de logements. Les unités d'échantillonnage supplémentaires ont été regroupées à l'échantillon principal afin de produire un grand fichier de données.

Les échantillons d'unités supplémentaires ont été sélectionnés entièrement à partir de la base liste de numéros de téléphone. Pour cela, on a apparié l'annuaire téléphonique du Canada à des fichiers administratifs internes afin de stratifier les numéros de téléphone publiés dans les régions infrarégionales (8 pour le Bas-St-Laurent, 12 pour Montréal-Centre et 2 pour Laval). Les tailles de l'échantillon sélectionné par région infraprovinciale ont été établies d'après les fonds disponibles et les exigences des RSS quant à l'obtention d'estimations fiables selon les régions infrarégionales (Bas St-Laurent a ajouté 2 400 unités, Montréal-Centre a ajouté 2 295 et Laval a ajouté 1 080). Le tableau 5.4 donne la répartition des échantillons supplémentaires selon les régions infrarégionales.

Tableau 5.4 Répartition des unités d'échantillonnage supplémentaires parmi les RSS du Bas-St-Laurent, Montréal-Centre et de Laval

Région infraprovinciale	Taille totale (visée)
Région de Bas-St-Laurent	2 400
La Matapédia	300
Matane	300
La Mitis	300
Rimouski-Neigette	300
Les Basques	300
Rivière-du-Loup	300
Témiscouata	300
Kamouraska	300
Région de Montréal-Centre	2 295
Pierrefonds et Lac St-Louis	191
LaSalle et du Vieux Lachine	191
Verdun/Côte-St-Paul, St-Henri et Pointe-St-Charles	191
René-Cassin, NDG/Montréal-Ouest	191
Côte-des-Neige, Métro et Parc Extension	191
Nord de l'île et St-Laurent	191
Ahuntsic et Montréal-Nord	191
Petite-Patrie et Villeray	191
Des Faubourgs, Plateau Mont-Royal et St-Louis-du Parc	191
St-Michel et St-Léonard	191
Hochelaga-Maisonneuve, Olivier-Guimond et Rosemont	191
Rivière-des-Prairies, Mercier-Est/Anjou et Pointe-aux- Trembles/Montréal-Est	191
Région de Laval	1 080
Est	540
Ouest	540

6. Collecte des données

6.1 Interviews assistées par ordinateur

La collecte de données aux fins du cycle 3.1 de l'ESCC a eu lieu de janvier à décembre 2005. Au cours de la période, 132 947 interviews valables assistées par ordinateur (IAO) ont été effectuées. Environ la moitié ont eu lieu au moyen de la méthode de l'interview sur place assistée par ordinateur (IPAO), l'autre moitié ayant consisté en des interviews téléphoniques assistées par ordinateur (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 dans un cas particulier et produit des rapports aux fins de la gestion du processus de collecte. L'IAO comprend également un ordonnanceur automatique d'appels, c'est-à-dire un système central qui optimise l'horaire des rappels et le calendrier des rendez-vous à l'appui de la collecte par ITAO.

Le système de gestion des cas achemine les applications de questionnaire et les fichiers d'échantillons du Bureau central de Statistique Canada aux bureaux régionaux de collecte dans le cas de l'ITAO et des bureaux régionaux aux ordinateurs portables des intervieweurs dans celui de l'IPAO. 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, grâce à l'IAO une interview personnalisée peut être conçue à l'intention de chaque répondant en fonction de ses caractéristiques particulières et des réponses d'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 la fourchette 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

Deux applications d'IAO distinctes ont été développées en prévision du cycle 3.1, l'une pour les interviews téléphoniques (ITAO), l'autre pour les interviews sur place (IPAO). Cette façon de faire permettait de profiter de composantes ou de procédures d'entrée et de sortie, nouvellement normalisées, conçues par Statistique Canada et d'adapter la fonctionnalité de chaque application au type d'interview menée. Les applications comportent les composantes Entrée, C2 (contenu sur la santé) et Sortie.

Les composantes Entrée et Sortie comprennent des séries standard de questions auxquelles l'intervieweur a pu se référer pour prendre contact avec un répondant, recueillir de l'information importante sur l'échantillon, choisir les répondants et évaluer l'état des cas. La composante C2 consiste en les modules sur la santé et représente la plus grande partie des applications. Il s'agit, notamment, des modules communs posés à tous les répondants et du contenu optionnel, qui variaient d'une région sociosanitaire à l'autre.

Deux applications expérimentales ont été développées qui comportent un sous-ensemble des modules C2, de même que les composantes Entrée et Sortie. Chacune a été l'objet d'essais pilotes distincts à la fin de l'été 2004.

Les essais pilotes visaient principalement :

- à vérifier les modifications apportées aux composantes Entrée, Sortie et C2;
- à évaluer les réactions des répondants aux nouvelles questions du cycle 3.1 sur le diabète et la thérapie hormonale de remplacement;
- à expérimenter l'infrastructure technique, y compris la transmission de données sur les serveurs des bureaux régionaux et les procédures propres à l'ITAO menée au moyen du logiciel BLAISE.

Les applications expérimentales ont été modifiées ou améliorées sur la base des résultats des essais pilotes. Les derniers essais des applications intégrales ont été entrepris au terme des modifications. Ils se sont déroulés en trois étapes : 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 tests intégrés portent sur l'ensemble des modules expérimentés, regroupés en applications intégrées avec les composantes Entrée et Sortie. La deuxième étape des essais vise à assurer que des renseignements clés, par exemple l'âge et le sexe, sont transmis de la composante Entrée aux sous-programmes C2 et Sortie. Elle 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 C2. Étant donné que, à ce moment, le fonctionnement des applications est essentiellement identique à ce qu'il sera sur le terrain, tous les scénarios possibles auxquels feront face les intervieweurs sont simulés afin d'en assurer la fonctionnalité rigoureuse. Les scénarios servent à tester 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 répond à la fourchette des valeurs acceptables, le listage de ménages, la prise de rendez-vous et la sélection de répondants. Les tests servent également à confirmer que, au cours d'une interview, les modules de contenu optionnel choisis pour une région sociosanitaire donnée 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 des essais permet d'expérimenter tous les aspects techniques de la saisie, de la transmission et de l'extraction des données pour chacune des applications du cycle 3.1 de l'ESCC. Il s'agit, par ailleurs, de la dernière occasion de déceler des erreurs dans les composantes Entrée, C2 et Sortie.

6.3 Formation des intervieweurs

Des gestionnaires de projet des bureaux régionaux de collecte ont assisté à des séances de formation sur le cycle 3.1 de l'ESCC à Statistique Canada en décembre 2004. Les séances, animées par des membres de l'équipe responsable du projet de l'ESCC, ont consisté en des descriptions de la formation qui serait donnée aux intervieweurs en région. Les gestionnaires ont ensuite animé des séances de formation personnalisées à l'intention des intervieweurs. Des membres de l'équipe responsable de l'enquête ont assisté aux séances afin d'y offrir un soutien et de répondre aux questions et aux préoccupations des participants.

L'objet des séances était de familiariser les intervieweurs avec les applications du cycle 3.1 de l'ESCC, notamment le contenu. La formation était centrée sur :

- les buts et objectifs de l'enquête;
- les techniques d'enquête;
- les fonctionnalités des applications;
- le contenu des questionnaires, qui a été l'objet d'exercices;
- les techniques à appliquer par l'intervieweur pour mener à bien l'interview, soit des exercices complets dont l'objet était de 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

Des unités d'échantillonnage sélectionnées à partir de la liste de numéros de téléphone et de la base CA ont répondu aux questions posées, à partir de centres d'appel centralisés, par des intervieweurs selon la méthode de l'ITAO. Un intervieweur principal affecté au même centre d'appels assurait la surveillance des intervieweurs. Des intervieweurs sur place décentralisés ont interviewé, au moyen de la méthode de l'IPAO, des unités d'échantillonnage sélectionnées dans la base aréolaire. Bien que, dans certaines situations, les intervieweurs sur place aient été autorisés à

mener tout ou partie d'une interview par téléphone, près des trois quarts (72,9 %) ont été effectuées exclusivement sur place. Les intervieweurs sur place ont effectué leur travail en autonomie à la maison, au moyen d'ordinateurs portables, et ils étaient surveillés à distance par des intervieweurs principaux. La variable SAME_TYP apparaissant dans le FMGD signifie qu'un cas a été choisi soit dans la base aréolaire (IPAO), soit dans la liste des numéros de téléphone ou la base CA (ITAO).

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 C2.

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 ADME_N09 indique si l'interview a été effectuée sur place, par téléphone ou au moyen d'un ensemble des deux techniques.

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 le fichier de microdonnées, des indicateurs signalent si une personne autre que le répondant était présente à l'interview (ADME_N10) et si, de l'avis de l'intervieweur, la présence de l'autre personne a influencé les réponses du répondant (ADME_N11).

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

Lettres d'introduction

Avant le début de chaque période de collecte, les ménages échantillonnés ont reçu des lettres d'introduction 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 du cycle 3.1 de l'ESCC.

Prise de contact

Les intervieweurs ont reçu instruction de mettre en œuvre tous les moyens raisonnables pour obtenir des interviews. Lorsque l'appel (ou la visite) de l'intervieweur était prévu à un moment peu commode, il fixait le moment d'un rappel qui convenait au répondant. Si l'intervieweur ne parvenait pas à prendre rendez-vous par téléphone, il devait effectuer une visite de suivi sur place. S'il n'y avait personne à la maison lors de la première visite, l'intervieweur laissait à la porte une brochure qui expliquait l'enquête et annonçait l'intention de l'intervieweur de prendre contact ultérieurement. De nombreux rappels ont été effectués, à divers moments et différents jours.

Conversion des cas de refus

Si une personne refusait d'abord 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 collaboration du ménage. Ensuite, un intervieweur principal, un surveillant de projet ou un autre intervieweur rappelait le répondant (ou lui rendait visite) pour faire valoir l'importance de sa participation.

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. De plus, les questions ont été traduites en chinois, panjabi et inuktitut. Le chinois et le panjabi sont les deux langues qui ont le plus souvent posé problème pour les bureaux régionaux. La traduction en inuktitut a servi à la collecte de données au Nunavut.

Interviews de jeunes

Les intervieweurs ont dû obtenir l'autorisation verbale des parents ou tuteurs pour interviewer des répondants âgés de 12 à 15 ans. Ils ont fait appel à plusieurs procédures pour répondre aux inquiétudes des parents et mener à terme les interviews. Notamment, ils portaient sur eux une fiche intitulée « Note aux parents/tuteurs concernant les interviews de jeunes à l'intention de l'Enquête sur la santé dans les collectivités canadiennes ». La fiche expliquait les raisons pour lesquelles des renseignements étaient recueillis auprès de jeunes, énumérait les thèmes dont traitait l'enquête, demandait l'autorisation de communiquer et de coupler l'information obtenue et expliquait la nécessité de respecter la vie privée et la confidentialité des jeunes.

Si un parent ou tuteur demandait à voir les questions, les intervieweurs avaient pour consigne soit de les leur montrer, soit, si l'interview avait lieu au téléphone, 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 le jeune sélectionné, soit sur place, soit par téléphone (sans qu'une autre personne soit à l'écoute), le code de refus était attribué à l'interview. Cependant, dans le cas des interviews selon la méthode de l'IPAO, s'il était impossible d'interviewer en confidence le jeune sélectionné, l'intervieweur pouvait proposer au parent ou tuteur de lui permettre de lire à haute voix les questions, après quoi le jeune pouvait y répondre directement à l'ordinateur.

Lors de toutes les interviews de jeunes, le parent ou tuteur a répondu aux questions sur le revenu et la sécurité alimentaire. Les questions précitées étaient posées à la fin du questionnaire, de telle sorte que le parent ou tuteur pouvait, à ce moment, compléter l'interview.

Interviews par procuration

Dans les cas où le répondant sélectionné était, pour des raisons de santé physique ou mentale, incapable de répondre à l'interview, les renseignements à son sujet ont été fournis par un autre

membre bien informé du ménage. Cette façon de faire est qualifiée d'interview par procuration. Quoique les interviewés 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. Il fallait donc tout tenter pour réduire au minimum le nombre d'interviews par procuration. La variable ADME_PRX indique si l'interview a été réalisée par procuration ou non.

6.5 Opérations sur le terrain

La plus grande part de l'échantillon du cycle 3.1 a été répartie en onze périodes de collecte chevauchantes de deux mois chacune. Les bureaux régionaux de collecte ont reçu instruction de passer les quatre premières semaines de chaque période de collecte à interviewer la majorité de l'échantillon, puis de consacrer les quatre semaines suivantes aux interviews restantes et au suivi des cas de non-réponse. À la deuxième semaine de chaque période, des tentatives devaient avoir été effectuées relativement à tous les cas.

Des techniques de collecte légèrement différentes ont dû être mises en œuvre aux fins de certaines parties de l'échantillon du cycle 3.1. Des périodes *trimestrielles* distinctes de collecte ont été établies en vue de l'exécution de l'enquête dans les collectivités éloignées du Nunavut, des T.-N.-O. et du Yukon. Des périodes de collecte trimestrielles ont également été prévues aux fins des sous-échantillons de mesures de taille et de poids (voir l'annexe A) pour garantir l'application en règle et uniforme des techniques de collecte spécialisées et la prise en charge des échantillons visés par des intervieweurs formés à la tâche. En dernier lieu, la collecte de renseignements auprès de l'échantillon tiré de la base CA a été effectuée au cours de périodes trimestrielles, car ses caractéristiques différaient de celles de l'échantillon provenant de la liste de numéros de téléphone, notamment les cibles et les calendriers de collecte.

Les bureaux de collecte centralisés ont reçu les fichiers d'échantillons 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.

Le surveillant de projet, l'intervieweur principal et l'équipe de soutien technique du bureau régional étaient chargés de transmettre les cas de chaque bureau responsable des interviews effectuées selon la méthode de l'ITAO au Bureau central. Les transmissions ont eu lieu la nuit, et tous les cas menés à bien ont été acheminés au Bureau central de Statistique Canada. Les interviews menées selon la méthode de l'IPAO ont été transmises chaque jour du domicile de l'intervieweur directement au Bureau central de Statistique Canada par voie d'une ligne de téléphone sécurisée.

En juin et de nouveau en novembre, les cas de non-réponse (refus exceptés) des périodes de collecte antérieures ont été retournés aux responsables aux fins de suivi. Il s'agissait de situations dans lesquelles les répondants sélectionnés avaient été absents pendant l'ensemble de la période de collecte initiale. Les répondants concernés ont été relancés et encouragés à participer à l'enquête.

Au terme de la collecte de données, un taux de réponse national de 79 % a été atteint. La section 9 offre le détail des taux de réponse.

6.6 Contrôle de la qualité et gestion de la collecte

Plusieurs méthodes ont servi à assurer la qualité des données et l'optimisation de la collecte au cours du cycle 3.1. Il s'est agi, 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.

