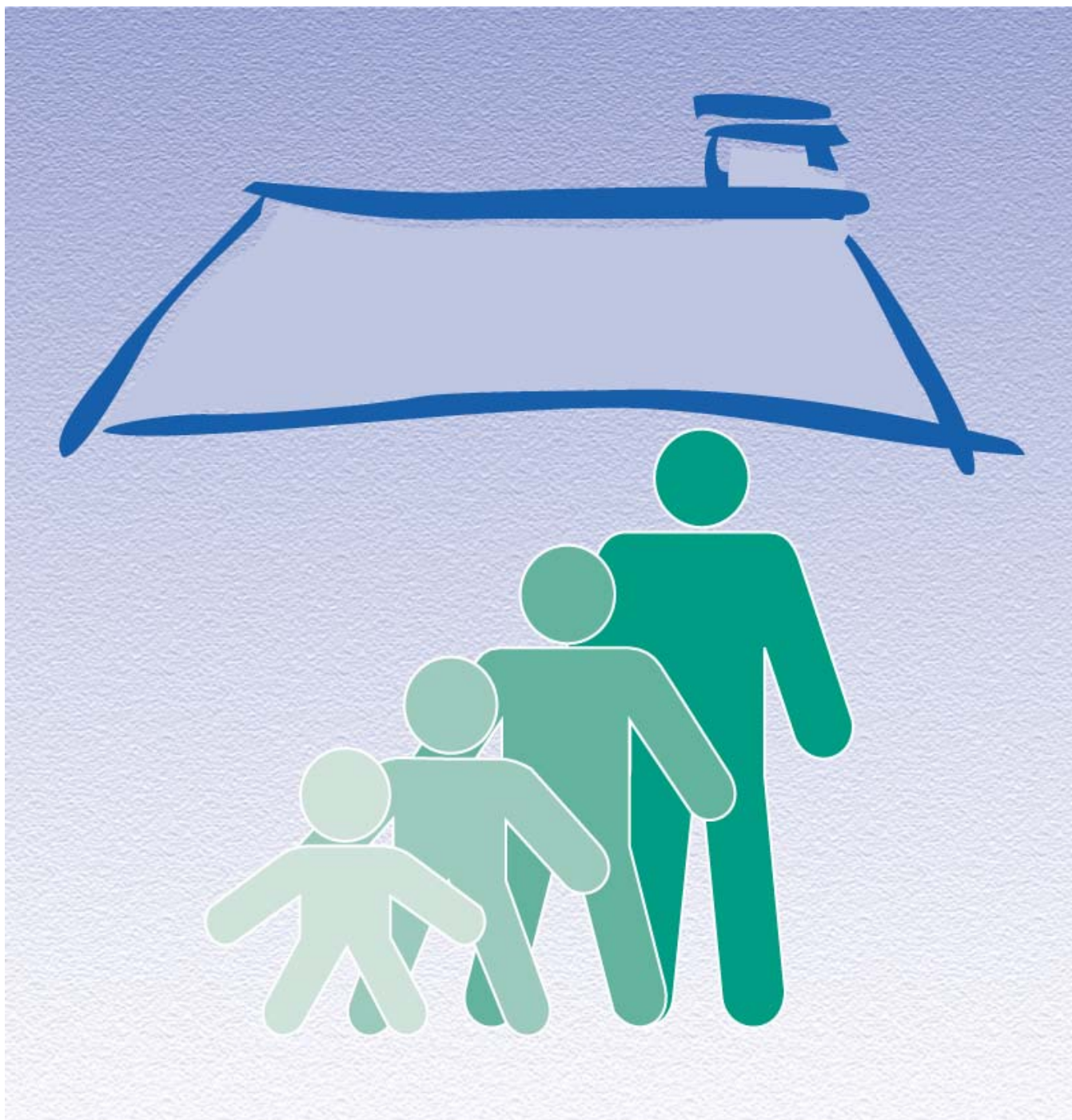


Enquête sur la santé dans les collectivités canadiennes (ESCC) – Composante annuelle

Guide de l'utilisateur Fichiers de microdonnées 2007-2008

Juin 2009



Statistics
Canada

Statistique
Canada

Canada

Table des matières

1.	INTRODUCTION	3
2.	HISTORIQUE	5
3.	REMANIEMENT DE L'ESCC EN 2007	6
4.	CONTENU DE L'ESCC	8
4.1	CONTENU COMMUN.....	8
4.2	CONTENU OPTIONNEL	9
4.3	CONTENU DE RÉPONSE RAPIDE.....	9
4.4	CONTENU DES FICHIERS DE MICRODONNÉES	9
5.	PLAN D'ÉCHANTILLONNAGE	11
5.1	POPULATION CIBLE	11
5.2	RÉGIONS SOCIO-SANITAIRES.....	11
5.3	TAILLE ET RÉPARTITION DE L'ÉCHANTILLON	11
5.4	BASES DE SONDAGE ET STRATÉGIES D'ÉCHANTILLONNAGE DES MÉNAGES	13
5.4.1	ÉCHANTILLONNAGE DES MÉNAGES À PARTIR DE LA BASE ARÉOLAIRE	13
5.4.2	ÉCHANTILLONNAGE DES MÉNAGES À PARTIR DE LA BASE LISTE DE NUMÉROS DE TÉLÉPHONE	15
5.4.3	ÉCHANTILLONNAGE DES MÉNAGES À PARTIR DE LA BASE DE SONDAGE À CA DE NUMÉROS DE TÉLÉPHONE	16
5.5	RÉPARTITION DE L'ÉCHANTILLON PAR PÉRIODE DE COLLECTE DES DONNÉES	16
5.6	ÉCHANTILLONNAGE DES PERSONNES INTERVIEWÉES	17
5.7	ACHAT D'ÉCHANTILLON POUR L'ONTARIO	17
6.	COLLECTE DES DONNÉES	19
6.1	INTERVIEWS ASSISTÉES PAR ORDINATEUR.....	19
6.2	DÉVELOPPEMENT DES APPLICATIONS DE L'ESCC.....	19
6.3	FORMATION DES INTERVIEWEURS	21
6.4	L'INTERVIEW	21
6.5	OPÉRATIONS SUR LE TERRAIN	24
6.6	CONTRÔLE DE LA QUALITÉ ET GESTION DE LA COLLECTE.....	24
7.	TRAITEMENT DES DONNÉES	26
7.1	VÉRIFICATION	26
7.2	CODAGE	26
7.3	CRÉATION DE VARIABLES DÉRIVÉES	26
7.4	PONDÉRATION.....	27
7.5	CONVERSION DU FICHIER MAÎTRE DU CYCLE 2007-2008 DE L'ESCC EN FICHIER DE MICRODONNÉES À GRANDE DIFFUSION (FMGD).....	27
8.	PONDÉRATION	30
8.1	INTRODUCTION.....	30
8.2	PONDÉRATION DE L'ÉCHANTILLON PROVENANT DE LA BASE ARÉOLAIRE.....	31
8.3	PONDÉRATION DE L'ÉCHANTILLON PROVENANT DE LA BASE TÉLÉPHONIQUE.....	33
8.4	INTÉGRATION DES BASES ARÉOLAIRE ET TÉLÉPHONIQUE	35
8.5	LES ÉTAPES DE PONDÉRATION POST-INTÉGRATION	35
8.6	PARTICULARITÉS DE LA PONDÉRATION POUR LES TROIS TERRITOIRES.....	36
9.	QUALITÉ DES DONNÉES	38
9.1	TAUX DE RÉPONSE.....	38
9.2	ERREURS DANS LES ENQUÊTES	39
9.2.1	ERREURS NON DUES À L'ÉCHANTILLONNAGE	40