Le bureau régional validait régulièrement le travail des intervieweurs sur place. Des cas choisis ont été repérés aléatoirement dans les échantillons à chaque période de collecte. Les gestionnaires et les surveillants des bureaux régionaux ont dressé des listes de cas à valider, lesquels ont été confiés à l'équipe de validation, qui communiquait avec les ménages concernés afin de confirmer la tenue d'une interview en règle. Par souci de repérer promptement les problèmes, la validation avait normalement lieu au cours des premières semaines d'une période de collecte. Ensuite, les surveillants adressaient régulièrement une rétroaction aux intervieweurs.

Les intervieweurs travaillant par téléphone ont également été l'objet d'une validation aléatoire. En l'espèce, les intervieweurs principaux des bureaux de collecte responsables des interviews selon la méthode de l'ITAO contrôlaient les interviews afin de vérifier que l'intervieweur appliquait les techniques et les procédures prévues (c'est-à-dire qu'il lisait le libellé des questions tel qu'il figurait dans les applications, qu'il ne posait pas de questions incitatives, et ainsi de suite).

Les responsables ont produit une série de rapports dont l'objet était de contrôler et de gérer efficacement les cibles de collecte et de mettre au jour les problèmes posés par la collecte.

À la fin de chaque période, des rapports cumulatifs ont été produits qui précisaient les taux de réponse, de couplage, de partage et d'interview par procuration, ventilés par échantillon ITAO et IPAO, de même que par région sociosanitaire. Les rapports ont servi à cerner les régions où les niveaux de collecte étaient inférieurs aux cibles, de sorte que les bureaux régionaux puissent y concentrer leurs efforts. S'ajoutaient aux rapports cumulatifs, une série de rapports faisant état de cibles de collecte hebdomadaires pour chaque bureau régional, qui ont servi à la planification des ressources et à l'atteinte des cibles de collecte.

Le Bureau central a effectué des analyses complémentaires au moyen de l'information tirée des applications d'IAO afin de recenser les interviews de durée excessivement courte. Ces dernières ont été signalées au moyen d'indicateurs, supprimées des microdonnées et classées parmi les cas de non-réponse.

Par ailleurs, des rapports personnalisés ont été établis et utilisés pour examiner des questions particulières de qualité des données qui se sont manifestées au cours de la collecte. À titre d'exemple, l'une des principales cibles de collecte du sous-échantillon des mesures de taille et de poids était le taux de recueil par les intervieweurs de mesures valables. Il s'agissait d'un taux d'importance capitale, dont ne traitaient pas les rapports cumulatifs. Lorsqu'il a été constaté que les taux étaient inférieurs aux cibles, des rapports personnalisés ont été établis, et les responsables s'y sont référés pour identifier les intervieweurs qui avaient du mal à obtenir des mesures valables. Ces derniers ont bénéficié ensuite d'une formation supplémentaire axée sur des procédures particulières de collecte.

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 l'instruction de contrôle programmée « 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 Codage

On a fourni des catégories de réponses précodées 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. 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.

7.3 Création de variables dérivées et groupées

Pour faciliter l'analyse des données, on a dérivé un certain nombre de variables à partir des éléments disponibles sur le questionnaire du cycle 3.1 de l'ESCC. Le cinquième 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 12.2.

7.4 Pondération

Le principe de base de l'estimation dans un échantillon aléatoire comme celui du cycle 3.1 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 le fichier de microdonnées à grande diffusion 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.

Vous trouverez les détails sur la façon dont on calcule les poids d'échantillonnage à la section 8.

7.5 Conversion du fichier maître du cycle 3.1 de l'ESCC en fichier de microdonnées à grande diffusion (FMGD)

La technique d'élaboration d'un FMGD vise à concilier l'impératif d'assurer la confidentialité des répondants, tout en réduisant au minimum le risque de divulgation, et la nécessité de produire des données d'utilité maximale à l'échelon géographique de la région sociosanitaire. Les paragraphes qui suivent décrivent à grands traits certaines des procédures de conversion du fichier maître de l'ESCC en FMGD.

Régions sociosanitaires : Les régions sociosanitaires sont de compétence provinciale. Comme il a été mentionné à la section 5.2, il existait 125 régions sociosanitaires au Canada au moment de la conception du plan d'échantillonnage du cycle 3.1. Au cours de la collecte, Terre-Neuve-et-Labrador a fondu 2 régions avec 2 autres et l'Ontario a scindé 1 région et fusionné les deux parties ainsi obtenues avec 2 régions existantes, pour un total de 122 régions sociosanitaires auxquelles le cycle 3.1 de l'ESCC serait diffusé. Par ailleurs, 3 régions du Québec ont acquis des unités d'échantillonnage supplémentaires pour pouvoir obtenir des estimations sous-régionales. Il s'agit de Montréal (12 sous-régions de CLSC), du Bas-Saint-Laurent (8) et de Laval (2). Ainsi, le fichier de données maître contient de l'information représentative de 122 régions sociosanitaires et 22 sous-régions de CLSC (la prise en compte des sous-régions fait suite à l'acquisition d'unités d'échantillonnage supplémentaires).

La population de certaines régions sociosanitaires a été jugée trop modeste pour être présentée distinctement dans le FMGD. Les régions dont c'est le cas ont donc été fusionnées avec d'autres.

La valeur minimale de 70 000 habitants a déterminé l'autonomie ou la fusion des régions sociosanitaires dans le FMGD. Il en a résulté :

- i) le regroupement de 33 régions sociosanitaires, dans toutes les provinces sauf le Québec et la Nouvelle-Écosse, en 15 ensembles;
- ii) le regroupement des 3 territoires (qui provoque la suppression de certains contenus optionnels);

- iii) l'exclusion du FMGD d'une région sociosanitaire du Québec (Nord-du-Québec) qui ne se prêtait pas au regroupement à cause de sa faible population et de ses caractéristiques démographiques;
- iv) l'exclusion de toutes variables de plan d'échantillonnage.

Après regroupement, le FMGD comptait 101 régions sociosanitaires, particulières ou regroupées, d'un océan à l'autre. L'annexe B illustre les régions, avant et après regroupement. Les sous-régions de CLSC, particulières ou regroupées, ont été pris en compte dans le calcul des données aux fins de l'échantillon supplémentaire ajouté à 3 régions sociosanitaires du Québec. La population desservie par tous les CLSC de Laval et Montréal est supérieure à 70 000 et aucun regroupement n'a donc été nécessaire à ces endroits. Comme le montre l'annexe B, les sous-régions de 8 CLSC du Bas-Saint-Laurent ont été fondues en deux régions. Par conséquent, le FMGD fournit de l'information sur 101 régions sociosanitaires, particulières ou regroupées, et 16 sous-régions de CLSC, également particulières ou regroupées.

Suppression de contenu optionnel : À la suite du regroupement des régions sociosanitaires, il aurait normalement fallu supprimer le contenu optionnel qui n'avait pas été retenu aux fins de chaque région d'un regroupement. Bien que la sélection de contenu optionnel puisse varier d'une région à l'autre d'une province, toutes les régions des provinces sans exception ont choisi le même contenu optionnel dans le cadre du cycle 3.1. Il n'a donc été nécessaire de supprimer aucun contenu optionnel dans une province donnée. Par contre, comme les trois territoires ont été réunis, le contenu optionnel a été éliminé là où les trois n'ont pas tous opté pour le contenu visé.

Contrôle de la divulgation : Comme il a été mentionné déjà, le FMGD du cycle 3.1 a été conçu de manière à sauvegarder au maximum la valeur analytique des données tout en réduisant au minimum le risque d'identification des répondants particuliers.

Le FMGD est assorti de plusieurs contrôles. Parmi les mesures de protection utilisées figurent la suppression des identificateurs directs (tels identificateur de l'échantillon, nom, numéro de téléphone) et la limitation, le recodage ou la suppression de particularités découlant de fréquences modestes correspondant à des caractéristiques sociodémographiques précises. Les variables du fichier maître exclues du FMGD à cause d'un risque élevé de divulgation (là où la variable présente un risque en soi ou lorsque conjuguée à d'autres) comprennent les tentatives de suicide au cours des 12 derniers mois, les mesures précises de taille et de poids et l'état de grossesse des répondantes.

Certaines catégories de réponses réputées présenter un risque de divulgation ont été regroupées et intégrées au FMGD. Il s'agit, notamment, de la partie du corps touchée par la blessure la plus grave, du nombre de consultations de professionnels de la santé, du nombre d'années écoulées depuis que le répondant a cessé de fumer et de la principale source de revenu du ménage. Ainsi, le FMGD à la fois contient moins de variables et est moins détaillé que le fichier maître du cycle 3.1, sans toutefois que la valeur analytique des données en soit altérée.

Âge des répondants : Le FMGD du cycle 3.1 de l'ESCC comprend les groupes d'âge des répondants. La plupart des groupes représentent un intervalle de 5 ans, allant des 20 à 24 ans aux 75 à 79 ans. Les répondants âgés de 80 ans et plus ont également été réunis. Les FMGD des cycles

antérieurs de l'ESCC distinguaient des tranches de 12 à 14 ans et de 15 à 19 ans. L'actuel fichier répartit le dernier groupe mentionné en deux tranches, soit les 15 à 17 ans et les 18 et 19 ans. Le groupe des 15 à 19 ans a été scindé pour deux principales raisons :

- 1) l'indice de la masse corporelle (IMC) est passé au premier rang des priorités dans le domaine de la santé publique. Des calculs d'IMC sont disponibles pour la première fois pour les jeunes de 12 à 17 ans, mais les utilisateurs n'auraient pas eu accès pleinement aux données si le groupe des 15 à 19 ans avait été conservé;
- 2) à l'occasion des diffusions antérieures, il a fallu supprimer des données concernant les répondants âgés de 18 ou 19 ans à cause de variables recueillies uniquement auprès de répondants âgés de 18 ans et plus.

Ainsi, la nouvelle répartition des jeunes par groupes d'âge présente au moins trois avantages importants :

- i) les données sur l'IMC des répondants âgés de 12 à 17 ans peuvent désormais être présentées;
- ii) les données correspondant à des variables recueillies uniquement auprès des répondants âgés de 18 ans et plus ne sont plus supprimées automatiquement si elles se rapportent à des répondants âgés de 18 ou 19 ans;
- iii) les utilisateurs ont quand même la possibilité d'effectuer des comparaisons avec des données sur les jeunes tirées de FMGD antérieurs en constituant le groupe des 15 à 19 ans par la fusion de ceux des 15 à 17 ans et des 18 et 19 ans.

Bien que l'information sur certaines variables ait été recueillie auprès de personnes appartenant à une catégorie d'âge particulière, il a néanmoins fallu, pour des motifs de confidentialité, supprimer des données se rapportant à des personnes d'âges donnés. À titre d'exemple, des données sur la population active ont été recueillies auprès de répondants âgés de 15 à 75 ans. Comme le FMGD comporte un groupe de 75 à 79 ans, la publication de données sur la population active visant des répondants de 75 ans aurait pour effet de révéler implicitement leur âge exact. Pour cette raison, le FMGD ne présente des données sur la population active (LBSE) que pour les répondants âgés de 15 à 74 ans. Dans le même ordre d'idées, les données sur l'éducation (EDUE) des répondants âgés de 14 ans et celles qui concernent les expériences maternelles (MEXE) des répondantes de 55 ans ont aussi été éliminées.

Exceptionnellement, les données sur les expériences maternelles (MEXE) des adolescentes de 15 à 17 ans ont été exclues à cause du risque élevé de divulgation. Ainsi, bien que le fichier maître présente des données sur les expériences maternelles de toutes les femmes âgées de 15 à 55 ans, le FMGD ne contient que celles qui ont trait aux femmes âgées de 18 à 54 ans.

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 facteurs de pondération, appelés ici les poids d'enquête, dans ses calculs. Un poids d'enquête est attribué à chaque personne incluse dans l'échantillon final, c'est-à-dire dans l'échantillon de personnes ayant répondu à l'enquête. Ce poids correspond au nombre de personnes représentées par le répondant dans l'ensemble de la population.

Tel que décrit dans la section 5, le cycle 3.1 de l'ESCC a eu recours à trois bases de sondage pour la sélection de son échantillon : une base aréolaire de logements agissant comme base principale, puis deux bases formées de numéros de téléphone utilisées pour compléter la base aréolaire. Puisque seulement quelques différences mineures distinguent les deux bases de numéros de téléphone pour la pondération, elles ont été traitées ensemble. On réfère à celles-ci comme faisant partie de la base téléphonique.

La stratégie de pondération a été développée en traitant séparément la base aréolaire et la base téléphonique. Les poids résultant de ces deux bases sont ensuite combinés en un seul ensemble de poids lors d'une étape appelée « intégration ». Suite à quelques ajustements, ce poids intégré devient le poids final. Noter que dépendamment du besoin, une seule ou deux bases pouvaient être utilisées pour la sélection de l'échantillon dans une région sociosanitaire donnée. La stratégie de pondération s'occupe de cette particularité lors de l'étape d'intégration.

8.1 Pondération de l'échantillon

Tel que mentionné plus haut, les unités des bases aréolaire et téléphonique sont traitées séparément jusqu'à l'étape d'intégration. La sous-section 8.1.1 fournit les détails de la stratégie de pondération pour la base aréolaire, puis la sous-section 8.1.2, ceux pour la base téléphonique. L'intégration des deux bases est traitée en 8.1.3, puis suivent les deux étapes finales de la pondération, c'est-à-dire l'ajustement pour contrôler la saisonnalité des données puis la poststratification, qui sont expliquées dans les sous-sections 8.1.4 et 8.1.5 respectivement.

Malgré que les deux bases aient été utilisées pour couvrir les trois territoires, certaines modifications ont dû être faites relativement à leur utilisation. Ces modifications affectent substantiellement la pondération pour ces trois régions, et celles-ci sont rapportées dans la sous-section 8.1.6.

Le diagramme A présente un sommaire des différents ajustements faisant partie de la stratégie de pondération dans l'ordre qu'ils sont appliqués. Un système de numérotation est utilisé pour identifier chaque ajustement apporté au poids et sera utilisé tout au long de la section. Les lettres *A* et *T* sont utilisées comme préfixes pour référer aux ajustements appliqués aux unités des bases Aréolaire et Téléphonique respectivement. Le préfixe *I* est quant à lui utilisé pour identifier l'ajustement d'Intégration et ceux qui suivent.

Diagramme A Sommaire de la stratégie de pondération

Base aréolaire	Base téléphonique
A0 - Poids initial	T0 - Poids initial
A1 - Accroissement de l'échantillon	T1 - Nombre de mois
A2 - Stabilisation	T2 - Retrait des unités hors champ
A3 - Retrait des unités hors champ	T3 - Couverture des bases listes
A4 - Non-réponse ménage	T4 - Combinaison des bases listes
A5 - Création du poids-personne	T5 - Non-réponse ménage
A6 - Non-réponse personne	T6 - Ménages sans téléphone
Poids final de la base aréolaire	T7 - Création du poids-personne
↗	T8 - Non-réponse personne
	T9 - Lignes multiples
	Poids final de la base téléphonique
	↘
	I1 – Intégration
	I2 - Effet saisonnier
	I3 – Poststratification
	Poids final du cycle 3.1 de l' ESCC

8.1.1 Pondération de l'échantillon provenant de la base aréolaire

A0 – Poids initial

Puisque le mécanisme utilisé pour sélectionner l'échantillon de la base aréolaire a été celui établi pour l'EPA, le poids initial a dû être calculé selon les particularités de cette enquête. D'abord, à l'intérieur de chacune des strates définies par l'EPA, des grappes (unités primaires) sont sélectionnées avec probabilités proportionnelles au nombre de ménages (selon les comptes du Recensement de 2001). À l'intérieur de chacune des grappes sélectionnées, un échantillon de logements est ensuite choisi à l'aide d'un échantillonnage systématique. Le produit des probabilités de chacune de ces sélections représente la probabilité de sélection globale et son inverse représente le poids initial du cycle 3.1 de l'ESCC. Pour plus de détails sur le mécanisme de sélection, de même qu'une définition plus complète des strates et des grappes, se référer à Statistique Canada (1998)⁶.

A1 – Accroissement de l'échantillon

Certaines modifications ont dû être faites au mécanisme standard de l'EPA lors de la sélection de l'échantillon pour le cycle 3.1 de l'ESCC. Le plan de l'EPA peut fournir un échantillon d'environ 60 200 logements à l'échelle du pays alors que les besoins du cycle 3.1 de l'ESCC sont plus grands dans certaines régions. Les modifications apportées afin de pouvoir obtenir la taille d'échantillon

⁶ Statistique Canada (1998). *Méthodologie de l'enquête sur la population active du Canada*. Statistique Canada numéro 71-526-XPB au catalogue.

désirée ont été, en résumé, de répéter le même processus de sélection une deuxième fois à l'intérieur de toutes les grappes sélectionnées dans la RSS où le besoin en échantillon se faisait sentir. Ceci a eu l'effet d'accroître l'échantillon et on a dû en tenir compte dans la pondération afin de correctement représenter la probabilité de sélection. Un facteur d'ajustement représentant le taux d'accroissement de l'échantillon a donc été calculé. Cependant, cet accroissement de l'échantillon n'a pas été observé dans toutes les régions. En effet, pour certaines régions, le plan de l'EPA conduit à des tailles d'échantillon supérieures à celles requises par l'ESCC. Pour ces régions, le taux d'accroissement de l'échantillon qui est calculé représente plutôt un taux de décroissement. Le poids initial calculé en A0 est multiplié par ce facteur d'ajustement, ce qui permet d'obtenir le poids A1.

A2 – Stabilisation

Dans certaines RSS, l'accroissement de l'échantillon tel que décrit au paragraphe précédent résultait en un échantillon beaucoup plus grand que nécessaire. Une stabilisation a donc été instaurée afin de ramener la taille de l'échantillon au niveau désiré. Le processus de stabilisation consistait à sous-échantillonner des logements aléatoirement à l'intérieur de la RSS. Un facteur d'ajustement représentant l'effet de la stabilisation a donc été calculé afin de corriger la probabilité de sélection. Ce facteur, multiplié par le poids A1, produit le poids A2.

A3 – Retrait des unités hors champ

Parmi tous les logements échantillonnés, une certaine proportion de ceux-ci est, lors de la collecte, identifiée comme étant hors du champ de l'enquête. Des logements détruits ou en construction, des logements vacants, saisonniers ou secondaires, de même que des établissements, sont tous des exemples de cas hors champ pour le cycle 3.1 de l'ESCC. Ces logements sont tout simplement retirés de l'échantillon, ne laissant plus que les logements faisant partie du champ de l'enquête. Ces derniers conservent donc le même poids qu'à l'étape précédente que l'on appelle maintenant poids A3.

A4 – Non-réponse ménage

Lors de la collecte, une certaine proportion des ménages interviewés a inévitablement résulté en non-réponse. Ceci survient habituellement lorsque le ménage refuse de participer à l'enquête, fournit des données inutilisables, ou encore, ne peut être rejoint pour réaliser l'interview. Le poids des ménages non-répondants est redistribué aux répondants à l'aide de classes de réponse. L'algorithme CHAID (Chi-Square Automatic Interaction Detector), disponible dans Knowledge Seeker⁷, permet d'identifier les caractéristiques qui divisent le mieux l'échantillon en groupes selon leurs propensions à répondre. Noter que ces groupes sont formés de façon indépendante à l'intérieur de chaque RSS. Puisque l'information disponible auprès des non-répondants est très limitée, seules quelques caractéristiques telles que la période de collecte et un indicateur du milieu rural/urbain ont pu être utilisées pour la création des classes. Il s'est avéré que l'indicateur rural/urbain était la caractéristique la plus significative pour la création des classes pour chacune

⁷ ANGOSS Software (1995). Knowledge Seeker IV for Windows - User's Guide. ANGOSS Software International Limited.

des RSS. La période de collecte (avec 4 périodes: janvier à mars, avril à mai, juin à août et septembre à novembre) était également une caractéristique significative pour un petit nombre de RSS. Un facteur d'ajustement a donc été calculé à l'intérieur de chaque classe de la façon suivante:

$$\frac{\text{Somme des poids A3 pour tous les ménages}}{\text{Somme des poids A3 pour tous les ménages répondants}}$$

Le poids A3 des ménages répondants a donc été multiplié par ce facteur d'ajustement pour produire le poids A4. Les ménages non-répondants sont éliminés du processus de pondération à partir de ce point.

A5 – Création du poids-personne

Puisque l'unité d'échantillonnage finale pour l'ESCC est la personne, le poids-ménage calculé jusqu'ici doit être converti en un poids-personne. Celui-ci est obtenu en multipliant le poids A4 par l'inverse de la probabilité de sélection de la personne choisie dans le ménage. Nous obtenons ainsi le poids A5. Rappelons que la probabilité de sélection de la personne change en fonction du nombre de personnes dans le ménage et de l'âge des individus (voir section 5.5 pour plus de détails).

A6 – Non-réponse personne

Dans le cadre du cycle 3.1 de l'ESCC, une interview peut être vue comme un processus en deux étapes. Dans un premier temps, l'intervieweur obtient la liste complète des personnes vivant dans le ménage, puis par la suite interviewe la personne sélectionnée dans le ménage. Dans certains cas, les intervieweurs ne réussissent qu'à compléter la première étape, soit parce qu'ils ne peuvent entrer en contact avec la personne sélectionnée, ou encore parce que la personne sélectionnée refuse d'être interviewée. De tels cas sont définis comme étant des non-réponses à l'échelle de la personne, et un facteur d'ajustement doit être appliqué aux poids des personnes répondantes pour compenser cette non-réponse. Tout comme pour la non-réponse à l'échelle du ménage, l'ajustement est appliqué à l'intérieur de classes définies à partir des caractéristiques disponibles pour les répondants et non-répondants. Toutes les caractéristiques recueillies lors du listage des membres du ménage étaient en fait disponibles pour créer ces classes. L'algorithme CHAID, disponible dans Knowledge Seeker, a encore une fois été utilisé pour obtenir la définition des classes et le résultat final présente quelques variations dans la définition des classes d'une RSS à l'autre. Selon la RSS, les caractéristiques suivantes ont pu être utilisées pour former les classes d'ajustement : les tailles des ménages, le sexe, l'indicateur de milieu rural/urbain, la période de collecte, l'état matrimonial, l'âge, les dispositions de vie et le nombre de personnes de 12 ans et moins dans le ménage. Un facteur d'ajustement est calculé à l'intérieur de chaque classe de la façon suivante:

$$\frac{\text{Somme des poids A5 pour toutes les personnes sélectionnées}}{\text{Somme des poids A5 pour toutes les personnes sélectionnées répondantes}}$$

Le poids A5 des personnes répondantes a donc été multiplié par ce facteur d'ajustement pour produire le poids A6. Les personnes non-répondantes sont éliminées de la pondération à partir de ce point.

Étant donné que cet ajustement est le dernier nécessaire pour l'échantillon provenant de la base aréolaire, le poids A6 représente donc le *poids final de la base aréolaire*. Ce poids sera plus tard intégré au poids final de la base téléphonique (section 8.1.3) pour créer le poids final du cycle 3.1 de l'ESCC.

8.1.2 Pondération de l'échantillon provenant de la base téléphonique

Tel que mentionné précédemment, la base téléphonique est en fait composée de deux bases : la base de sondage à composition aléatoire (CA), puis une base liste de numéros de téléphone. Noter qu'une seule de ces deux bases peut être utilisée à l'intérieur d'une RSS. La base liste est toujours utilisée comme complément à la base aréolaire tandis que la base CA est toujours utilisée seule. Les unités provenant de ces deux bases téléphoniques sont toutefois traitées ensemble et sont donc toutes soumises aux mêmes ajustements. Il existe toutefois trois exceptions: d'abord, puisque la probabilité de sélection est relative à la base utilisée pour faire la sélection, cette probabilité sera légèrement différente dépendamment que l'unité provienne de la base CA ou de la base liste. Les autres exceptions impliquent les ajustements T3 et T4. Les détails de ces exceptions sont donnés dans les sous-sections réservées aux ajustements impliqués.

Une autre particularité propre aux unités provenant de la base téléphonique affecte la façon dont l'échantillon a été pondéré. Cette particularité concerne l'emplacement géographique des unités échantillonnées. En effet, la géographie utilisée pour sélectionner l'échantillon à partir de la base téléphonique ne répliquait pas parfaitement la géographie des RSS, ce qui a forcé certaines unités à être sélectionnées dans une certaine région alors que l'information fournie lors de l'interview les localisait plutôt dans une région avoisinante. Cette particularité a été contournée lors de la pondération en appliquant tous les ajustements relativement à la RSS assignée lors de la sélection de l'échantillon. Toutefois, puisque les unités devaient en bout de ligne appartenir à leur vraie RSS, telle qu'identifiée lors de la collecte, on a dû ajuster les poids de celles-ci comme si elles avaient fait partie de leur vraie région dès la sélection de l'échantillon. Cet ajustement a été fait via la poststratification (I3) qui est traitée plus tard dans cette section.

T0 – Poids initial

Le poids initial est calculé quelque peu différemment selon que l'échantillon provienne de la base CA ou de la base liste. Dans les deux cas, le poids initial est défini comme étant l'inverse de la probabilité de sélection, mais puisque les méthodes de sélection diffèrent, les probabilités diffèrent aussi. Pour la base CA, la sélection des numéros est faite à l'intérieur de chaque strate CA. Une strate CA représente un agrégat d'indicatifs régionaux et préfixes (IRP: les six premiers chiffres du numéro à 10 chiffres), contenant chacune des banques valides de cent numéros (voir Norris et Paton⁸ pour plus de détails). Conséquemment, la probabilité de sélection est le ratio entre le

⁸ Norris, D.A. et Paton, D.G. (1991). L'Enquête sociale générale canadienne: bilan des cinq premières années. *Techniques d'enquête*. 17, 245-260.

nombre d'unités échantillonnées et cent fois le nombre de banques présentes dans la strate CA. Pour la base liste, les numéros de téléphone sont sélectionnés parmi tous les numéros disponibles dans la liste, et ce indépendamment pour chaque RSS. Ainsi, la probabilité de sélection correspond au ratio entre le nombre d'unités échantillonnées et le nombre de numéros de téléphone dans la liste pour la RSS. Puisque l'échantillonnage pour la base téléphonique est fait sur une base mensuelle (voir ajustement T1) et que la base liste a été mise à jour durant l'enquête, le nombre de numéros disponibles dans chaque RSS a pu légèrement varier affectant ainsi la probabilité de sélection dans le temps. L'inverse de ces probabilités de sélection représente le poids initial T0.

T1 – Nombre de mois

Contrairement à la base aréolaire, pour laquelle l'échantillon a été sélectionné entièrement au début du processus d'échantillonnage, des échantillons ont été tirés à chaque mois pour la base téléphonique. À chacun de ces échantillons mensuels correspond un poids initial faisant en sorte que chaque échantillon soit représentatif de la RSS. Toutefois, pour que l'échantillon total ne représente qu'une seule fois la population, un facteur d'ajustement a dû être appliqué pour réduire les poids de chaque échantillon mensuel. Le facteur d'ajustement appliqué à chaque échantillon mensuel était égal à la proportion que représentait cet échantillon mensuel parmi l'échantillon total. Or, puisque deux versions différentes de la base liste ont été utilisées et que chaque version possède une couverture différente, l'ajustement du nombre de mois a été calculé indépendamment pour chaque version de la base liste. À partir de ce moment, l'échantillon de la base liste représente deux fois la population totale, soit une fois pour chaque version de la base liste. Les échantillons provenant des deux bases listes sont combinés à l'étape T4 de telle sorte que la base téléphonique représente alors une seule fois la population. Le poids T1 est donc obtenu en multipliant le poids T0 par le facteur d'ajustement défini ci-dessus.

T2 - Retrait des unités hors champ

Les numéros de téléphone associés à des entreprises, des établissements ou à d'autres logements hors du champ de l'enquête, de même que les numéros hors service sont tous des exemples de cas hors champ pour la base téléphonique. Comme pour la base aréolaire, ces cas sont simplement retirés de l'échantillon, ne laissant ainsi dans l'échantillon que les logements dans le champ de l'enquête. Ces derniers conservent le même poids qu'à l'étape précédente que l'on appelle maintenant poids T2.

T3 – Couverture des bases listes

Puisque la base liste ne couvre pas certains numéros de téléphone qui sont toutefois couverts par la base CA, un ajustement doit être apporté au poids initial des unités de la base liste pour que les deux bases soient comparables en ce qui a trait à la couverture. Cet ajustement consiste à gonfler le poids des unités de la base liste proportionnellement au taux de couverture dans chaque RSS. L'estimation de ce taux de couverture a été une tâche ardue, et a pu être faite à l'aide des données recueillies auprès de l'échantillon de la base aréolaire. En effet, le questionnaire utilisé pour l'interview des personnes sélectionnées par la base aréolaire incluait un ensemble de questions vérifiant la présence d'un téléphone dans le logement du répondant, le nombre de lignes utilisées à

des fins personnelles, puis le numéro pour chacune de ces lignes. Pour dériver le taux de couverture désiré, on a simplement calculé le pourcentage des numéros de téléphone recueillis étant présents sur la base liste. Ce taux a été calculé pour chaque version de la base liste étant donné que la couverture diffère d'une version à l'autre. L'inverse de ce taux représente le facteur utilisé pour cet ajustement. Le facteur, une fois multiplié par le poids T2, produit le poids T3.