9.2.2	ERREURS DUES À L'ÉCHANTILLONNAGE	40
10.	LIGNES DIRECTRICES POUR LA TOTALISATION, L'ANALYSE ET LA DIFFUSION	42
10.1	LIGNES DIRECTRICES POUR L'ARRONDISSEMENT	42
10.2	LIGNES DIRECTRICES POUR LA PONDÉRATION DE L'ÉCHANTILLON EN VUE DE LA TOTALISATION	43
10.2.1	DÉFINITIONS DES CATÉGORIES D'ESTIMATIONS : DE TYPE NOMINAL PAR OPPOSITION À QUANTITATIVES	43
10.2.2	TOTALISATION D'ESTIMATIONS DE TYPE NOMINAL	44
10.2.3	TOTALISATION D'ESTIMATIONS QUANTITATIVES	44
10.3	LIGNES DIRECTRICES POUR L'ANALYSE STATISTIQUE	45
10.4	LIGNES DIRECTRICES POUR LA DIFFUSION	46
11.	TABLEAUX DE LA VARIABILITÉ D'ÉCHANTILLONNAGE APPROXIMATIVE	48
11.1	COMMENT UTILISER LES TABLEAUX DE CV POUR LES ESTIMATIONS DE TYPE NOMINAL	48
11.2	EXEMPLES D'UTILISATION DES TABLEAUX DE CV POUR DES ESTIMATIONS DE TYPE NOMINAL	51
11.3	COMMENT UTILISER LES TABLEAUX DE CV POUR CALCULER LES LIMITES DE CONFIANCE	54
11.4	EXEMPLE D'UTILISATION DE TABLEAUX DE CV POUR OBTENIR DES LIMITES DE CONFIANCE	55
11.5	COMMENT UTILISER LES TABLEAUX DE CV POUR EFFECTUER UN TEST Z	55
11.6	EXEMPLE D'UTILISATION DES TABLEAUX DE CV POUR EFFECTUER UN TEST Z	56
11.7	VARIANCES OU COEFFICIENTS DE VARIATION EXACTS	56
11.8	SEUILS POUR LA DIFFUSION DES ESTIMATIONS RELATIVES À L'ESCC	57
12	FICHIERS DE MICRODONNÉES : DESCRIPTION, ACCÈS ET UTILISATION.....	58
12.1	FICHIERS MAÎTRES	58
12.1.1	CENTRE DE DONNÉES DE RECHERCHE	58
12.1.2	TOTALISATIONS PERSONNALISÉES	58
12.1.3.	TÉLÉ-ACCÈS	58
12.2	FICHIERS DE PARTAGE.....	59
12.3	FICHIERS DE MICRODONNÉES À GRANDE DIFFUSION.....	59
12.4	UTILISATION DES FICHIERS DE L'ESCC : FICHIER ANNUEL OU FICHIER DEUX ANS ?	59
12.5	UTILISATION DE LA VARIABLE DE PONDÉRATION	60
12.6	CONVENTION APPLIQUÉE POUR NOMMER LES VARIABLES À PARTIR DE 2007	60
12.7	CONVENTION APPLIQUÉE POUR NOMMER LES VARIABLES AVANT 2007	63
Annexe A	65
Annexe B	69
Annexe C	75
Annexe D	81
Annexe E	87
Annexe F	96

1. INTRODUCTION

L'Enquête sur la santé dans les collectivités canadiennes (ESCC) est une enquête transversale qui vise à recueillir des renseignements sur l'état de santé, l'utilisation des services de santé et les déterminants de la santé de la population canadienne. Elle est réalisée auprès d'un grand échantillon de répondants et conçue pour fournir des estimations fiables à l'échelle de la région sociosanitaire. En 2007, des changements importants ont été apportés à la conception de l'ESCC. Les données sont dorénavant recueillies sur une base continue et leur diffusion est prévue chaque année, plutôt que chaque deux ans comme c'était le cas avant 2007. Les objectifs de l'enquête ont été révisés. Ces objectifs sont :

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

Les autres changements qui ont découlé du remaniement sont détaillés à la section 3.

Les données de l'ESCC sont toujours recueillies auprès de personnes âgées de 12 ans et plus vivant dans des logements privés dans 121 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 des fichiers de microdonnées de l'ESCC et de décrire la méthodologie utilisée. L'enquête produit 3 types de fichiers de microdonnées : des fichiers maîtres, des fichiers de partage et des fichiers de microdonnées à grande diffusion (FMGD). Les caractéristiques de chacun de ces fichiers sont présentées dans ce guide. Le FMGD sera diffusé à l'été 2009 et comprendra les données du contenu collecté au cours des deux années 2007 et 2008.

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) 951-1653

Courriel : hd-ds@statcan.gc.ca

Renseignements sur le télé-accès :

(613) 951-1653

Courriel : cchs-escc@statcan.gc.ca

Télécopieur : (613) 951-0792

2. **HISTORIQUE**

En 1991, le Groupe de travail national sur l'information en matière de santé a relevé plusieurs questions et problèmes posés par le système d'information sur la santé. Selon ses membres, les données étaient fragmentées, elles étaient incomplètes, elles ne pouvaient être partagées facilement et elles n'étaient pas analysées aussi pleinement que possible; en outre, les résultats des études réalisées n'atteignaient pas de façon régulière la population 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 l'enquête ont été définis après avoir mené des consultations approfondies auprès de spécialistes et d'intervenants fédéraux, provinciaux et de régions sociosanitaires communautaires en vue de déterminer leurs exigences en matière de données².

Afin de remplir les nombreux besoins en données, le cycle de collecte des données de l'ESCC s'étendait sur deux années. Jusqu'au remaniement de 2007, la première année du cycle, indiquée par la notation « .1 », correspondait à une enquête générale sur la santé de la population conçue pour fournir des estimations fiables à l'échelle de la région sociosanitaire. La deuxième année du cycle, représentée par la notation « .2 », avait un plus petit échantillon et était conçue pour fournir des données à l'échelle provinciale sur des sujets particuliers ayant trait à la santé.

Nouvelles désignations des Cycles .1 et .2

À partir de 2007, la composante régionale du programme de l'ESCC est dorénavant collectée sur une base continue. Pour éviter toute confusion avec les enquêtes de santé thématiques, les deux composantes ont cessé d'utiliser les notations «.1» et «.2» pour se distinguer entre elles. Dorénavant, on désigne les cycles x.1 de l'ESCC comme « la composante annuelle » de l'ESCC. Le titre au long est « Enquête sur la santé dans les collectivités canadiennes – Composante annuelle, 2008 » et le titre abrégé est simplement « ESCC – 2008 ». La composante thématique de l'enquête demeure, quant à elle, inchangée. Elle continuera d'explorer plus en profondeur des sujets ou des populations plus précises. On la désignera par le nom de l'enquête suivi du sujet des thèmes couverts par chacune des enquêtes (par exemple, « Enquête sur la santé dans les collectivités canadiennes portant sur le vieillissement en santé » ou « ESCC – Vieillissement en santé »).

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. REMANIEMENT DE L'ESCC EN 2007

Jusqu'en 2005, les données de l'ESCC étaient recueillies à chaque deux ans sur une période annuelle et diffusées aux deux ans, environ 6 mois après la fin de la collecte. Le remaniement de l'ESCC de 2007 visait deux principaux points : mieux répondre aux besoins des partenaires qui désiraient augmenter le contenu de l'enquête et la fréquence des diffusions de données et assurer une meilleure utilisation des ressources opérationnelles. C'est ainsi que des changements à la conception de l'ESCC ont été proposés de manière à augmenter l'efficacité et la flexibilité de l'enquête par une collecte des données sur une base continue.

Des consultations approfondies ont été menées d'un bout à l'autre du Canada auprès de spécialistes et intervenants fédéraux, provinciaux et de régions sociosanitaires. Elles visaient à recueillir les commentaires sur les changements proposés ainsi qu'à recueillir des renseignements détaillés sur les besoins en données et produits de données des divers partenaires.