T4 - Combinaison des bases listes

L'échantillon provenant de chaque version de la base liste représente jusqu'à cette étape, la population totale des RSS où la base liste est utilisée. Les unités des deux versions doivent donc être combinées afin qu'elles ne représentent qu'une seule fois la population. Pour se faire, un facteur de combinaison tenant compte de l'importance de chaque version est calculé de la façon suivante :

$$\frac{\textit{Taille d' échantillon provenant de la version de la base liste}}{\textit{Taille totale d' échantillon provenant de la base liste}}$$

Ce facteur est calculé et appliqué indépendamment dans chaque RSS où la base liste a été utilisée. Pour les régions où c'est la base CA qui est utilisée, le facteur d'ajustement est égal à 1. Le poids T4 est obtenu en multipliant le poids T3 par le facteur de combinaison.

T5 - Non-réponse ménage

L'ajustement fait ici pour compenser l'effet de la non-réponse ménage est identique à celui appliqué pour la base aréolaire (ajustement A4). Comme c'était le cas pour A4, la période de collecte s'est avérée une caractéristique significative pour expliquer la non-réponse. C'est donc cette variable qui a été utilisée pour définir les classes d'ajustement. Le facteur d'ajustement calculé à l'intérieur de chaque classe a été obtenu de la façon suivante:

$$\frac{\textit{Somme des poids T4 pour tous les ménages}}{\textit{Somme des poids T4 pour tous les ménages répondants}}$$

Le poids T4 des ménages répondants a donc été multiplié par ce facteur d'ajustement pour produire le poids T5. Les ménages non-répondants sont éliminés à partir de ce point.

T6 - Ménages sans téléphone

Une certaine proportion de la population canadienne n'a pas accès à une ligne téléphonique résidentielle privée. Tel qu'expliqué à l'étape T3, de l'information concernant la présence d'un téléphone dans le logement du répondant est recueillie auprès de l'échantillon de la base aréolaire. Cette information a été utilisée pour estimer la proportion de ménages n'ayant pas le téléphone à l'échelle de chaque RSS. Tout comme pour T3, cette proportion est ensuite utilisée pour gonfler le poids des unités de la base téléphonique, ajustant ainsi pour la sous-représentation de la base due à cette sous-population non couverte. Le facteur utilisé pour cet ajustement correspond à l'inverse de la proportion estimée, et une fois multiplié par le poids T5, procure le poids T6.

T7 – Création du poids-personne

Tout comme l'ajustement A5, cet ajustement permet de convertir ce qui était jusqu'à cette étape-ci un poids-ménage en un poids-personne. L'algorithme de sélection de la personne à l'intérieur du ménage étant le même que pour la base aréolaire, le calcul du facteur d'ajustement est effectué de la même façon. Ce facteur, multiplié par le poids T6, donne le poids T7.

T8 - Non-réponse personne

Cet ajustement est similaire à l'ajustement A6 utilisé pour la base aréolaire. Il consiste à compenser pour l'effet de la non-réponse à l'échelle de la personne. Tout comme pour A6, une approche par classes d'ajustement a été utilisée. Ces classes étaient définies à partir des variables disponibles pour toutes les personnes sélectionnées, répondantes ou non (voir A6 pour la liste des variables disponibles). Un facteur d'ajustement a donc été calculé à l'intérieur de chaque classe de la façon suivante:

$$\frac{\text{Somme des poids T7 pour toutes les personnes sélectionnées}}{\text{Somme des poids T7 pour toutes les personnes sélectionnées répondantes}}$$

Le poids T7 des personnes répondantes a donc été multiplié par ce facteur d'ajustement pour produire le poids T8. Les personnes non-répondantes sont éliminées à partir de ce point.

T9 - Lignes multiples

Le fait que certains ménages possèdent plus d'une ligne téléphonique résidentielle a un impact sur la pondération: plus le ménage a de lignes, meilleure est sa probabilité d'être sélectionné. Conséquemment, les poids doivent être ajustés pour tenir compte du nombre de lignes résidentielles que le ménage possède. Malgré que cette caractéristique soit relative au ménage, l'information n'est recueillie que durant l'interview auprès de la personne. C'est pour cette raison que l'ajustement est fait à ce stade-ci de la pondération. Le facteur d'ajustement représente l'inverse du nombre de lignes dans le ménage et le poids T9 est obtenu en multipliant ce facteur par le poids T8.

Puisque cet ajustement est le dernier nécessaire pour l'échantillon provenant de la base téléphonique, le poids T9 représente donc le **poids final de la base téléphonique**. Ce poids sera par la suite, à l'étape II, intégré au poids final de la base aréolaire pour finalement créer le poids final du cycle 3.1 de l'ESCC.

8.1.3 Intégration des bases aréolaire et téléphonique (I1)

Cette étape consiste à intégrer les poids finaux des échantillons aréolaire et téléphonique créés jusqu'à maintenant, en un seul poids en appliquant une méthode d'intégration⁹. Un facteur d'ajustement, compris entre 0 et 1, est déterminé au niveau de RSS de façon à représenter l'importance relative de chaque échantillon dans l'échantillon total. Cette importance relative est mesurée en termes de taille d'échantillon et d'effet de plan. Plus la proportion d'échantillon qu'une base représente dans l'échantillon total est grande, plus grande sera son importance relative dans l'échantillon total. Pour ce qui de l'effet de plan, l'importance relative sera plus grande pour les unités provenant de la base dont l'effet de plan est plus petit. Pour obtenir le facteur d'ajustement d'intégration, on calcule d'abord un facteur α , obtenu de la façon suivante:

$$\alpha = \frac{n_A}{R} \bigg/ \left(\frac{n_A}{R} + n_T \right)$$

où n_A et n_T représentent respectivement les tailles d'échantillon des bases aréolaire et téléphonique, alors que R représente le rapport médian des effets de plan estimés de plusieurs caractéristiques clés pour chacune des deux bases. Le poids des unités de la base aréolaire est multiplié par ce facteur α , alors que le poids des unités de la base téléphonique est multiplié par $1 - \alpha$. Il est à noter que dans les cas où une RSS n'est couverte que par une seule base, le facteur d'ajustement est égal à 1. Le produit du facteur d'ajustement dérivé ici, par le poids final calculé auparavant (A6 ou T9 dépendant de quelle base provient l'unité), procure le poids intégré I1.

8.1.4 Effet saisonnier et Winsorization (I2)

Le cycle 3.1 de l'ESCC avait initialement planifié de répartir la collecte des données également sur les douze mois de l'année de référence de l'enquête afin de contrôler entre autres l'effet saisonnier des données recueillies. Certains événements ont toutefois affecté ce plan, de sorte qu'un ajustement additionnel a dû être ajouté pour assurer qu'il n'y ait pas d'effet saisonnier dans les estimations produites à l'aide des données du cycle 3.1 de l'ESCC. L'ajustement appliqué en I2 a été fait de façon à ce que la somme des poids des unités interviewées lors d'une des quatre saisons, représente exactement 25 % de la somme des poids de l'échantillon total. Bref, après l'application de cet ajustement, la portion d'échantillon interviewée à chaque saison représente 25 % de la population totale de chaque RSS.

Les quatre saisons définies pour l'ESCC sont les périodes couvrant septembre à novembre, décembre à février, mars à mai, puis juin à août. Le facteur d'ajustement utilisé pour contrôler l'effet saisonnier d'une personne interviewée lors de la saison S , est défini comme suit:

$$\frac{\text{Somme des poids } I1 \text{ pour l'échantillon total}}{4 \times \text{somme des poids } I1 \text{ de l'échantillon interviewé lors de la saison } S}$$

⁹ Skinner, C.J. et Rao, J.N.K. (1996). Estimation in Dual Frame Surveys with Complex Designs. *Journal of the American Statistical Association*. 91, 433, 349-356.

Cet ajustement saisonnier appliqué au poids I1 permet d'obtenir le poids I2.

Noter que suite à la série d'ajustements appliqués sur les poids, il est possible que certaines unités se retrouvent avec des poids se démarquant des autres poids de leur RSS au point même de devenir aberrants. Certains répondants peuvent effectivement représenter une proportion anormalement élevée de leur RSS et ainsi influencer fortement les estimations de ces RSS. Afin d'éviter cette situation, le poids des répondants qui contribuent de façon aberrante à leur groupe RSS-âge-sexe est ajusté à la baisse selon une méthode « winsorization ».

8.1.5 Poststratification (I3)

La dernière étape nécessaire afin d'obtenir le poids final du cycle 3.1 de l' ESCC est la poststratification. La poststratification est appliquée afin d'assurer que la somme des poids finaux corresponde aux estimations de populations définies à l'échelle des RSS, pour chacun des 10 groupes d'âge-sexe d'intérêt, c'est-à-dire les cinq groupes d'âge 12-19, 20-29, 30-44, 45-64, 65+, pour chacun des deux sexes. Il est à noter que la poststratification a été faite en utilisant une géographie révisée contenant 4 régions au lieu des 6 utilisées initialement lors de la planification et du déroulement de l'enquête à Terre-Neuve-et-Labrador, et une région sociosanitaire de moins pour l'Ontario. Aussi, pour les trois régions du Québec qui ont fait l'achat d'échantillon supplémentaire (2403, 2407 et 2415), la poststratification a été appliquée à l'échelle de la région de CLSC plutôt que par RSS.

Les estimations de population de 2005 utilisées sont basées sur les comptes du Recensement de 2001, de même que sur les comptes de naissance, décès, immigration et émigration depuis ce temps. La moyenne des estimations mensuelles de 2003 pour chacun des croisements RSS-âge-sexe a été retenue pour réaliser la poststratification. Le poids I2 a donc été ajusté afin d'obtenir le poids final I3, à l'aide du facteur d'ajustement I3 défini comme suit:

$$\frac{\text{Estimation de population pour le groupe RSS - âge - sexe du répondant}}{\text{Somme des poids I2 pour le groupe RSS - âge - sexe du répondant}}$$

Le poids I3 correspond au **poids final du cycle 3.1 de l' ESCC** que l'on retrouve dans le fichier de données portant le nom de variable WTSE_M.

8.1.6 Particularités de la pondération pour les trois territoires

Tel que décrit à la section 5, le plan d'échantillonnage utilisé pour les trois territoires était quelque peu différent de celui utilisé dans les 10 provinces. La stratégie de pondération a donc dû être adaptée pour répondre à ces différences. Cette section résume les changements apportés à la stratégie expliquée aux sous-sections 8.1.1 à 8.1.5.

D'abord pour la base aréolaire, tel que mentionné à la sous-section 5.4.1, une étape additionnelle de sélection a été ajoutée pour les territoires. Chaque territoire était initialement stratifié selon des regroupements de communautés à l'intérieur desquels on a sélectionné aléatoirement une communauté. Noter que les capitales de chaque territoire formaient une strate à elles seules, et

étaient donc toutes trois sélectionnées automatiquement à cette première sélection. Cette particularité n'a eu d'effet que dans le calcul de la probabilité de sélection, et donc dans la valeur du poids initial (A0). Une fois ce poids initial calculé, la même série d'ajustements (A1 à A6) a été appliquée aux unités de la base aréolaire. Les classes d'ajustement pour les non-réponses ménage et personne ont été construites à l'aide du même ensemble de variables disponibles pour les provinces.

Pour ce qui est de la pondération des unités de la base téléphonique, mentionnons tout d'abord que seule la base CA a été utilisée, et ce, uniquement à l'intérieur des capitales du Yukon et des Territoires du Nord-Ouest. Ceci élimine donc le besoin d'avoir recours aux ajustements T3 (couverture des bases listes) et T4 (combinaison des bases listes). Les autres ajustements de la base téléphonique ont tous été appliqués. Finalement, l'ajustement T6 (ménages sans téléphone) a aussi subi une légère modification puisque la base CA était utilisée uniquement dans les capitales. Les proportions de ménages sans téléphone ont été dérivées, tout comme pour les provinces, à partir des données de la base aréolaire, mais en excluant toutefois du calcul les données des ménages situés à l'extérieur des capitales.

Les deux ensembles de poids (aréolaire et téléphonique) ont ensuite été intégrés, puis ajustés pour la saisonnalité et finalement poststratifiés de façon semblable à ce qui a été fait pour les provinces, à l'exception de deux détails. D'abord, l'intégration a été appliquée uniquement pour les unités situées dans les capitales du Yukon et des Territoires du Nord-Ouest; les autres communautés ayant été couvertes uniquement par la base aréolaire. Un autre chose à noter pour le Nunavut est que les comptes de la population utilisés pour la calibration représentent 70% de la population seulement étant donné la sous couverture de la base aréolaire, telle que décrite à la section 5.4.1.

9. Qualité des données

9.1 Taux de réponse

Au total et après avoir retiré les unités hors du champ de l'enquête, 168 464 ménages ont été sélectionnés pour participer au cycle 3.1 de l'ESCC. De ce nombre, 143 076 ont accepté de participer à l'enquête ce qui résulte en un taux de réponse à l'échelle du ménage de 84,9 %. Parmi ces ménages répondants, 143 076 personnes ont été sélectionnées (une personne par ménage) pour participer à l'enquête parmi lesquelles 132 947 ont accepté de le faire ce qui résulte en un taux de réponse à l'échelle de la personne de 92,9 %. À l'échelle canadienne, un taux de réponse combiné de 78,9 % a donc été observé au cycle 3.1 de l'ESCC. Le tableau 9.1 donne les taux de réponse combinés ainsi que l'information pertinente au calcul de ceux-ci pour chaque région sociosanitaire ou regroupement de régions sociosanitaires.

On décrit dans ce qui suit de quelle façon les différentes composantes de l'équation doivent être manipulées afin de calculer correctement les taux de réponse combinés.

Taux de réponse à l'échelle du ménage

$$\text{HHRR} = \frac{\text{\# de ménages répondants provenant des 2 bases}}{\text{tous les ménages faisant partie du champ de l'enquête provenant des 2 bases}}$$

Taux de réponse à l'échelle de la personne

$$\text{PPRR} = \frac{\text{\# de répondants provenant des 2 bases}}{\text{toutes les personnes sélectionnées provenant des 2 bases}}$$

$$\text{Taux de réponse combiné} = \text{HHRR} \times \text{PPRR}$$

Voici maintenant un exemple de calcul du taux de réponse combiné pour le Canada en utilisant l'information fournie dans le tableau 9.1.

$$\text{HHRR} = \frac{69\,417 + 73\,659}{78\,396 + 90\,068} = \frac{143\,076}{168\,464} = 0,849$$

$$\text{PPRR} = \frac{65\,039 + 67\,908}{69\,417 + 73\,659} = \frac{132\,947}{143\,076} = 0,929$$

$$\begin{aligned} \text{Taux de réponse combiné} &= 0,849 \times 0,929 \\ &= 0,789 \\ &= \mathbf{78,9\%} \end{aligned}$$

Table 9.1 Tableau 9.1		Area frame / Base aréolaire							Phone frames / Bases téléphoniques							All cases / Tous les cas
Prov./ Terr.	Health Region	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp. pers.	Pers. resp. rates	Resp. rates	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp. pers.	Pers. resp. rates	Resp. rates	Combined resp. rates
Prov./ Terr.	Région socio- sanitaire	# mén. cibles	# mén. rép.	Taux de rép. mén.	# pers. sélect.	# rép.	Taux de rép. pers.	Taux de rép.	# mén. cibles	# mén. rép.	Taux de rép. mén.	# pers. sélect.	# rép.	Taux de rép. pers.	Taux de rép.	Taux de rép. combiné
CA	Total	78396	69417	88.5	69417	65039	93.7	83.0	89040	72850	81.8	72850	67182	92.2	75.5	79.0
NL	Total	2283	2112	92.5	2112	2001	94.7	87.6	2516	2271	90.3	2271	2110	92.9	83.9	85.7
	10911	1050	966	92.0	966	917	94.9	87.3	1076	979	91.0	979	897	91.6	83.4	85.3
	10912	416	383	92.1	383	362	94.5	87.0	465	420	90.3	420	385	91.7	82.8	84.8
	10913*	817	763	93.4	763	722	94.6	88.4	975	872	89.4	872	828	95.0	84.9	86.5
PE	Total	1157	1049	90.7	1049	995	94.9	86.0	1279	1120	87.6	1120	1036	92.5	81.0	83.4
	11901*	691	636	92.0	636	608	95.6	88.0	851	749	88.0	749	687	91.7	80.7	84.0
	11903	466	413	88.6	413	387	93.7	83.0	428	371	86.7	371	349	94.1	81.5	82.3
NS	Total	2944	2607	88.6	2607	2417	92.7	82.1	3098	2806	90.6	2806	2649	94.4	85.5	83.8
	12901	438	409	93.4	409	378	92.4	86.3	495	450	90.9	450	422	93.8	85.3	85.7
	12902	408	369	90.4	369	350	94.9	85.8	345	317	91.9	317	306	96.5	88.7	87.1
	12903	414	378	91.3	378	350	92.6	84.5	439	396	90.2	396	368	92.9	83.8	84.2
	12904	363	330	90.9	330	312	94.5	86.0	479	431	90.0	431	401	93.0	83.7	84.7
	12905	481	429	89.2	429	394	91.8	81.9	512	458	89.5	458	442	96.5	86.3	84.2
	12906	840	692	82.4	692	633	91.5	75.4	828	754	91.1	754	710	94.2	85.7	80.5
NB	Total	2982	2706	90.7	2706	2542	93.9	85.2	3117	2748	88.2	2748	2558	93.1	82.1	83.6
	13901	599	525	87.6	525	493	93.9	82.3	572	506	88.5	506	475	93.9	83.0	82.7
	13902	558	501	89.8	501	469	93.6	84.1	616	550	89.3	550	511	92.9	83.0	83.5
	13903	567	511	90.1	511	486	95.1	85.7	558	486	87.1	486	461	94.9	82.6	84.2
	13904*	558	518	92.8	518	487	94.0	87.3	642	563	87.7	563	519	92.2	80.8	83.8
	13906*	700	651	93.0	651	607	93.2	86.7	729	643	88.2	643	592	92.1	81.2	83.9
QC	Total	14674	12895	87.9	12895	12124	94.0	82.6	23498	18973	80.7	18973	17041	89.8	72.5	76.4
	24901	840	785	93.5	785	765	97.5	91.1	3579	3012	84.2	3012	2753	91.4	76.9	79.6
	24902	745	674	90.5	674	645	95.7	86.6	737	638	86.6	638	585	91.7	79.4	83.0
	24903	1181	1020	86.4	1020	973	95.4	82.4	1219	1003	82.3	1003	910	90.7	74.7	78.5
	24904	983	883	89.8	883	851	96.4	86.6	1101	904	82.1	904	831	91.9	75.5	80.7
	24905	827	701	84.8	701	655	93.4	79.2	765	646	84.4	646	578	89.5	75.6	77.4
	24906	2017	1698	84.2	1698	1570	92.5	77.8	5793	4346	75.0	4346	3820	87.9	65.9	69.0
	24907	793	663	83.6	663	626	94.4	78.9	831	692	83.3	692	630	91.0	75.8	77.3
	24908	619	570	92.1	570	530	93.0	85.6	799	687	86.0	687	640	93.2	80.1	82.5
	24909	763	665	87.2	665	633	95.2	83.0	702	603	85.9	603	542	89.9	77.2	80.2
	24911	701	656	93.6	656	619	94.4	88.3	784	654	83.4	654	595	91.0	75.9	81.8
	24912	854	776	90.9	776	736	94.8	86.2	927	763	82.3	763	678	88.9	73.1	79.4

Table 9.1 Tableau 9.1		Area frame / Base aréolaire							Phone frames / Bases téléphoniques							All cases / Tous les cas
Prov./ Terr.	Health Region	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp. pers.	Pers. resp. rates	Resp. rates	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp. pers.	Pers. resp. rates	Resp. rates	Combined resp. rates
Prov./ Terr.	Région socio- sanitaire	# mén. cibles	# mén. rép.	Taux de rép. mén.	# pers. sélect.	# rép.	Taux de rép. pers.	Taux de rép.	# mén. cibles	# mén. rép.	Taux de rép. mén.	# pers. sélect.	# rép.	Taux de rép. pers.	Taux de rép.	Taux de rép. combiné
	24913	950	832	87.6	832	723	86.9	76.1	2454	1941	79.1	1941	1709	88.0	69.6	71.4
	24914	914	815	89.2	815	773	94.8	84.6	934	762	81.6	762	674	88.5	72.2	78.3
	24915	939	803	85.5	803	757	94.3	80.6	1069	841	78.7	841	775	92.2	72.5	76.3
	24916	1548	1354	87.5	1354	1268	93.6	81.9	1804	1481	82.1	1481	1321	89.2	73.2	77.2
ON	Total	25781	22457	87.1	22457	20869	92.9	80.9	28458	22823	80.2	22823	20897	91.6	73.4	77.0
	35926	453	359	79.2	359	334	93.0	73.7	662	526	79.5	526	497	94.5	75.1	74.5
	35927	492	425	86.4	425	390	91.8	79.3	551	447	81.1	447	408	91.3	74.0	76.5
	35930	978	831	85.0	831	759	91.3	77.6	1110	879	79.2	879	820	93.3	73.9	75.6
	35931	467	406	86.9	406	382	94.1	81.8	448	353	78.8	353	326	92.4	72.8	77.4
	35933	634	556	87.7	556	530	95.3	83.6	600	509	84.8	509	481	94.5	80.2	81.9
	35934	451	382	84.7	382	355	92.9	78.7	520	413	79.4	413	374	90.6	71.9	75.1
	35935	600	515	85.8	515	479	93.0	79.8	654	552	84.4	552	511	92.6	78.1	78.9
	35936	820	732	89.3	732	689	94.1	84.0	969	793	81.8	793	710	89.5	73.3	78.2
	35937	1145	961	83.9	961	875	91.1	76.4	1156	885	76.6	885	807	91.2	69.8	73.1
	35938	522	477	91.4	477	437	91.6	83.7	629	505	80.3	505	471	93.3	74.9	78.9
	35939*	738	655	88.8	655	623	95.1	84.4	760	637	83.8	637	583	91.5	76.7	80.5
	35940	435	418	96.1	418	404	96.7	92.9	515	423	82.1	423	390	92.2	75.7	83.6
	35941	645	543	84.2	543	506	93.2	78.4	636	528	83.0	528	495	93.8	77.8	78.1
	35942	469	430	91.7	430	392	91.2	83.6	630	494	78.4	494	448	90.7	71.1	76.4
	35943	634	562	88.6	562	518	92.2	81.7	615	525	85.4	525	479	91.2	77.9	79.8
	35944	922	750	81.3	750	692	92.3	75.1	978	792	81.0	792	726	91.7	74.2	74.6
	35945	334	289	86.5	289	274	94.8	82.0	467	393	84.2	393	367	93.4	78.6	80.0
	35946	957	861	90.0	861	822	95.5	85.9	1005	805	80.1	805	722	89.7	71.8	78.7
	35947*	701	649	92.6	649	607	93.5	86.6	745	644	86.4	644	595	92.4	79.9	83.1
	35949	381	324	85.0	324	301	92.9	79.0	420	346	82.4	346	323	93.4	76.9	77.9
	35951	1228	1017	82.8	1017	930	91.4	75.7	1357	1122	82.7	1122	1045	93.1	77.0	76.4
	35952	462	409	88.5	409	368	90.0	79.7	467	402	86.1	402	385	95.8	82.4	81.1
	35953	1356	1147	84.6	1147	1055	92.0	77.8	1596	1203	75.4	1203	1070	88.9	67.0	72.0
	35955	553	471	85.2	471	430	91.3	77.8	540	456	84.4	456	416	91.2	77.0	77.4
	35956	399	362	90.7	362	334	92.3	83.7	477	389	81.6	389	359	92.3	75.3	79.1
	35957	475	434	91.4	434	408	94.0	85.9	451	371	82.3	371	345	93.0	76.5	81.3
	35958	621	557	89.7	557	524	94.1	84.4	659	549	83.3	549	508	92.5	77.1	80.6
	35960	824	730	88.6	730	689	94.4	83.6	876	710	81.1	710	644	90.7	73.5	78.4

Table 9.1 Tableau 9.1		Area frame / Base aréolaire								Phone frames / Bases téléphoniques								All cases / Tous les cas
Prov./ Terr.	Health Region	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	Combined resp. rates		
Prov./ Terr.	Région socio- sanitaire	# mén. cibles	# mén. rép.	Taux de rép. mén.	# pers. sélect.	# rép.	Taux de rép. pers.	Taux de rép.	# mén. cibles	# mén. rép.	Taux de rép. mén.	# pers. sélect.	# rép.	Taux de rép. pers.	Taux de rép.	Taux de rép. combiné		
	35961	646	589	91.2	589	558	94.7	86.4	713	607	85.1	607	560	92.3	78.5	82.3		
	35962	567	516	91.0	516	482	93.4	85.0	631	540	85.6	540	494	91.5	78.3	81.5		
	35965	999	865	86.6	865	780	90.2	78.1	1015	841	82.9	841	781	92.9	76.9	77.5		
	35966	646	590	91.3	590	555	94.1	85.9	730	599	82.1	599	548	91.5	75.1	80.2		
	35968	870	779	89.5	779	742	95.3	85.3	1014	808	79.7	808	732	90.6	72.2	78.2		
	35970	1114	989	88.8	989	924	93.4	82.9	1354	1016	75.0	1016	900	88.6	66.5	73.9		
	35995	2243	1877	83.7	1877	1721	91.7	76.7	2508	1761	70.2	1761	1577	89.6	62.9	69.4		
MB	Total	4378	3973	90.7	3973	3785	95.3	86.5	4443	3758	84.6	3758	3567	94.9	80.3	83.3		
	46910	1231	1055	85.7	1055	993	94.1	80.7	1297	1087	83.8	1087	1037	95.4	80.0	80.3		
	46915*	694	640	92.2	640	617	96.4	88.9	755	639	84.6	639	607	95.0	80.4	84.5		
	46920*	762	710	93.2	710	667	93.9	87.5	606	521	86.0	521	479	91.9	79.0	83.8		
	46930	422	387	91.7	387	365	94.3	86.5	371	316	85.2	316	306	96.8	82.5	84.6		
	46940	456	425	93.2	425	407	95.8	89.3	453	392	86.5	392	376	95.9	83.0	86.1		
	46960*	813	756	93.0	756	736	97.4	90.5	961	803	83.6	803	762	94.9	79.3	84.4		
SK	Total	4431	4053	91.5	4053	3883	95.8	87.6	4804	4080	84.9	4080	3882	95.1	80.8	84.1		
	47901*	1074	999	93.0	999	965	96.6	89.9	989	828	83.7	828	783	94.6	79.2	84.7		
	47904	812	707	87.1	707	679	96.0	83.6	674	576	85.5	576	549	95.3	81.5	82.6		
	47905*	590	548	92.9	548	524	95.6	88.8	764	643	84.2	643	620	96.4	81.2	84.5		
	47906	860	778	90.5	778	738	94.9	85.8	705	591	83.8	591	567	95.9	80.4	83.4		
	47907*	689	643	93.3	643	608	94.6	88.2	651	573	88.0	573	543	94.8	83.4	85.9		
	47909*	406	378	93.1	378	369	97.6	90.9	1021	869	85.1	869	820	94.4	80.3	83.3		
AB	Total	7139	6388	89.5	6388	5970	93.5	83.6	7331	6145	83.8	6145	5830	94.9	79.5	81.5		
	48920	595	551	92.6	551	530	96.2	89.1	615	524	85.2	524	501	95.6	81.5	85.2		
	48921	507	470	92.7	470	450	95.7	88.8	446	375	84.1	375	363	96.8	81.4	85.3		
	48922	1658	1409	85.0	1409	1308	92.8	78.9	1691	1405	83.1	1405	1340	95.4	79.2	79.1		
	48923	786	713	90.7	713	654	91.7	83.2	812	682	84.0	682	649	95.2	79.9	81.5		
	48924	524	485	92.6	485	461	95.1	88.0	519	455	87.7	455	423	93.0	81.5	84.8		
	48925	1576	1395	88.5	1395	1291	92.5	81.9	1565	1312	83.8	1312	1252	95.4	80.0	81.0		
	48926	559	518	92.7	518	491	94.8	87.8	670	556	83.0	556	528	95.0	78.8	82.9		
	48927*	934	847	90.7	847	785	92.7	84.0	1013	836	82.5	836	774	92.6	76.4	80.1		
BC	Total	9687	8513	87.9	8513	7976	93.7	82.3	10247	7925	77.3	7925	7431	93.8	72.5	77.3		
	59911	348	332	95.4	332	321	96.7	92.2	324	278	85.8	278	258	92.8	79.6	86.2		

Table 9.1 Tableau 9.1		Area frame / Base aréolaire							Phone frames / Bases téléphoniques							All cases / Tous les cas
Prov./ Terr.	Health Region	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp. pers.	Pers. resp. rates	Resp. rates	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp. pers.	Pers. resp. rates	Resp. rates	Combined resp. rates
Prov./ Terr.	Région socio- sanitaire	# mén. cibles	# mén. rép.	Taux de rép. mén.	# pers. sélect.	# rép.	Taux de rép. pers.	Taux de rép.	# mén. cibles	# mén. rép.	Taux de rép. mén.	# pers. sélect.	# rép.	Taux de rép. pers.	Taux de rép.	Taux de rép. combiné
	59912	328	305	93.0	305	292	95.7	89.0	377	303	80.4	303	287	94.7	76.1	82.1
	59913	679	629	92.6	629	607	96.5	89.4	714	559	78.3	559	519	92.8	72.7	80.8
	59914	553	483	87.3	483	443	91.7	80.1	601	482	80.2	482	457	94.8	76.0	78.0
	59921	591	533	90.2	533	506	94.9	85.6	660	513	77.7	513	486	94.7	73.6	79.3
	59922	916	851	92.9	851	816	95.9	89.1	968	727	75.1	727	669	92.0	69.1	78.8
	59923	956	844	88.3	844	792	93.8	82.8	1069	802	75.0	802	741	92.4	69.3	75.7
	59931	500	425	85.0	425	403	94.8	80.6	575	400	69.6	400	379	94.8	65.9	72.7
	59932	1085	822	75.8	822	780	94.9	71.9	1134	797	70.3	797	735	92.2	64.8	68.3
	59933	583	518	88.9	518	478	92.3	82.0	723	571	79.0	571	543	95.1	75.1	78.2
	59941	853	759	89.0	759	691	91.0	81.0	881	694	78.8	694	660	95.1	74.9	77.9
	59942	678	605	89.2	605	572	94.5	84.4	662	538	81.3	538	513	95.4	77.5	81.0
	59943	277	245	88.4	245	235	95.9	84.8	355	290	81.7	290	279	96.2	78.6	81.3
	59951*	834	702	84.2	702	650	92.6	77.9	693	545	78.6	545	505	92.7	72.9	75.6
	590052	506	460	90.9	460	390	84.8	77.1	511	426	83.4	426	400	93.9	78.3	77.7
Terr.	Total	2940	2664	90.6	2664	2477	93.0	84.3	249	201	80.7	201	181	90.0	72.7	83.3

* = régions sociosanitaires regroupé

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.