Les principaux changements qui ont découlé du remaniement de l'ESCC sont les suivants:

- Dans le passé, les données de l'ESCC étaient collectées auprès de 130 000 répondants sur une période de 12 mois. Dorénavant, la collecte de données se déroule sur une base continue. L'échantillon, qui préserve la même taille, est réparti en 12 période de collecte de 2 mois chacune. Chacune des périodes de collecte est représentative de la population vivant dans les 10 provinces canadiennes au cours des deux mois. Pour des raisons opérationnelles, l'échantillon dans les territoires est représentatif de leur population après 12 mois.
- La composante de contenu commun a été scindée en deux : le contenu de base et le contenu thématique. La composante thématique a été créée afin de tirer profit de l'approche de collecte continue. La durée de collecte des données de cette composante pourra être adaptée selon la prévalence des estimations désirées et leur niveau géographique souhaité. Le contenu de base demeurera relativement stable dans le temps. À la discrétion des provinces et des régions, le contenu optionnel peut aussi être modifié sur une base annuelle, plutôt qu'à chaque deux ans.
- Les changements à la collecte et au contenu ont inévitablement un impact sur la stratégie de diffusion. Dans le passé, les données étaient diffusées à tous les deux ans. À compter de 2008, les données de l'ESCC seront diffusées sur une base annuelle. À tous les deux ans, un fichier regroupant l'échantillon des deux années (taille de 130 000) sera également produit. En plus des ces fichiers réguliers, d'autres fichiers spéciaux seront rendus disponibles lorsque du contenu supplémentaire aura été recueilli pendant des périodes de collecte qui ne correspondent pas aux périodes annuelles standard, c'est-à-dire de janvier à décembre.
- La collecte annuelle de données est divisée en 6 périodes de 2 mois. Contrairement à ce qui se faisait avant, ces périodes ne se chevauchent plus, ce qui permet d'assurer une

surveillance plus efficace de la collecte et d'effectuer au besoin des modifications à l'interface de collecte à tous les deux mois.

4. CONTENU DE L'ESCC

En plus de données sociodémographiques et administratives, le contenu de l'ESCC comporte trois composantes qui répondent chacune à des besoins différents : la composante de contenu commun inclut le contenu de base et le contenu thématique, la composante de contenu optionnel et la composante de réponse rapide. L'annexe A présente la liste des modules qui ont été inclus dans le questionnaire de 2008 selon leur composante.

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

Tableau 4.1 Durée de l'enquête selon les composantes

Composante de l'ESCC	Durée moyenne
Contenu commun <ul style="list-style-type: none">• Contenu de base• Contenu thématique	30 minutes (20 minutes) (10 minutes)
Contenu optionnel	10 minutes
Contenu de réponse rapide (facultatif)	2 minutes

4.1 Contenu commun

La composante de contenu commun de l'ESCC comprend des questions qui sont demandées aux répondants de toutes les provinces et territoires (à moins d'exception). Elle est divisée en deux volets : le contenu de base et le contenu thématique.

Le contenu de base est formé de questions qui sont demandées à tous les répondants de l'enquête. Ces questions demeureront relativement inchangées dans le questionnaire pour une période d'environ 6 ans, à moins qu'un souci important vis-à-vis la qualité le requiert.

Le contenu thématique est formé de questions qui sont reliées à un sujet spécifique. Le contenu thématique est formé d'un thème de deux ans et d'un thème annuel et compte pour 10 minutes de l'entrevue. Certains thèmes pourraient être réintroduits dans l'enquête à tous les 2, 4 ou 6 ans au besoin. Cette composante permet à l'ESCC une meilleure planification du contenu à moyen terme.

Certains modules inclus dans la composante thématique peuvent être demandés à un sous-échantillon de répondants si l'objectif de ces questions est de fournir des estimations fiables à l'échelle nationale ou provinciale, plutôt qu'à l'échelle de la région sociosanitaire. Cette approche est utilisée dans le but de minimiser le fardeau de réponse et les coûts qui s'y rattachent.

4.2 Contenu optionnel

La composante de contenu optionnel offre aux régions sociosanitaires de choisir du contenu qui répond aux priorités provinciales et régionales en matière de santé publique. Le contenu optionnel est sélectionné à partir d'une longue liste de modules disponibles pour inclusion dans l'ESCC. Les modules de contenu choisis par une région ne sont demandés qu'aux résidents des régions qui ont sélectionné ces modules. En réalité, depuis le cycle 3.1 de l'ESCC, les régions et les provinces ont choisi de coordonner la sélection du contenu optionnel de manière à uniformiser la sélection des modules optionnels à l'échelle provinciale. Le contenu optionnel peut varier d'une année à l'autre selon les besoins et doit être revu chaque deux ans.

Il convient de mentionner que contrairement au contenu commun, les données provenant des modules de contenu optionnel ne peuvent être généralisées à l'échelle du Canada³.

L'annexe B présente les résultats de la sélection du contenu optionnel pour l'année en cours selon la province de résidence.

4.3 Contenu de réponse rapide

La composante de réponse rapide est offerte contre recouvrement des coûts aux organisations désirant obtenir des estimations nationales sur un sujet émergent ou particulier lié à la santé de la population. La réponse rapide est formée d'un maximum de deux minutes de contenu. Les questions apparaissent au questionnaire pendant une seule période de collecte (deux mois) et sont demandées à tous les répondants de l'ESCC au cours de cette période.

4.4 Contenu des fichiers de microdonnées

L'enquête produit des fichiers annuels. Un fichier combinant les deux années les plus récentes est également diffusé à chaque deux ans. En juin 2009, deux fichiers de microdonnées ont été diffusés : un fichier annuel fondé sur la période de référence de 2008, qui s'apparente au fichier de données de 2007, ainsi qu'un fichier de données fondé sur la période de référence de 2007-2008. Le fichier principal de 2007-2008 est de la même taille que les fichiers des cycles précédents (2001, 2003 et 2005), soit environ 130 000 répondants.

Le fichier de deux ans comprend l'ensemble des questions qui sont demeurées dans l'enquête au cours des deux années. À moins d'exception, il s'agit de la composante des questions du contenu de base, du contenu thématique de deux ans et le contenu optionnel choisi pour deux ans. Les thèmes d'un an et le contenu optionnel choisi pour une année seulement ne sont pas disponibles dans le fichier de données de deux ans.

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

Par ailleurs, les modules thématiques collectés auprès d'un sous-échantillon de la population continueront d'être diffusés dans des fichiers séparés. Ces derniers incluent le contenu de base et le contenu des modules thématiques collectés auprès d'un sous-échantillon de répondants. Le tableau 4.2 fournit des précisions concernant la disponibilité du contenu dans les fichiers de 2007, 2008 et 2007-2008 combinées.

Tableau 4.2 Composantes de contenu comprises dans les fichiers de données de 2007 et 2008

Fichiers		Contenu de base	Contenu thématique de 2007 ¹	Contenu thématique de 2008 ²	Contenu thématique de 2007-2008	Contenu optionnel ³
Fichiers de 2007	Principal	Oui	Non	S/O	Oui	Oui
	Sous-échantillon (3 modules)	Oui	Oui	S/O	Non	Non
Fichiers de 2008	Principal	Oui	S/O	Oui (sauf le module de sous-échantillon)	Oui	Oui
	Sous-échantillon (1 module)	Oui	S/O	Un module (Poids et taille mesurés)	Non	Non
Fichier de 2007-2008	Principal	Oui	Non	Non	Oui	Oui

¹ Le thème de 2007 était composé de 3 modules (Satisfaction des patients, Accès aux services de santé et Temps d'attentes) qui ont été posés à un sous-échantillon de répondants.

² Le thème de 2008 comprend un groupe de modules reliés à la détection des maladies chroniques ainsi qu'un module sur le poids et la taille mesurés. Ce dernier est demandé à un sous-échantillon de répondants.

³ Le contenu optionnel est inclus dans le fichier de 2007-2008 s'il a été demandé aux répondants d'une province pendant les deux années. Autrement, il sera inclus uniquement dans le fichier pour l'année pendant laquelle il a été recueilli. Il est à noter que si un module thématique un an est choisi dans le contenu optionnel d'une juridiction au cours de la deuxième année, le module est inclus dans le fichier de deux ans et traité comme contenu optionnel.

5. PLAN D'ÉCHANTILLONNAGE

5.1 Population cible

L'ESCC vise la population de 12 ans et plus vivant à domicile et résidant dans les dix provinces et 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. En collaboration avec les provinces, Statistique Canada est parfois appelé à modifier les limites de certaines RSS afin qu'elles correspondent aux données géographiques du Recensement ou encore afin de mieux tenir compte des besoins en données sur la santé selon de nouvelles bornes géographiques. Pour l'ESCC 2007-2008, des données ont été recueillies pour 118 RSS dans les 10 provinces, ainsi que pour une RSS par territoire, soit, en tout, pour 121 RSS (Annexe C).

5.3 Taille et répartition de l'échantillon

Afin de produire des estimations fiables pour chaque RSS et compte tenu du budget accordé pour l'ESCC, il a été établi que cette composante de l'enquête doit être réalisée auprès d'un échantillon d'environ 130 000 personnes sur une période de 2 ans. La production d'estimations fiables pour chaque RSS était l'objectif principal, mais la qualité des estimations pour certaines caractéristiques importantes pour les provinces a été jugée importante également. Par conséquent, la stratégie de répartition de l'échantillon, qui comporte trois étapes, accorde une importance plus ou moins égale aux RSS et aux provinces. A la première étape, on impose une taille minimum de 500 répondants par RSS. C'est le minimum considéré afin d'obtenir un niveau de qualité de données raisonnable. Par contre, pour des raisons de fardeau de réponse, on impose une fraction de sondage maximum de 1 sur 20 logements afin d'éviter d'échantillonner trop de logements dans des petites régions qui sont aussi interpellées par d'autres enquêtes. Notons que très peu de RSS ont une taille de moins de 500 due à la limite de la fraction de sondage. Cette première étape a réparti au total 60,350 unités. La deuxième étape consiste à répartir le reste de l'échantillon disponible par province en suivant une répartition proportionnelle à la taille de la population par province. La taille totale de l'échantillon par province est donc la somme des tailles établies aux deux premières étapes. Il est à noter que la stratégie de répartition de l'échantillon a été utilisée pour l'ESCC au cycle 3.1 et que les tailles alors établies ont restées sensiblement les mêmes depuis. L'échantillon est ensuite divisé également pour chacune des 2 années de collecte. Le tableau 5.1 donne la taille d'échantillon totale pour les années 2007 et 2008.

Tableau 5.1 Nombre de régions sociosanitaires et tailles visées d'échantillon selon la province/territoire pour l'ESCC 2007-2008

Province	Nombre de RSS	Taille totale de l'échantillon (visée)
Terre-Neuve-et-Labrador	4	4 010
Île-du-Prince-Édouard	3	2 480
Nouvelle-Écosse	6	5 040
Nouveau-Brunswick	7	5 160
Québec	16	24 290
Ontario ¹	36	44 460
Manitoba	10	7 510
Saskatchewan	11	7 720
Alberta	9	12 210
Colombie-Britannique	16	16 100
Yukon	1	1 200
Territoires du Nord-Ouest	1	1 200
Nunavut	1	700
Canada	121	132 080

¹La taille de l'échantillon pour l'Ontario inclut l'achat d'échantillons supplémentaires par réseau local d'intégration des services de santé (RLISS). La taille initiale pour l'Ontario avant l'achat était de 20 900 unités (se référer à la section 5.7 pour plus de détails)

À la troisième étape, l'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 proportionnelle à la taille par province.

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. Au total, pour 2007-2008, 1200 unités d'échantillonnage ont été attribuées au Yukon, 1200 aux Territoires du Nord-Ouest et 700 au Nunavut. Ces tailles sont déterminées selon le budget disponible. La répartition de l'échantillon dans les territoires est faite proportionnellement à la taille de la population dans les strates. Les strates utilisées sont les mêmes que celles définies par l'enquête sur la population active (EPA) qui regroupent un ensemble de collectivités. Pour plus de détails, se reporter à la section 5.4.1.

L'échantillon est ensuite divisé également en 2, de sorte à obtenir les mêmes tailles d'échantillon entre l'échantillon de la base aréolaire et l'échantillon de la base liste pour chaque RSS⁴ tel qu'élaboré dans la section suivante. Mentionnons finalement que l'effectif des échantillons tirés de chaque base a été augmenté avant la collecte des données afin de tenir compte des unités hors du champ de l'enquête et du taux de non-réponse anticipés à partir des taux obtenus des cycles

⁴ À l'exception de 2 régions socio-sanitaires utilisant la base de sondage à composition aléatoire (CA) seulement (section 5.4.3) et les 3 territoires utilisant seulement la base aréolaire et la base de sondage à composition aléatoire (sections 5.4.1 et 5.4.3).

précédents de l'ESCC. Les tailles d'échantillons par RSS et par base de sondage sont données à l'annexe D.

5.4 Bases de sondage et stratégies d'échantillonnage des ménages

L'échantillon de ménages 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'échantillonnage⁵. À la première étape, des strates homogènes sont formées et des échantillons indépendants de grappes sont sélectionnés dans chaque strate. À la deuxième étape, une liste de logements pour chaque grappe est créée, puis des logements sont sélectionnés dans chaque liste.

Pour les besoins du plan d'échantillonnage de l'EPA, chaque province est répartie en trois catégories de région, soit 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 aires de diffusion (AD) du recensement pour cibler les ménages à haut revenu, les immigrants et les autochtones. Dans chaque strate, six grappes ou des bâtiments d'habitation (parfois 12 ou 18 appartements) sont sélectionnées par une méthode d'échantillonnage aléatoire avec 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, six grappes (habituellement des AD du recensement) sont sélectionnées 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 la liste des nouvelles grappes établie, l'échantillon est tiré par échantillonnage systématique des logements. La taille de chaque échantillon systématique est appelée le 'rendement'. Le tableau 5.2 donne un aperçu des catégories d'UPE utilisées dans l'échantillon de l'EPA et le rendement prévu par échantillon systématique. Comme les taux d'échantillonnage sont prédéterminés, il existe souvent un écart entre la taille prévue d'échantillon et les chiffres obtenus. Par exemple, le rendement de l'échantillon est parfois excessif. Cette situation peut se présenter

⁵ Statistique Canada (1998). *Méthodologie de l'enquête sur la population active du Canada*. Statistique Canada numéro 71-526-XPB au catalogue.

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, la production excessive est corrigée en éliminant, dès le départ, une partie des unités sélectionnées. Cette modification est considérée lors de la pondération.

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 de l'ESCC, certaines modifications ont dû être apportées à cette stratégie d'échantillonnage. Pour obtenir un échantillon annuel de 33,000 répondants pour l'ESCC par année, il fallait sélectionner près de 47 000 logements de la base aréolaire afin de 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 économiques des dix provinces, alors que, pour l'ESCC, il fallait obtenir un échantillon d'environ 47 000 logements réparti par RSS dont les limites géographiques différaient de celles des régions économiques de l'EPA. Globalement, l'ESCC nécessitait la sélection d'un nombre inférieur de logements que produit le mécanisme de sélection de l'EPA, ce qui correspond à un *facteur de redressement* de 0,78 (47 000/60 000). Toutefois, comme ce facteur de redressement variait de 0,3 à 3,0 au niveau des RSS, certains ajustements ont été nécessaires.