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 non-réponse partielle (le fait de ne pas répondre à une ou plusieurs questions) à une non-réponse totale. Dans le cas du cycle 3.1 de l'ESCC, il n'y a presque pas eu 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. On a traité les cas de non-réponse totale 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.

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. La mesure résultante, 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 du cycle 3.1 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

Cette section du guide décrit les lignes directrices que doivent suivre les utilisateurs qui totalisent, analysent, publient ou diffusent de quelque autre façon des données provenant du fichier de microdonnées à grande diffusion (FMGD) de l'enquête. Ces lignes directrices devraient leur permettre de reproduire les chiffres déjà publiés par Statistique Canada et de produire aussi des chiffres non encore publiés conformes aux lignes directrices établies.

10.1 Lignes directrices pour l'arrondissement

Afin que les estimations calculées d'après ce fichier de microdonnées à grande diffusion en vue d'être publiées ou diffusées de toute autre façon 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 classique.
- 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.
- 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 FMGD 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 progiciels ne permettent pas d'obtenir des estimations qui coïncident exactement avec celles de Statistique Canada.

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 FMGD.

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? (SMKE_202)

- 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 est-ce que ... fume(z) chaque jour?
(SMKE_204)

||| Nombre de cigarettes

10.2.2 Totalisation d'estimations de type nominal

On peut obtenir, à partir du fichier de microdonnées à grande diffusion, des estimations du nombre de personnes qui possèdent une caractéristique donnée en additionnant les poids finals 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 :

- additionner les poids finals des enregistrements contenant la caractéristique voulue pour le numérateur (\hat{X});
- additionner les poids finals des enregistrements contenant la caractéristique voulue pour le dénominateur (\hat{Y});
- 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) :

- multiplier la valeur de la variable étudiée par le poids final, puis faire la somme de cette quantité pour tous les enregistrements visés pour obtenir le numérateur (\hat{X});
- faire la somme des poids finals 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 sommant le produit entre la valeur de la variable **SMKE_204** et le poids **WTSE_M**. Ensuite additionnez cette valeur pour les enregistrements pour lesquels la valeur de la variable **SMKE_202** 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 **SMKE_202** 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.

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. Par conséquent, si les estimations calculées au moyen de ces progiciels sont souvent exactes, les variances n'ont, quant à elles, pratiquement aucune signification.

Dans le cas de nombreuses méthodes d'analyse (par exemple la régression linéaire, la régression logistique, l'analyse de la variance), une méthode permet de corriger les résultats obtenus des progiciels courants de façon à ce qu'il soit plus adéquat. Cette méthode consiste à rééchelonner les poids qui figurent dans les enregistrements de façon à ce que le poids moyen soit égal à un (1). Les résultats produits par les progiciels classiques sont ainsi plus raisonnables puisque, même s'ils ne reflètent toujours pas la stratification et la mise en grappes du plan d'échantillonnage, ils tiennent compte de la sélection avec probabilités inégales. On peut effectuer cette transformation en utilisant dans l'analyse un poids égal au poids original divisé par la moyenne des poids originaux pour les unités échantillonnées (personnes) qui contribuent à l'estimation en question.

Pour permettre à l'utilisateur d'évaluer la qualité des totalisations estimées d'après les données, Statistique Canada a produit un ensemble de tableaux de variabilité d'échantillonnage approximative (couramment appelées « Tableaux des CV ») pour l'ESCC. On peut employer ces tableaux pour obtenir des coefficients de variation approximatifs pour les estimations de type nominal et les proportions. Pour plus de détails, consulter la section 11.

10.4 Lignes directrices pour la diffusion

Avant de diffuser et/ou de publier des estimations tirées des fichiers de microdonnées, l'utilisateur doit d'abord déterminer le nombre de répondants dans l'échantillon ayant la caractéristique à l'étude (par exemple, le nombre de répondants qui fument lorsqu'on s'intéresse à la proportion de fumeurs pour une population donnée). Si ce nombre est inférieur à 30, l'estimation pondérée ne doit pas être diffusée, quelle que soit la valeur de son coefficient de variation. Pour les estimations pondérées basées sur des échantillons d'au moins 30 personnes, l'utilisateur doit calculer le coefficient de variation de l'estimation arrondie et suivre les lignes directrices qui suivent.

Table 10.1 Lignes directrices relatives à la variabilité d'échantillonnage

Type d'estimation	CV (en %)	Lignes directrices
Acceptable	$0,0 \leq CV \leq 16,5$	On peut envisager une diffusion générale non restreinte des estimations. Aucune annotation particulière n'est nécessaire.
Marginale	$16,6 < CV \leq 33,3$	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	$CV > 33,3$	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 permettre aux utilisateurs d'avoir facilement accès à des coefficients de variation qui s'appliqueraient à une multitude d'estimations de type nominal obtenues à partir de ce fichier de microdonnées à grande diffusion, Statistique Canada a produit un ensemble de tableaux de la variabilité d'échantillonnage approximative. 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.

Les coefficients de variation (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 du plan*, on a d'abord calculé les effets du plan pour une vaste gamme de caractéristiques, puis pour chaque tableau, choisi une valeur conservatrice parmi tous les effets du 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 effets du plan, les tailles d'échantillon et les comptes de population qui ont servi à produire les tableaux de la variabilité d'échantillonnage approximative de même que les tableaux, sont disponibles à l'Annexe E. 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.

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.

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 estimative de personnes qui fument tous les jours parmi les fumeurs est plus fiable que le nombre estimatif 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 (p. ex., 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'écart-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.

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

Supposons qu'un utilisateur estime à 4 722 617 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 (4 722 617) 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 4 000 000.
- 3) Le coefficient de variation d'un agrégat estimé (exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 1,0 %.
- 4) Le coefficient de variation approximatif de l'estimation est donc 1,0 %. Par conséquent, selon les « Lignes directrices relatives à la variabilité d'échantillonnage » présentées à la section 10.4, l'estimation selon laquelle 4 722 617 personnes fument tous les jours peut être diffusée sans réserve.

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

Supposons qu'un utilisateur estime à $4\,722\,617/6\,081\,453 = 77,7\%$ 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 (77,7 %) et la partie numérateur du pourcentage (4 722 617) pour déterminer le coefficient de variation.
- 3) Le numérateur (4 722 617) 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 4 000 000. 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 0,6 %, est le coefficient de variation (exprimé en pourcentage) à employer.
- 5) Le coefficient de variation de l'estimation est donc 0,6 %. Par conséquent, selon les « Lignes directrices relatives à la variabilité d'échantillonnage » présentées à la section

10.4, l'estimation selon laquelle 77,7 % 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, $2\,535\,367/13\,078\,499 = 19,4\%$ fument tous les jours (estimation 1), alors que chez les femmes, ce pourcentage est estimé à $2\,187\,250/13\,476\,931 = 16,2\%$ (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 1,5 % le CV de l'estimation 1 (exprimé en pourcentage) et à 1,5 % 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 est l'estimation 1, \hat{X}_2 est 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 $\hat{d} = (0,194 - 0,162) = 0,032$ est donc :

$$\begin{aligned} \sigma_{\hat{d}} &= \sqrt{[(0,194)(0,015)]^2 + [(0,162)(0,015)]^2} \\ &= 0,004 \end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,004/0,032 = 0,125$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 12,5 % (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 4 722 617 le nombre de personnes qui fument tous les jours et à 1 358 836 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 1 358 836. 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 2,3 %.
- 4) Le dénominateur de cette estimation de rapport 4 722 617. Le nombre le plus petit qui se rapproche le plus de ce nombre est 4 000 000. Le coefficient de variation de cette estimation (exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 1,0 %.
- 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,

$$\begin{aligned} \alpha_{\hat{R}} &= \sqrt{(0,023)^2 + (0,01)^2} \\ &= 0,025 \end{aligned}$$

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 1 358 836/4 722 617, soit 0,29:1. Le coefficient de variation de cette estimation est 2,5 % (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 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 %

Nota : 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).

$$\begin{aligned} \hat{X} &= 0,777 \\ z &= 2 \end{aligned}$$

$\alpha_{\hat{x}}$ = 0,006 est le coefficient de variation de cette estimation selon les tableaux.

$$CI_{\hat{x}} = \{0,777 - (2) (0,777) (0,006), 0,777 + (2) (0,777) (0,006)\}$$

$$CI_{\hat{x}} = \{0,768, 0,786\}$$

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 $z = (\hat{X}_1 - \hat{X}_2) / \sigma_{\hat{d}}$ 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 0,05.

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 0,004. Par conséquent,

$$z = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0,194 - 0,162}{0,004} = \frac{0,032}{0,004} = 8$$

Puisque $z = 8$ est supérieur à 2, on doit conclure qu'il existe une différence significative entre les deux estimations au niveau de signification de 0,05. 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 de 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 précisions, telles que des méthodes par rééchantillonnage. Parmi celles-ci, la méthode du bootstrap est celle recommandée pour l'analyse des données de l'ESCC.

Le calcul de coefficients de variation (ou tout autre mesure de précision) fait à l'aide de la méthode du bootstrap nécessite toutefois l'accès à de l'information considérée confidentielle qui n'est évidemment pas disponible dans le fichier de microdonnées à grande diffusion. Le calcul doit donc se faire à l'aide du fichier maître. L'accès au fichier maître est discuté à la section 12.3.

Pour le calcul de coefficients de variation, il est conseillé d'utiliser la méthode du bootstrap. Un programme macro, appelé le « Bootvar », a été développé pour faciliter le calcul à l'aide de la méthode bootstrap. Le programme Bootvar est offert en formats SAS et SPSS, et est constitué de macros qui calculent les variances de totaux, ratios, différences entre ratios, et pour des régressions linéaires et logistiques.

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 (le « Bootvar »).

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.

Enfin, un utilisateur qui peut se servir des tableaux de CV, mais obtient ainsi un coefficient de variation compris dans la fourchette marginale (de 16,6 % à 33,3 %), devrait diffuser les estimations associées en y joignant une mise en garde aux utilisateurs quant à la variabilité d'échantillonnage élevée liée aux estimations. Dans ce cas, il serait bon de recalculer le coefficient de corrélation à l'aide du programme de variance exacte pour vérifier si ces estimations peuvent être diffusées sans mise en garde. 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, donc, jugée grossière, alors que le programme de calcul de la variance réelle produit le coefficient de variation précis associé à la variable en question.

11.8 Seuils pour la diffusion des estimations relatives à l'ESCC

L'Annexe E présente les tableaux indiquant les seuils de diffusion des totaux selon les estimations pour le Canada, les provinces, les régions sociosanitaires, les régions de CLSC ainsi que pour les différents groupes d'âges (pour le Canada seulement). Les estimations inférieures à la valeur indiquée dans la colonne « Marginal » ne peuvent en aucun cas être diffusées.

12. Utilisation du fichier

La présente section débute par une description de la variable de pondération présente sur le fichier de microdonnées à grande diffusion, et fournit des explications sur la façon de l'utiliser lorsqu'on effectue des totalisations. Suit une explication de la convention appliquée pour nommer les variables de l'ESCC. Ensuite, vient la description des diverses méthodes d'accès aux données que peuvent adopter les analystes.

12.1 Utilisation de la variable de pondération

La variable de pondération **WTSE_M** représente le poids d'échantillonnage pour le cycle 3.1 de l'ESCC. Pour un répondant donné, ce poids d'échantillonnage peut être interprété comme étant le nombre de personnes que le répondant représente dans la population. 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.

12.2 Convention appliquée pour nommer les variables

On a adopté, pour nommer les variables de l'ESCC, une convention qui permet aux utilisateurs des données de repérer et d'utiliser facilement celles-ci en fonction du module et du cycle. Les exigences qui suivent doivent être satisfaites : limiter les noms des variables à huit caractères au plus pour qu'il soit facile de les utiliser avec les logiciels d'analyse, préciser l'édition de l'enquête (cycle 1.1, 1.2...) dans le nom, et permettre de repérer facilement les variables conceptuellement identiques d'un cycle à l'autre de l'enquête. Les noms des variables correspondant à des modules ou à des questions identiques ne devraient différer qu'en ce qui concerne la position réservée dans le nom à l'identification du cycle particulier durant lequel les données ont été recueillies.

12.2.1 Structure élémentaire des noms des variables de l'ESCC

Chacun des huit caractères du nom d'une variable fournit des renseignements sur le type de données que contient la variable.

Positions 1 à 3 :	Nom du module/de la section du questionnaire
Position 4 :	Cycle de l'enquête
Position 5 :	Type de variable
Positions 6 à 8 :	Numéro de la question

Par exemple, la structure du nom de la variable correspondant à la question 202, module Usage du tabac, cycle 3.1, c'est-à-dire SMKE_202 est la suivante :

Positions 1 à 3 : SMK Module de la dépression
Position 4 : E Cycle 3.1
Position 5 : _ (_ = données recueillies)
Position 6 à 8 : 202 numéro de la question et option de réponse

12.2.2 Positions 1 à 3 : Nom de la variable/section du questionnaire

On se sert des valeurs suivantes pour la composante du nom de la variable correspondant à la section du questionnaire :

ACC	Accès aux services de soins de santé	MAM	Mammographie
ADM	Administration	MED	Médicaments
ALC	Consommation d'alcool	MEX	Expériences maternelles
BPC	Tension artérielle	NDE	Dépendance à la nicotine
BRX	Examen des seins	OH2	Santé bucco-dentaire 2
BSX	Auto-examen des seins	ORG	Organismes à but non lucratif
CCC	Problèmes de santé chroniques	PAC	Activités physiques
CCS	Dépistage du cancer du côlon et du rectum	PAP	Test PAP
CIH	Changements pour améliorer la santé	PAS	Satisfaction des patients
CMH	Utilisation des services – santé mentale	PCU	Examen général
CPG	Jeu excessif	PSA	Dépistage du cancer de la prostate
CST	Facteurs de stress durant l'enfance et la vie adulte	RAC	Limitation des activités
DEN	Visites chez le dentiste	REP	Blessures (mouvement répétitif)
DHH	Données démographiques et composition du ménage	SAC	Activités sédentaires
DIA	Les soins pour le diabète	SAM	Identificateurs d'échantillon
DIS	Détresse	SCA	Outils pour cesser de fumer
DPS	Dépression	SCH	Usage du tabac – les étapes de changement
EDU	Education	SDC	Renseignements socio-démographiques
ETS	Exposition à la fumée des autres	SFE	Estime de soi
EYX	Examens de la vue	SFR	État de santé – ST-36
FDC	Choix alimentaires	SMK	Usage du tabac
FLU	Vaccins contre la grippe	SLP	Sommeil
FSC	Sécurité alimentaire	SPC	Usage du tabac – consultation d'un médecin
FVC	Consommation de fruits et de légumes	SSA	Soutien social - disponibilité

GEN	État de santé général	SSB	Protection contre le soleil
GEO	Identificateurs géographiques	SSU	Soutien social – utilisation
HCS	Satisfaction du système des soins de santé	STC	Faire face au stress
HCU	Utilisation des soins de santé	STS	Des sources du stress
HMC	Soins de santé à domicile	SUI	Pensées suicidaires et tentative de suicide
HUI	L'indice de l'état de santé	SWL	Satisfaction de la vie
HWT	Taille et poids	SXB	Comportement sexuel
IDG	Drogues illicites	TWD	Incapacité des deux dernières semaines
INC	Revenu	UPE	Utilisation de l'équipement protecteur
INJ	Blessures	WST	Stress au travail
INS	Couverture d'assurance	WTM	Temps d'attente
LBF	Population active	WTS	Poids de sondage
		YSM	Usage du tabac chez les jeunes

12.2.3 Position 4 : Cycle

Cycle	Description
-------	-------------

- | | |
|----------|---|
| A | <p><u>Cycle 1.1 : Enquête sur la santé dans les collectivités canadiennes</u></p> <ul style="list-style-type: none"> • enquête à l'échelle régionale, échantillon stratifié selon la région sociosanitaire • contenu commun et contenu optionnel sélectionnés par les régions sociosanitaires • estimations à l'échelle régionale (régions sociosanitaires), (provincial, territorial et national) |
| B | <p><u>Cycle 1.2 : Enquête sur la santé dans les collectivités canadiennes, santé mentale et bien-être</u></p> <ul style="list-style-type: none"> • enquête à l'échelle provinciale • contenu thématique et contenu général supplémentaire • estimations aux échelle provinciale, territoriale et nationale |
| C | <p><u>Cycle 2.1 : Enquête sur la santé dans les collectivités canadiennes</u></p> <ul style="list-style-type: none"> • enquête à l'échelle régionale, échantillon stratifié selon la région sociosanitaire • contenu commun et contenu optionnel sélectionnés par les régions sociosanitaires • estimations à l'échelle régionale (régions sociosanitaires), (provincial, territorial et national) |

- D** Cycle 2.2 : Enquête sur la santé dans les collectivités canadiennes, nutrition
- enquête à l'échelle provinciale
 - contenu thématique et contenu général supplémentaire
 - estimations aux échelle provinciale, territoriale et nationale
- E** Cycle 3.1 : Enquête sur la santé dans les collectivités canadiennes
- enquête à l'échelle régionale, échantillon stratifié selon la région sociosanitaire
 - contenu commun et contenu optionnel sélectionnés par les régions sociosanitaires
 - estimations à l'échelle régionale (régions sociosanitaires), (provincial, territorial et national)

12.2.4 Position 5 : Type de variable

–	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)

12.2.5 Positions 6 à 8 : Nom de la variable

En général, les trois dernières positions 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 chiffres. Par exemple, la question Q01A du questionnaire devient simplement 01A et la question Q15, simplement 15.

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 Q2 a 4 réponses possibles, les nouvelles questions seraient Q2A pour la première possibilité, Q2B pour la deuxième, Q2C pour la troisième et ainsi de suite. Si seulement les options 2 et 3 sont choisies, alors Q2A = Non, Q2B = Oui, Q2C = Oui et Q2D = Non.

12.3 Accès au fichier maître

Afin de respecter le droit à la vie privée des répondants qui participent à l'enquête, 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 destiné à confirmer qu'aucune personne ne pourra être identifiée. Les valeurs rares pour certaines variables susceptibles de permettre l'identification d'une personne sont supprimées du fichier ou agrégées en catégories moins détaillées, de façon à réduire au minimum le risque de divulgation de renseignements personnels. Fréquemment, ces variables sont les plus essentielles à l'analyse complète des données d'enquête. Puisqu'une quantité importante de ressources est investie dans la collecte de ces données, il est important de prendre des mesures pour tirer le plein potentiel analytique des fichiers de microdonnées à grande diffusion afin de bien rentabiliser l'investissement statistique.

Une première méthode offerte à tous les utilisateurs consiste à demander au personnel des Services personnalisés à la clientèle de la Division de la statistique de la santé de produire des totalisations personnalisées. Ce service 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. Ce service est offert contre remboursement des frais.

Deuxièmement, le Programme des centres de données de recherche permet aux chercheurs de soumettre à Statistique Canada un projet de recherche fondé sur les données des fichiers maîtres. Un ensemble particulier de règles est appliqué afin de décider quels projets seront acceptés. Lorsque le projet est accepté, le chercheur est considéré comme étant « employé réputé » par Statistique Canada pour la durée de l'étude et se voit accorder l'accès au fichier maître de l'enquête dans des locaux désignés de Statistique Canada. Pour plus de renseignements, consultez la page web suivante : http://www.statcan.ca/francais/rdc/index_f.htm

En dernier lieu, le service de télé-accès aux fichiers maîtres de l'enquête est un moyen d'accéder à ces données si il est impossible de passer par un Centre. On peut fournir à l'acheteur d'un produit de microdonnées un fichier maître "fictif" d'essai et le cliché d'enregistrement correspondant. Grâce à ces outils, il peut mettre au point son propre ensemble de programmes analytiques en se servant du fichier fictif pour confirmer que les routines fonctionnent convenablement. Il ne lui reste plus qu'à envoyer le code pour les totalisations personnalisées par courrier électronique à cchs-escc@statcan.ca. Le code est transmis au réseau interne protégé de Statistique Canada et traité en regard du fichier maître approprié de données du cycle 3.1 de l'ESCC. Les estimations générées seront communiquées à l'utilisateur, sujet aux directives sur l'analyse et la communication des

données tel qu'exposé dans les grandes lignes à la section 10 de ce document. Les résultats sont filtrés pour vérifier s'ils sont conformes aux normes de confidentialité et de fiabilité, puis, les données de sortie sont renvoyées au client. Ce service est gratuit.

ANNEXE A

**ENQUÊTE SUR LA SANTÉ DANS LES COLLECTIVITÉS CANADIENNES (ESCC)
CYCLE 3.1 (2005)**

FICHER MAÎTRE

**LIGNES DIRECTRICES POUR L'UTILISATION DES VARIABLES DES
SOUS-ÉCHANTILLONS**

Statistique Canada

Juin 2006

Sous-échantillons du cycle 3.1 de l'ESCC

Le questionnaire du cycle 3.1 de l'ESCC comprend trois types de modules :

1. **Contenu commun** - modules du questionnaire posés à tous les répondants dans toutes les régions sociosanitaires;
2. **Contenu optionnel** - modules inclus dans le questionnaire à la demande des régions sociosanitaires. Tout module inclus dans le questionnaire à titre optionnel pour une région sociosanitaire particulière a été posé à tous les répondants de la région. Bien que l'objet du contenu optionnel soit de permettre le calcul d'estimations au niveau de la région sociosanitaire, toutes les régions de chacune des dix provinces ont choisi le même contenu optionnel du cycle 3.1, ce qui offre la possibilité de calculer des estimations provinciales;
3. **Contenu des sous-échantillons** - modules du questionnaire posés uniquement à un sous-ensemble de répondants. L'objet du sous-échantillonnage est de permettre le calcul d'estimations provinciales et nationales en réduisant au minimum le fardeau de réponse. Trois ensembles de modules ont été posés à trois sous-échantillons (distincts mais chevauchants) :

Sous-échantillon 1 : Indice de l'état de santé (IES), Consommation de fruits et de légumes, Population active (questionnaire détaillé);

Sous-échantillon 2 : Mesures de la taille et du poids;

Sous-échantillon 3 : Accès aux services de santé, Temps d'attente, Satisfaction des patients.

Nota : Le sous-échantillon 3 remplace l'Enquête sur l'accès aux services de santé (EASS). En 2000-2001, l'EASS a constitué un suivi de l'ESCC. En 2003, elle y a été intégrée en partie.

Convention de dénomination des variables des modules de sous-échantillon

Les variables des modules de contenu commun et de contenu optionnel sont désignées par la lettre « E » à la position 4 du nom de la variable, p. ex. SMKE_201.

Les variables des modules de contenu des sous-échantillons sont désignées par la lettre « Z » à la position 4 du nom de la variable, p. ex. ACCZ_01.

Exception faite des Mesures de la taille et du poids, chacun des modules de sous-échantillon a aussi été mis à la disposition des régions sociosanitaires à titre de contenu optionnel. Pour chacun des modules susmentionnés, il existe deux ensembles de variables : HUIE_28 et HUIZ_28. Les variables s'appliquent ou non à un répondant donné, selon qu'il fait partie ou non du sous-échantillon pertinent **et** qu'il vit ou non dans une région sociosanitaire où le module a été choisi comme contenu optionnel. Le tableau 1 illustre les possibilités.

Tableau 1. Valeurs potentiellement non applicables à une variable type de sous-échantillon

Répondant du sous-échantillon 1	Répondant vit dans une région sociosanitaire où Indice de l'état de santé (IES) faisait partie du contenu optionnel	
	OUI	NON
OUI	HUIE_28 et HUIZ_28 s'appliquent	HUIE_28 ne s'applique pas et HUIZ_28 s'applique
NON	HUIE_28 s'applique et HUIZ_28 ne s'applique pas	HUIE_28 et HUIZ_28 ne s'appliquent pas

Contenu des sous-échantillons figurant dans des fichiers distincts

Afin d'encourager l'utilisation appropriée des données de l'ESCC, le contenu des sous-échantillons est fourni dans des fichiers distincts. Chacun comporte un poids d'échantillonnage et un ensemble de poids bootstrap correspondants, qui doivent être utilisés pour produire des estimations valables pour toutes les variables du fichier. Le tableau 2 décrit l'ensemble des quatre fichiers de données diffusés au cours du cycle 3.1.

Tableau 2. Noms et contenu des fichiers de données de l'ESCC 3.1

Fichier	Nom du fichier	Poids d'échantillonnage	Fichier des poids bootstrap	Variables incluses
Fichier maître principal	HS.txt	WTSE_M	b5.txt	Tous les modules communs et optionnels. Dans le cas des modules qui font partie du contenu optionnel et du contenu des sous-échantillons, l'ensemble de variables « E » est inclus.
Fichier maître du sous-échantillon 1	HS_S1.txt	WTSE_S1M	b5_s1.txt	Tous les modules communs, plus l'ensemble de variables « Z » pour Indice de l'état de santé (IES), Consommation de fruits et de légumes, Population active (questionnaire détaillé)
Fichier maître du sous-échantillon 2	HS_S2.txt	WTSE_S2M	b5_s2.txt	Tous les modules communs, plus l'ensemble de variables « Z » pour le module des mesures de la taille et du poids.
Fichier maître du sous-	HS_S3.txt	WTSE_S3M	b5_s3.txt	Tous les modules communs, plus l'ensemble de variables « Z » pour le module Accès aux services de

échantillon 3				santé, Temps d'attente, Satisfaction des patients.
------------------	--	--	--	---

Analyse combinant le contenu des sous-échantillons et le contenu optionnel

Les modules des sous-échantillons de l'ESCC visent à fonder le calcul d'estimations au niveau national. Des estimations à l'échelon provincial ou territorial sont également offertes pour tous les modules de sous-échantillon, sauf Mesures de la taille et du poids. Les sous-échantillons, et les poids connexes, ne sont pas destinés au calcul d'estimations au niveau de la région sociosanitaire.

Au cours du cycle 3.1, certains modules de sous-échantillon ont également été choisis par l'ensemble des régions sociosanitaires d'au moins une province à titre de contenu optionnel, comme le montre le tableau 3.

Tableau 3. Modules des sous-échantillons qui ont aussi été sélectionnés comme contenu optionnel

Sous-échantillon	Provinces où les modules des sous-échantillons ont été choisis par toutes les régions sociosanitaires
1	<ul style="list-style-type: none"> • IES (Colombie-Britannique) • Consommation de fruits et de légumes (ÎP-E, Ontario, Alberta, C-B.)
3	<ul style="list-style-type: none"> • Accès aux services de santé, Temps d'attente (Nouveau Brunswick)

Si toutes les régions sociosanitaires d'une province ont choisi un module du **sous-échantillon 1** ou du **sous-échantillon 3** comme contenu optionnel, il nous est possible de calculer des estimations au niveau de la région sociosanitaire et de la province pour les variables du module en utilisant le **fichier maître principal** et le **poids d'échantillonnage maître**. Cela a pour avantage de fonder les calculs sur un échantillon important et donc de réduire l'erreur d'échantillonnage. (Les tableaux 4 à 7 résument la taille des échantillons du fichier maître principal et de chacun des fichiers des sous-échantillons.) L'estimation provinciale peut par la suite être comparée à l'estimation concernant d'autres provinces calculée au moyen du fichier de sous-échantillon pertinent et du poids d'échantillonnage de sous-échantillon.

Nota : Il n'est pas possible de calculer des estimations valables au moyen de totalisations croisées ou de modèles multivariés en utilisant un ensemble de variables tirées de différents sous-échantillons. La taille des sous-échantillons est distincte et leur poids leur est donc propre.