Les modifications apportées au processus de sélection dans les régions variaient selon le facteur de redressement. Pour les RSS où le facteur était égal ou inférieur à 1, le nombre d'UPE choisi a été réduit si nécessaire. Par exemple, si le facteur était de 0,5, alors seulement 3 UPE ont été choisies dans chaque strate de l'EPA au lieu du nombre habituel de 6 UPE. Pour les RSS avec un facteur supérieur à 1 mais égal ou inférieur à 2, le processus d'échantillonnage des logements à l'intérieur d'une UPE a été répété pour un sous-ensemble des UPE sélectionnées appartenant à la RSS en question. Par exemple, si le facteur était de 1,6, alors la sélection des logements a été répétée dans 4 des 6 UPE pour chaque strate dans la RSS. Lorsque la répétition de la sélection de logements à l'intérieur d'une UPE était nécessaire mais qu'aucun autre logement n'était disponible dans cette UPE, alors une autre UPE a été choisie. Lorsque le facteur était supérieur à 2, le processus d'échantillonnage des logements a été répété à l'intérieur d'autres UPE appartenant à la RSS en question⁶.

⁶ Pour réduire les coûts de listage, le processus de sélection a été répété jusqu'à 3 fois dans les UPE déjà choisies, pour certaines régions urbaines seulement. Il s'agit par contre de cas d'exception.

Finalement, lorsque le nombre de logements disponibles dans les UPE sélectionnées était supérieur au nombre de logements requis, un sous-échantillon a été tiré. Ce processus est appelé la 'stabilisation'.

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

Pour des raisons opérationnelles, le plan d'échantillonnage à partir de la base aréolaire de l'EPA pour les trois territoires est différent. Pour chaque territoire, les collectivités (municipalités) faisant partie du champ de l'enquête sont groupées en strates en fonction de diverses caractéristiques (population, données géographiques, proportion d'Inuit et/ou d'Autochtones et revenu médian du ménage). L'EPA a défini 5 strates pour le Yukon, 10 pour les Territoires du Nord-Ouest et 6 pour le Nunavut. Le premier degré d'échantillonnage a consisté à sélectionner aléatoirement une collectivité avec probabilité proportionnelle à la taille de la population dans chaque strate. Puis, à l'intérieur de chaque collectivité, une stratégie d'échantillonnage des ménages est mise en place de façon identique à celle décrite plus haut. L'ESCC a sélectionné son échantillon à partir des mêmes collectivités sélectionnées par l'EPA tout en s'assurant de sélectionner des logements différents. Si plus ou pas assez de logements étaient disponibles pour une collectivité à l'intérieur d'une strate, l'EPA a choisi une autre collectivité pour l'ESCC.

Il convient de mentionner que la base de sondage de l'ESCC 2007-2008 couvrait 90 % des ménages privés du Yukon, 97 % de ceux des Territoires du Nord-Ouest et 71 % de ceux du Nunavut⁷.

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

À l'exception de 5 RSS (les 2 RSS provenant de la base CA et les trois territoires), la base liste de numéros de téléphone a été utilisée dans toutes les régions pour compléter la base aréolaire. La base liste est l'annuaire téléphonique du Canada, une base administrative externe de numéros de téléphones contenant les noms, les adresses et les numéros de téléphone répertoriés dans les annuaires de téléphone du Canada et qui est mise à jour à tous les 6 mois. Elle a été stratifiée par RSS en utilisant un fichier de conversion de codes postaux de sorte à pouvoir associer un RSS à chaque numéro de téléphone. Dans chaque RSS, un échantillon de numéros de téléphone a été tiré par échantillonnage aléatoire simple. Comme pour la base de sondage à CA, des numéros de téléphone supplémentaires ont été sélectionnés pour tenir compte des numéros hors service ou hors du champ d'observation.

Il importe de souligner que la sous-couverture de la base liste de numéros de téléphone est plus 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 base liste est toujours utilisée en complément de la base aréolaire, l'effet de la sous-couverture dû à l'utilisation de la base liste de numéros de téléphone est minimal et est corrigé lors de la pondération.

⁷ Au Nunavut, pour des raisons d'ordre opérationnel inhérentes aux collectivités éloignées, seulement les 10 communautés les plus grandes sont couvertes par l'enquête soient Iqaluit, Cambridge Bay, Baker Lake, Arviat, Rankin Inlet, Kugluktuk, Pond Inlet, Cape Dorset, Pangnirtung et Igloodik.

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 générale⁸. 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, une banque de cent numéros a été choisie au hasard et un numéro compris entre 00 et 99 a été généré aléatoirement 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 25 % à 50% parmi les 4 RSS qui y ont eu recours.

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

Afin d'équilibrer la charge de travail des intervieweurs et de réduire au minimum les effets saisonniers éventuels sur les estimations de 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 pour chaque période de collecte de 2 mois.

Pour l'échantillon de la base aréolaire, chaque UPE sélectionnée dans chaque RSS a été répartie au hasard à une période de collecte en tenant compte de plusieurs contraintes reliées aux opérations sur le terrain ou encore à la pondération tout en conservant une taille égale par période. Par exemple, on s'est assuré d'avoir un échantillon représentatif de la population canadienne aux 6 mois en s'assurant d'avoir un échantillon de logements couvrant toutes les strates de l'EPA sur cette période.

⁸ 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.

Pour les listes des numéros de téléphone, des échantillons indépendants ont été sélectionnés à chaque période de collecte. Cette stratégie permet d'assurer que chaque échantillon soit représentatif de la population canadienne faisant partie du champ d'observation de l'enquête à chaque 2 mois.

5.6 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 a tenu compte des besoins des utilisateurs de données, du coût, de l'efficacité du plan d'échantillonnage, du fardeau de réponse et des contraintes opérationnelles. Une personne est sélectionnée par ménage en utilisant diverses probabilités de sélection variant selon l'âge et selon la composition du ménage. Les probabilités choisies font suite à des résultats de simulations reposant sur divers paramètres dans le but de déterminer l'approche optimale sans générer de poids d'échantillonnage extrêmes en 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 groupe d'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 aurait 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

5.7 Achat d'échantillon pour l'Ontario

La province de l'Ontario a demandé une augmentation de l'échantillon afin de produire des estimations au niveau de la géographie des réseaux locaux d'intégration des services de santé (RLISS). L'Ontario compte 14 RLISS (voir Annexe Y). L'échantillon de l'ESCC a été augmenté de sorte à obtenir une taille minimum de 2000 par RLISS sur une période de 2 ans. Comme les limites des RSS et RLISS s'entrecoupent, le niveau de stratification utilisé a été le croisement RSS-RLISS. Les tailles d'échantillon préalablement alloué par RSS ont donc été conservées. Dans les cas où la répartition par RSS n'a pas permis d'atteindre des tailles de 2000 par RLISS, l'échantillon a été augmenté en conséquence et répartie proportionnellement à la taille de la population dans le croisement RSS-RLISS. Le tableau 5.4 donne la taille d'échantillon de répondants visés par RLISS pour les années 2007-2008.

Tableau 5.4 Tailles visées de répondants par réseau local d'intégration des services de santé (RLISS) dans l'ESCC 2007-2008

RLISS	Taille totale de l'échantillon (visée)
01-Erie St. Clair	3 104
02-South West	5 136
03-Waterloo Wellington	2 460
04-Hamilton Niagara Haldimand Brant	5 248
05-Central West	2 124
06-Mississauga Halton	2 272
07-Toronto Central	2 170
08-Central	2 808
09-Central East	4 152
10-South East	2 688
11-Champlain	4 116
12-North Simcoe Muskoka	2 116
13-North East	3 984
14-North West	2 082
Ontario	44 460

La taille totale de l'échantillon des croisements RSS-RLISS a ensuite été répartie également entre la base liste et la base aréolaire. Les procédures normales de sélection de l'échantillon dans chaque base ont été appliquées sur l'échantillon total. L'échantillon supplémentaire fait partie intégrante de l'échantillon de l'ESCC. Les tailles d'échantillon par réseau local d'intégration des services de santé et par base de sondage sont données à l'annexe D.

6. COLLECTE DES DONNÉES

6.1 Interviews assistées par ordinateur

Entre janvier 2007 et décembre 2008, un total de 131 959 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

L'ESCC utilise deux applications d'IAO distinctes pour la collecte de données, l'une pour les interviews téléphoniques (ITAO), l'autre pour les interviews sur place (IPAO). Cette façon de faire permet d'adapter la fonctionnalité de chaque application au type d'interview menée. Chaque application comporte 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. Chacune des applications a été l'objet de trois étapes de mise à l'essai: les tests modulaires, intégrés et de bout en bout.

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

Les 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 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

Les gestionnaires de projet, les intervieweurs principaux et les intervieweurs des bureaux régionaux responsables de la collecte des données de l'ESCC ont reçu une trousse de formation pour de l'auto-apprentissage avant le début de la collecte. Ces trousse ont été préparées par l'équipe projet de l'ESCC en vue d'être utilisées par les intervieweurs expérimentés de l'ESCC afin de leur procurer une mise à niveau de leurs connaissances. Les gestionnaires de projet et les intervieweurs principaux ont également mené, au besoin, des séances de formation pour les nouveaux intervieweurs. Enfin, des séances de formation mettant l'emphase sur des sujets spécifiques reliés à la collecte des données de l'ESCC se sont déroulées sur une base mensuelle.

L'objet des séances de formation était de familiariser les intervieweurs avec les applications de l'enquête, de les familiariser avec le contenu et de les introduire aux procédures d'entrevue spécifiques à l'ESCC 2007 et 2008. La formation était centrée sur :

- les buts et objectifs de l'enquête, incluant une partie centrée sur le remaniement de l'enquête;
- les techniques d'enquête;
- les fonctionnalités des applications;
- le contenu des questionnaires, qui a été l'objet d'exercices avec emphase sur les changements significatifs apportés au contenu;
- 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 (73,9 %) 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 SAM_TYP apparaissant dans les fichiers de microdonnées 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 ADM_N09 apparaissant dans les fichiers de microdonnées 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 les fichiers de microdonnées, des indicateurs signalent si une personne autre que le répondant était présente à l'interview (ADM_N10) et si, de l'avis de l'intervieweur, la présence de l'autre personne a influencé les réponses du répondant (ADM_N11).

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

a) Lettres d'introduction

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

b) Prise de contact

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.

c) 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.

d) Obstacles linguistiques

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

e) Interviews de jeunes

Les intervieweurs 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.

f) 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 ADM_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 de 2007/2008 a été répartie en douze périodes de collecte de deux mois chacune qui ne se chevauchent pas. 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.

Les bureaux de collecte centralisés ont reçu les fichiers d'échantillons environ deux semaines avant le début de chaque période de collecte. Chaque échantillon IPAO 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.

Au terme de la collecte de données, un taux de réponse national de 76 % a été atteint. L'annexe E 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 des données de 2007 et 2008. 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 IPAQ, 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.

Le Bureau central a effectué des analyses complémentaires au moyen de l'information tirée des applications d'IAQ 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, le protocole de l'ESCC permet des interviews par procuration lorsque le répondant choisi est mentalement ou physiquement incapable de mener l'interview. Les interviews par procuration ont une valeur limitée étant donné que plusieurs modules ne sont pas demandés pour des raisons reliées à la qualité des données. Il est donc important de déterminer les taux d'interview par procuration et les motifs derrière ces interviews. Un rapport d'interview par procuration a été créé afin de suivre leur progression. Ces rapports ont permis d'identifier les intervieweurs qui semblaient mal comprendre les circonstances qui rendaient la conduite d'un interview par procuration acceptable. Ces derniers ont bénéficié ensuite d'un rappel ou d'une formation supplémentaire axée sur les critères particuliers à appliquer dans le cas d'interview par procuration.

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

Pour faciliter l'analyse des données, on a dérivé un certain nombre de variables à partir des éléments disponibles sur le questionnaire de l'ESCC. Le quatrième caractère du nom des variables dérivées est en général un « D », « G » ou un « F ». Dans certains cas, les variables dérivées sont simples, donnant lieu à un regroupement des catégories de réponses. Dans d'autres cas, on a combiné plusieurs variables pour en créer une nouvelle. La *Documentation sur les variables dérivées (VD)* fournit des détails sur la façon de dériver ces variables plus complexes. Pour de plus amples renseignements concernant la nomenclature, veuillez vous référer à la section 12.5.

7.4 Pondération

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

L'étape de détermination des facteurs de pondération donne lieu au calcul du poids d'échantillonnage de chaque personne échantillonnée. Ce poids apparaît dans 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 2007-2008 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 121 régions sociosanitaires au Canada au moment de la conception du plan d'échantillonnage de 2007-2008. Ainsi, le fichier de données maître contient de l'information représentative de 121 régions sociosanitaires.

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 29 régions sociosanitaires, dans toutes les provinces sauf le Québec, la Nouvelle-Écosse, l'Alberta, et le Colombie Britannique en 13 ensembles;
- ii) le regroupement des 3 territoires (qui provoque la suppression de tout 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 102 régions sociosanitaires, particulières ou regroupées, d'un océan à l'autre. L'annexe F illustre les régions, avant et après regroupement.

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 de l'ESCC de 2007-2008. 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.

Depuis 2007, les provinces et les régions ont la possibilité de modifier la sélection de contenu optionnel à chaque année. Les fichiers maître et FMGD de 2007-2008 ne contiennent que les données du contenu optionnel pour les modules qui ont été demandés au cours des deux années, ce qui comprend la grande majorité du contenu optionnel.

Contrôle de la divulgation : Comme il a été mentionné déjà, le FMGD de 2007-2008 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, souffre du cancer de la peau 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 de 2007-2008, sans toutefois que la valeur analytique des données en soit altérée.

Âge des répondants : Le FMGD de 2007-2008 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 3.1 distinguaient des tranches de 12 à 14 ans et de 15 à 19 ans. Depuis l'ESCC 3.1 le FMGD 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 (LBS), le stress au travail (WST) et les activités physiques - Installations au travail (PAF) que pour les répondants âgés de 15 à 74 ans. Dans le même ordre d'idées, les données sur l'éducation (EDU) des répondants âgés de 14 ans et celles qui concernent les expériences maternelles (MEX, MXS, et MXA) des répondantes de 55 ans ont aussi été éliminées.

Exceptionnellement, les données sur les expériences maternelles (MEX, MXS, et MXA) des adolescentes de 15 à 19 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 20 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 de l'enquête.

Tel que décrit dans la section 5, l'ESCC a 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.

Selon les besoins, une seule ou deux bases peuvent être utilisées pour la sélection de l'échantillon dans une région sociosanitaire (RSS). Quand on utilise deux bases, la stratégie de pondération traite indépendamment la base aréolaire et la base téléphonique pour dériver les poids au niveau de ménage séparé pour chaque base utilisée. Ces poids des ménages sont ensuite combinés en un seul ensemble de poids des ménages lors d'une étape appelée « *intégration* ». Suite à la transformation des poids des ménages en poids de personne et à quelques autres ajustements, ce poids intégré devient le poids de personne final.