Tableau 4. Taille de l'échantillon par province/territoire pour le fichier maître

Province / Territoire	Taille de l'échantillon du fichier maître principal
CANADA	132 947
Terre-Neuve-et-Labrador	4 111
Île-du-Prince-Édouard	2 031
Nouvelle-Écosse	5 066
Nouveau-Brunswick	5,100
Québec	29 891
Ontario	41 766
Manitoba	7 352
Saskatchewan	7 765
Alberta	11 800
Colombie-Britannique	15 407
Yukon	868
Territoires du Nord-Ouest	1 007
Nunavut	783

Tableau 5. Taille du sous-échantillon 1 par province/territoire
(Indice de l'état de santé (IES), Consommation de fruits et de légumes, Population active (questionnaire détaillé))

Province / Territoire	Taille du sous-échantillon 1
CANADA	32 153
Terre-Neuve-et-Labrador	1 571
Île-du-Prince-Édouard	1 008
Nouvelle-Écosse	1 960
Nouveau-Brunswick	1 736
Québec	5 446
Ontario	6 622
Manitoba	2 104
Saskatchewan	2 075
Alberta	3 225
Colombie-Britannique	3 748
Yukon	868
Territoires du Nord-Ouest	1 007
Nunavut	783

1. En tout, 55 répondants qui ont refusé à la première question de chaque module (HUIZ_01, FVCZ_1A et LFSZ_01) ont été exclus du sous-échantillon 1.

Tableau 6. Taille du sous-échantillon 2 par province/territoire
(Mesures de la taille et du poids)

Province / Territoire	Taille du sous-échantillon 2 ¹
CANADA	4 735
Terre-Neuve-et-Labrador	85
Île-du-Prince-Édouard	17
Nouvelle-Écosse	127
Nouveau-Brunswick	93
Québec	1 185
Ontario	1 785
Manitoba	186
Saskatchewan	172
Alberta	458
Colombie-Britannique	627
Yukon	0
Territoires du Nord-Ouest	0
Nunavut	0

1. Inclut les femmes enceintes

Tableau 7. Taille du sous-échantillon 3 par province /territoire
(Accès aux services de santé, Temps d'attente, Satisfaction des patients)

Province / Territoire	Taille du sous-échantillon 3 ¹
CANADA	35 968
Terre-Neuve-et-Labrador	2 798
Île-du-Prince-Édouard	1 882
Nouvelle-Écosse	3 035
Nouveau-Brunswick	3 039
Québec	4 558
Ontario	4 692
Manitoba	3 076
Saskatchewan	3 199
Alberta	3 458
Colombie-Britannique	3 802
Yukon	804
Territoires du Nord-Ouest	925
Nunavut	700

1. Le sous-échantillon 3 comprend uniquement les personnes âgées de 15 ans et plus.

Autres considérations

Dictionnaires de données

Des dictionnaires de données distincts, y compris des descriptions de l'univers et des fréquences, sont fournis relativement au fichier maître principal et à chacun des trois fichiers de sous-échantillon.

Le traitement des modules de contenu optionnel au dictionnaire de données du fichier maître principal est identique à ce qu'il était lors des cycles 1.1 et 2.1 de l'ESCC. Pour chaque module, un indicateur signale si un répondant donné vit dans une région sociosanitaire où le module a été sélectionné comme contenu optionnel. Lorsque l'indicateur est 2 (non), toutes les variables du module ont des valeurs « sans objet ». Par exemple, la variable PASEFOPT indique si le module Satisfaction des patients s'applique à un répondant donné.

Totaux de population

Les poids d'échantillonnage des sous-échantillons 1 et 2 sont calibrés afin de représenter les Canadiens âgés de 12 ans et plus (27 131 965). Les poids d'échantillonnage du sous-échantillon 3 (dont les modules s'appliquent aux répondants âgés de 15 ans et plus) sont calibrés de façon à représenter les Canadiens âgés de 15 ans et plus (25 877 399).

Différences dans le calcul des variables de contenu commun fondé sur différents fichiers

Les variables tirées des modules de contenu commun peuvent être estimées à partir de l'un ou l'autre des quatre fichiers de données fournis. 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.

Questions

Pour toutes les questions concernant l'utilisation appropriée des données des sous-échantillons de l'ESCC, veuillez communiquer avec l'Unité d'accès aux données au :

1 (613) 951-1653
cchs-escc@statcan.ca

ANNEXE B : Régions sociosanitaires

Régions sociosanitaires – Diffusion (avant FMGD)

Terre-Neuve-et-Labrador

- 1011 Eastern Health Authority
- 1012 Health and Community Services Central Region
- 1013 Health and Community Services Western Region
- 1014 Labrador-Grenfell Health Authority

Île-du-Prince-Édouard

- 1101 West Prince Health Region
- 1102 East Prince Health Region
- 1104 Kings Health Region
- 1103 Queens Health Region

Nouvelle-Écosse

- 1201 Zone 1
- 1202 Zone 2
- 1203 Zone 3
- 1204 Zone 4
- 1205 Zone 5
- 1206 Zone 6

Nouveau-Brunswick

- 1301 Région 1
- 1302 Région 2
- 1303 Région 3
- 1304 Région 4
- 1305 Région n 5
- 1306 Région 6
- 1307 Région 7

Québec

- 2401 Région du Bas-Saint-Laurent
- 2402 Région du Saguenay - Lac-Saint-Jean
- 2403 Région de Québec
- 2404 Région de la Mauricie et du Centre-du-Québec
- 2405 Région de l'Estrie
- 2406 Région de Montréal-Centre
- 2407 Région de l'Outaouais
- 2408 Région de l'Abitibi-Témiscamingue
- 2409 Région de la Côte-Nord
- 2410 Région du Nord-du-Québec
- 2411 Région de la Gaspésie - Îles-de-la-Madeleine
- 2412 Région de la Chaudière-Appalaches
- 2413 Région de Laval
- 2414 Région de Lanaudière
- 2415 Région des Laurentides
- 2416 Région de la Montérégie

Régions sociosanitaires dans le FMGD

Terre-Neuve-et-Labrador

- 1011 Eastern Health Authority
- 1012 Health and Community Services Central Region
- 1013 Western Region/Labrador-Grenfell

Île-du-Prince-Édouard

- 1101 West Prince/East Prince/Kings

- 1103 Queens Health Region

Nouvelle-Écosse

- 1201 Zone 1
- 1202 Zone 2
- 1203 Zone 3
- 1204 Zone 4
- 1205 Zone 5
- 1206 Zone 6

Nouveau-Brunswick

- 1301 Région 1
- 1302 Région 2
- 1303 Région 3
- 1304 Région 4/5
- 1306 Région 6/7

Québec

- 2401 Région du Bas-Saint-Laurent
- 2402 Région du Saguenay - Lac-Saint-Jean
- 2403 Région de Québec
- 2404 Région de la Mauricie et du Centre-du-Québec
- 2405 Région de l'Estrie
- 2406 Région de Montréal-Centre
- 2407 Région de l'Outaouais
- 2408 Région de l'Abitibi-Témiscamingue
- 2409 Région de la Côte-Nord
- 2411 Région de la Gaspésie - Îles-de-la-Madeleine
- 2412 Région de la Chaudière-Appalaches
- 2413 Région de Laval
- 2414 Région de Lanaudière
- 2415 Région des Laurentides
- 2416 Région de la Montérégie

Régions sociosanitaires – Diffusion (avant FMGD)

Ontario

3526 Algoma
 3527 Brant
 3530 Durham
 3531 Elgin-St. Thomas
 3533 Grey Bruce
 3534 Haldimand-Norfolk
 3535 Haliburton, Kawartha, Pine Ridge
 3536 Halton
 3537 Hamilton
 3538 Hastings and Prince Edward
 3539 Huron
 3554 Perth
 3540 Chatham-Kent
 3541 Kingston, Frontenac and Lennox and Addington
 3542 Lambton
 3543 Leeds, Grenville and Lanark
 3544 Middlesex-London
 3546 Niagara
 3547 North Bay Parry Sound
 3563 Timiskaming
 3549 Northwestern
 3551 Ottawa
 3552 Oxford
 3553 Peel
 3555 Peterborough
 3556 Porcupine
 3557 Renfrew
 3558 Eastern Ontario
 3560 Simcoe Muskoka
 3561 Sudbury
 3562 Thunder Bay
 3565 Waterloo
 3566 Wellington-Dufferin-Guelph
 3568 Windsor-Essex
 3570 York
 3595 City of Toronto

Manitoba

4610 Winnipeg RHA
 4615 Brandon RHA
 4645 Assiniboine RHA
 4620 North Eastman RHA
 4625 South Eastman RHA
 4630 Interlake RHA
 4640 Central RHA

Régions sociosanitaires dans le FMGD

Ontario

3526 Algoma
 3527 Brant
 3530 Durham
 3531 Elgin-St. Thomas
 3533 Grey Bruce
 3534 Haldimand-Norfolk
 3535 Haliburton, Kawartha, Pine Ridge
 3536 Halton
 3537 Hamilton
 3538 Hastings and Prince Edward
 3539 Huron/Perth

 3540 Chatham-Kent
 3541 Kingston, Frontenac and Lennox and Addington
 3542 Lambton
 3543 Leeds, Grenville and Lanark
 3544 Middlesex-London
 3546 Niagara
 3547 North Bay Parry Sound/Timiskaming

 3549 Northwestern
 3551 Ottawa
 3552 Oxford
 3553 Peel
 3555 Peterborough
 3556 Porcupine
 3557 Renfrew
 3558 Eastern Ontario
 3560 Simcoe Muskoka
 3561 Sudbury
 3562 Thunder Bay
 3565 Waterloo
 3566 Wellington-Dufferin-Guelph
 3568 Windsor-Essex
 3570 York
 3595 City of Toronto

Manitoba

4610 Winnipeg RHA
 4615 Brandon/Assiniboine

 4620 North Eastman/South Eastman

 4630 Interlake RHA
 4640 Central RHA

Régions sociosanitaires – Diffusion (avant FMGD)

4660 Parkland RHA
 4670 Norman RHA
 4685 Burntwood RHA/Churchill RHA

Saskatchewan

4701 Sun Country RHA
 4702 Five Hills RHA
 4703 Cypress RHA
 4704 Regina Qu'Appelle RHA
 4705 Sunrise RHA
 4708 Kelsey trail RHA
 4706 Saskatoon RHA
 4707 Heartland RHA
 4710 Prairie North RHA

 4709 Prince Albert Parkland RHA
 4714 Mamawetan Churchill River RHA/Keewatin Yatthé RHA/Athabaska HA

Alberta

4820 Chinook Regional Health Authority
 4821 Palliser Health Region
 4822 Calgary Health Region
 4823 David Thompson Regional Health Authority
 4824 East Central Health
 4825 Capital Health
 4826 Aspen Regional Health Authority
 4827 Peace Country Health
 4828 Northern Lights Health Region

Colombie-Britannique

5911 East Kootenay
 5912 Kootenay-Boundary
 5913 Okanagan
 5914 Thompson/Cariboo
 5921 Fraser East
 5922 Fraser North
 5923 Fraser South
 5931 Richmond
 5932 Vancouver
 5933 North Shore/Coast Garibaldi
 5941 South Vancouver Island
 5942 Central Vancouver Island
 5943 North Vancouver Island
 5951 Northwest
 5953 Northeast
 5952 Northern Interior

Régions sociosanitaires dans le FMGD

4660 Parkland/Norman/Burntwood-Churchill

Saskatchewan

4701 Sun Country/Five Hills/Cypress

 4704 Regina Qu'Appelle RHA
 4705 Sunrise/Kelsey Trail

 4706 Saskatoon RHA
 4707 Heartland/Prairie North

 Prince Albert Parkland/Mamawetan-Keewatin
 4709 Yatthé-Athabaska

Alberta

4820 Chinook Regional Health Authority
 4821 Palliser Health Region
 4822 Calgary Health Region
 4823 David Thompson Regional Health Authority
 4824 East Central Health
 4825 Capital Health
 4826 Aspen Regional Health Authority
 4827 Peace Country/Northern Lights

Colombie-Britannique

5911 East Kootenay
 5912 Kootenay-Boundary
 5913 Okanagan
 5914 Thompson/Cariboo
 5921 Fraser East
 5922 Fraser North
 5923 Fraser South
 5931 Richmond
 5932 Vancouver
 5933 North Shore/Coast Garibaldi
 5941 South Vancouver Island
 5942 Central Vancouver Island
 5943 North Vancouver Island
 5951 Northwest/Northeast

 5952 Northern Interior

Régions sociosanitaires – Diffusion (avant FMGD)

Territoires

6001 Territoire du Yukon
6101 Territoires du Nord-Ouest
6201 Nunavut

Régions sociosanitaires dans le FMGD

Territoires

Territoire du Yukon / Territoires du
6001 Nord-Ouest/Nunavut

Régions sociosanitaires des échantillons supplémentaires du Québec (avant FMGD)

Échantillons supplémentaires du Québec

2401	Région du Bas-Saint-Laurent
1	La Matapédia
2	Matane
3	La Mitis
4	Rimouski-Neigette
5	Les Basques
6	Rivière-du-Loup
7	Témiscouata
8	Kamouraska
2406	Région de Montréal-Centre
1	Pierrefonds et Lac Saint-Louis
2	LaSalle et Vieux Lachine
3	Verdun/Côte Saint-Paul, Saint-Henri, Pointe Saint-Charles
4	René-Cassin et NDG/Montréal-Ouest
5	Côte-des-Neiges, Métro et Parc Extension
6	Nord de l'Île et Saint-Laurent
7	Ahuntsic et Montréal-Nord
8	La Petite Patrie et Villeray
9	Faubourgs, Plateau Mont-Royal et Saint-Louis du Parc
10	Saint-Michel et Saint-Léonard
11	Hochelaga-Maisonneuve, Olivier-Guimond et Rosemont
12	Rivière-des-Prairies, Mercier-Est/Anjou et Pointe-aux-Trembles/Montréal-Est
2413	Région de Laval
1	Est
2	Ouest

Régions sociosanitaires des échantillons du Québec dans le FMGD

2401	Région du Bas-Saint-Laurent
1	Est
5	Ouest
2406	Région de Montréal-Centre
1	Pierrefonds et Lac Saint-Louis
2	LaSalle et Vieux Lachine
3	Verdun/Côte Saint-Paul, Saint-Henri, Pointe Saint-Charles
4	René-Cassin et NDG/Montréal-Ouest
5	Côte-des-Neiges, Métro et Parc Extension
6	Nord de l'Île et Saint-Laurent
7	Ahuntsic et Montréal-Nord
8	La Petite Patrie et Villeray
9	Faubourgs, Plateau Mont-Royal et Saint-Louis du Parc
10	Saint-Michel et Saint-Léonard
11	Hochelaga-Maisonneuve, Olivier-Guimond et Rosemont
12	Rivière-des-Prairies, Mercier-Est/Anjou et Pointe-aux-Trembles/Montréal-Est
2413	Région de Laval
1	Est
2	Ouest