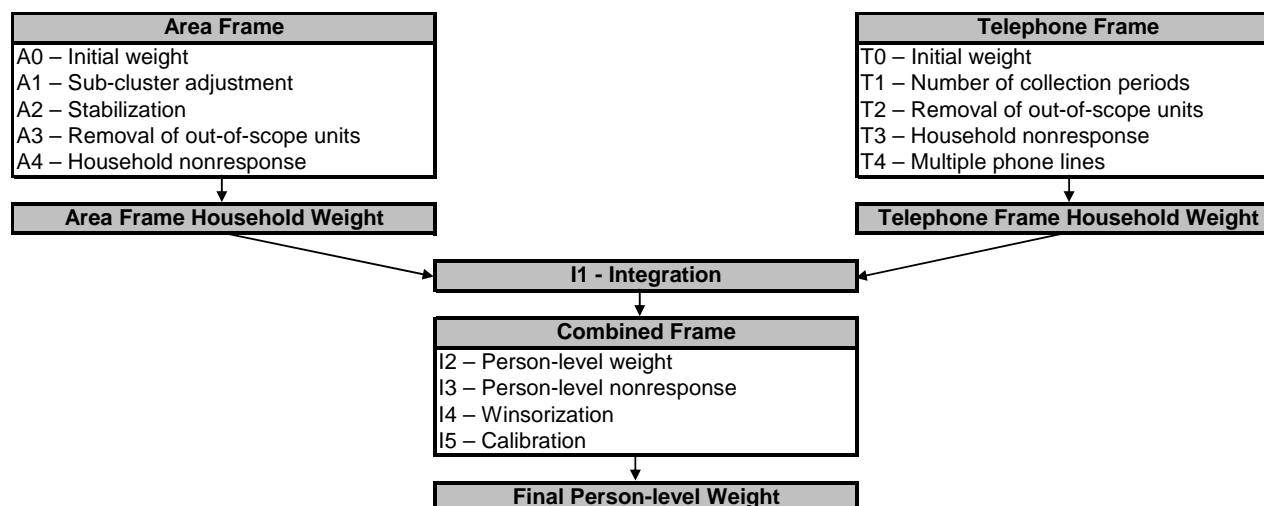
8.1 Introduction

Tel que mentionné plus haut, les unités des bases aréolaire et téléphonique sont traitées séparément jusqu'à l'étape d'intégration. Les sections suivantes décrivent la stratégie de pondération pour les provinces. La sous-section 8.2 fournit les détails de la stratégie de pondération pour la base aréolaire, puis la sous-section 8.3, ceux pour la base téléphonique. L'intégration des deux bases est traitée en 8.4. Puis, suivent les deux étapes finales de la pondération, c'est-à-dire l'ajustement pour contrôler la saisonnalité des données puis le calage aux marges, qui sont expliquées dans la sous-section 8.5.

Malgré que les deux bases aient été utilisées pour couvrir les trois territoires, les méthodes d'échantillonnage utilisées ont été légèrement modifiées pour les territoires. Ces modifications affectent substantiellement la pondération pour ces trois régions, et celles-ci sont rapportées dans la sous-section 8.6.

Le diagramme A présente un sommaire des différents ajustements faisant partie de la stratégie de pondération. 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



8.2 Pondération de l'échantillon provenant de la base aréolaire

A0 – Poids initial

La pondération pour la base aréolaire débute avec un poids fourni par l'Enquête sur la population active (EPA). Ce poids est basé sur le plan de sondage de l'EPA puisque le plan de sondage de base aréolaire de CCHS découle de celui de l'EPA. Le plan de sondage de l'EPA consiste en un échantillonnage de logements dans les grappes sélectionnées parmi les strates de l'EPA. Lors de l'ajustement initial, le poids de l'EPA est ajusté afin de tenir compte du fait que l'ESCC sélectionne un échantillon représentatif au niveau des régions sociosanitaires (RSS). De plus, l'ESCC sélectionne un nombre de grappes différent de l'EPA et peut répéter l'échantillonnage de logements dans les grappes sélectionnées. Le poids résultant est appelé A0. Pour plus de détails sur le mécanisme de sélection, de même qu'une définition plus complète des strates de l'EPA et des grappes, se référer à Statistique Canada (1998)⁹.

A1 – Ajustement de sous-poids de grappe

Dans les grappes où un accroissement significatif de la population est observé, une méthode de sous-échantillonnage est utilisée afin de ne pas augmenter indûment la tâche de l'interviewer. La méthode alors appliquée par l'EPA pourra être de sous-échantillonner les logements dans cette grappe, de diviser cette grappe en sous-grappes ou de redéfinir cette grappe comme étant une strate et de créer de nouvelles grappes dans cette strate. Dans tous ces cas, un ajustement de sous-poids de grappe est calculé et appliqué aux poids de CCHS. Cet ajustement est multiplié par le poids A0 pour produire le poids A1. Encore une fois, plus de détails sont disponibles dans la documentation de l'EPA (Statistique Canada (1998)).

⁹ Statistique Canada (1998). *Méthodologie de l'enquête sur la population active du Canada*. Statistique Canada numéro 71-526-XPB au catalogue.

A2 – Stabilisation

Dans certaines RSS, l'accroissement de l'échantillon, tel que décrit à la section 5, résulte 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 consiste à sous-échantillonner des logements aléatoirement à l'intérieur de la RSS parmi les logements originalement sélectionnés dans chaque grappe. Un facteur d'ajustement représentant l'effet de la stabilisation est donc 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 l'ESCC. Ces logements sont tout simplement retirés de l'échantillon, ne laissant plus que les logements faisant partie du champ de l'enquête. Les logements ou ménages qui restent dans l'échantillon conservent 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 échantillonnés 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 groupes homogènes de réponse (GHR). Dans le but de créer ces GHR, la méthode du score basée sur une régression logistique est utilisée afin de déterminer la probabilité de réponse puis, ces probabilités sont utilisées afin de diviser l'échantillon par groupes ayant des propriétés de réponse similaire. L'information disponible pour les non-répondants étant limitée, le modèle de régression comprend donc des variables comme la période de collecte et des informations géographiques ainsi que des para-données ce qui inclut le nombre d'essais pour contacter le ménage, l'heure/jour des essais et si les essais ont été faits au cours de la semaine ou de la fin de semaine. Un facteur d'ajustement a donc été calculé à l'intérieur de chaque GHR 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 est donc 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.

8.3 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 pour une RSS donnée. Les unités provenant de ces deux bases téléphoniques sont toutefois traitées ensemble et sont donc toutes soumises aux mêmes ajustements.

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 les premières trois ajustements (T0, T1 et T2) relativement à la RSS assignée lors de la sélection de l'échantillon. Les 2 ajustements restant (T3 et T4) sont appliqués à des RSS basé sur l'information recueillie de la répondant pour s'assurer que tous les unités font partie de leur vraie région dès la sélection de l'échantillon.

T0 – Poids initial

Le poids initial est défini comme l'inverse de la probabilité de sélection et il 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 Paton¹⁰ pour plus de détails). Conséquemment, la probabilité de sélection est le ratio entre le 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. Le ratio est basé sur la version de la base disponible et du nombre d'unités sélectionnées pour une période de collecte donnée. Pour cette raison, la probabilité de sélection peut changer selon la répartition de l'échantillon et les mises à jour de la base. L'inverse de ces probabilités de sélection représente le poids initial T0.

¹⁰ Norris, D.A. et Paton, D.G. (1991). L'Enquête sociale générale canadienne: bilan des cinq premières années. *Techniques d'enquête*. 17, 245-260.

T1 – Nombre de périodes de collecte

Contrairement à la base aréolaire, pour laquelle l'échantillon est sélectionné entièrement au début du processus d'échantillonnage, des échantillons sont tirés à chaque deux mois pour les bases téléphoniques. 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 doit être appliqué pour réduire les poids de chaque échantillon mensuel. Le facteur d'ajustement appliqué à chaque échantillon mensuel est égal à l'inverse du nombre d'échantillons combinés ou encore, du nombre de périodes de collecte. À partir de ce moment, l'échantillon de la base liste correspond à la moyenne des échantillons des périodes de collecte combinées. Les poids initiaux sont multipliés par cet facteur d'ajustement de produire le poids T1.

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 – 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). Par contre, les para-données utilisées diffèrent puisque des applications de collecte différentes sont utilisées pour les interviews en personne et par téléphone. 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:

$$\frac{\textit{Somme des poids T2 pour tous les ménages}}{\textit{Somme des poids T2 pour tous les ménages répondants}}$$

Le poids T2 des ménages répondants a donc été multiplié par ce facteur d'ajustement pour produire le poids T3. Les ménages non-répondants sont éliminés à partir de ce point.

T4 - 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. Le facteur d'ajustement représente l'inverse du nombre de lignes dans le ménage et le poids T4 est obtenu en multipliant ce facteur par le poids T3.

8.4 Intégration des bases aréolaire et téléphonique

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'intégration¹¹. Le poids des unités qui se trouvent sur la base aréolaire mais pas sur la base téléphonique n'est pas ajusté. Pour toutes les autres unités, un facteur d'ajustement α , compris entre 0 et 1, est appliqué au poids. 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 de ménage final calculé auparavant (A4 ou T4 dépendant de quelle base provient l'unité), procure le poids intégré I1.

8.5 Les étapes de pondération post-intégration

I2 – 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 I1 par l'inverse de la probabilité de sélection de la personne choisie dans le ménage. Nous obtenons ainsi le poids I2. 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 5.6 pour plus de détails).

I3 – Non-réponse personne

Dans le cadre 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 groupes homogènes de réponse. Dans ce processus, la méthode par score était utilisée pour définir une probabilité de réponses définies à partir des caractéristiques disponibles pour les répondants et non-répondants. Toutes les caractéristiques recueillies lors du listage des membres du ménage, en plus de l'information géographique et des para-données, étaient en fait disponibles pour estimer les probabilités de réponse. Ces probabilités étaient utilisées afin de définir les classes et un facteur d'ajustement est calculé à l'intérieur de chaque classe de la façon suivante:

$$\frac{\text{Somme des poids I2 pour toutes les personnes sélectionnées}}{\text{Somme des poids I2 pour toutes les personnes sélectionnées répondantes}}$$

¹¹ 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.

Le poids I2 des personnes répondantes a donc été multiplié par ce facteur d'ajustement pour produire le poids I3. Les personnes non-répondantes sont éliminées de la pondération à partir de ce point.

I4 – Winsorization

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. Ces unités peuvent effectivement représenter une proportion anormalement élevée de leur RSS et ainsi influencer fortement les estimations, et la variance, de ces RSS. Afin d'éviter cette situation, le poids des répondants qui contribuent de façon aberrante est ajusté à la baisse selon une méthode « winsorization ».

I5 - Calage aux marges

La dernière étape nécessaire afin d'obtenir le poids final de l'ESCC est la calage au marges (I5). La calage est appliquée en utilisant CALMAR¹² 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. Dans un même temps, les poids sont ajustés afin de s'assurer que chaque période de collecte (de deux mois) est également représenté par l'échantillon. Il est à noter que la calage au marges a été faite en utilisant une géographie la plus à jour possible qui diffère peut-être de la géographie utilisée lors de l'échantillonnage.

Les estimations de population sont basées sur les comptes du Recensement les plus récentes, de même que sur les comptes de naissance, décès, immigration et émigration depuis ce temps. La moyenne des estimations mensuelles pour chacun des croisements RSS-âge-sexe par période de collecte a été retenue pour réaliser la calage. Après la calage, l'ajustement de poids I5 est obtenu. Le poids I5 correspond au ***poids personne final de l'ESCC*** que l'on retrouve dans le fichier de données portant le nom de variable WTS_M.

8.6 Particularités de la pondération pour les trois territoires

Tel que décrit à la section 5, le plan d'échantillonnage utilisé pour les trois territoires est quelque peu différent de celui utilisé dans les provinces. La stratégie de pondération est donc adaptée pour répondre à ces différences. Cette section résume les changements apportés à la stratégie expliquée aux sous-sections 8.1 à 8.5.

D'abord pour la base aréolaire, tel que mentionné à la sous-section 5.4.1, une étape additionnelle de sélection est ajoutée pour les territoires. Chaque territoire est 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 sont donc toutes trois sélectionnées automatiquement à cette première sélection. Cette particularité n'a

¹² Sautory O. CALMAR 2: A New Version of the CALMAR Calibration Adjustment Program. *Proceedings of Statistics Canada Symposium* (Statistics Canada, Catalogue no. 11-522-XCB), 2003.

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 A4) est appliquée aux unités de la base aréolaire. Les classes d'ajustement sont construites, pour les non-réponses ménage et personne, 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 est utilisée, et ce, uniquement à l'intérieur des capitales du Yukon et des Territoires du Nord-Ouest. Tous les ajustements de la base téléphonique sont appliqués afin de calculer un poids pour les unités de la base téléphonique.

Les deux ensembles de poids (aréolaire et téléphonique) sont ensuite intégrés, puis poststratifiés de façon semblable à ce qui est fait pour les provinces, à l'exception de trois 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 étant couvertes uniquement par la base aréolaire. Un autre détail à noter pour le Nunavut est que les comptes de la population utilisés pour la calibration représentent les 10 plus grandes communautés (70% de la population) seulement étant donné la sous couverture de la base aréolaire, telle que décrite à la section 5.4.1. Finalement, en commençant avec les produits de diffusion des périodes de référence de 2008 et 2007-2008, les contrôles ont été mis en place pour s'assurer que la proportion des autochtones et la proportion d'individus vivant dans les régions capitales soient contrôlées dans les Territoires du Nord-Ouest et du Yukon. Un contrôle similaire basé sur le statut Inuit a été introduit pour le Nunavut. Ces contrôles s'assurent que la proportion des estimations représentées par ces groupes est cohérente avec les proportions indiquées par le recensement de 2006.

9. QUALITÉ DES DONNÉES

9.1 Taux de réponse

Au total, 84 973 des unités sélectionnées à l'ESCC 2007 faisaient parti du champ de l'enquête¹³. Parmi ces unités, 71 922 ont accepté de participer à l'enquête ce qui résulte en un taux de réponse à l'échelle du ménage de 84,6%. Parmi ces ménages répondants, 71 922 personnes ont été sélectionnées (une personne par ménage) pour participer à l'enquête parmi lesquelles 65 946 ont accepté de le faire ce qui résulte en un taux de réponse à l'échelle de la personne de 91,7 %. À l'échelle canadienne, un taux de réponse combiné de 77,6 % a donc été observé pour l'ESCC 2007. 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. Le tableau 9.2 donne la même information mais au niveau des réseaux locaux d'intégration des services de santé (RLISS).

Tableau 9.1 : Taux de réponse par région socio-sanitaire et par base de sondage de l'ESCC 2007-2008

[\(voir Annexe E\)](#)

Tableau 9.2 : Taux de réponse par réseau local d'intégration des services de santé (RLISS) et par base de sondage de l'ESCC 2007-2008 en Ontario

[\(voir Annexe E\)](#)

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.

¹³ Parmi les unités sélectionnées au départ, certaines ne font pas parti du champ de l'enquête. Ce sont par exemples des logements vacants ou démolis, des logements non-résidentiels ou encore des numéros de téléphones non-valides tels les numéros sans service ou non-résidentiels. Ces unités sont identifiées pendant la collecte, autrement elles auraient été exclues lors de la sélection. Ces unités ne sont pas considérées dans le calcul des taux de réponse.

Taux de réponse à l'échelle du ménage

HHRR =	# de ménages répondants provenant des 2 bases
	tous les ménages faisant partie du champ de l'enquête provenant des 2 bases

Taux de réponse à l'échelle de la personne

PPRR =	# de répondants provenant des 2 bases
	toutes les personnes sélectionnées provenant des 2 bases

Taux de réponse combiné = HHRR x 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.

HHRR =	35 748 + 36 174	=	71 922	= 0,846
	40 668 + 44 310		84 978	

PPRR =	33 473 + 32 473	=	65 946	= 0,917
	35 748 + 36 174		71 922	

$$\begin{aligned}
 \text{Taux de réponse combiné} &= 0,846 \times 0,917 \\
 &= 0,776 \\
 &= \mathbf{77,6 \%}
 \end{aligned}$$

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 l'ESCC 2007, il y a peu de non-réponse partielle car une fois le questionnaire débuté les répondants avaient tendance à le terminer. Il y a eu non-réponse totale lorsque la personne sélectionnée pour participer à l'enquête a refusé de le faire ou que l'intervieweur a été incapable d'entrer en contact avec elle. Dans certaines régions socio-sanitaires, l'échantillon initial a été augmenté au cours de l'année afin d'atteindre les tailles ciblées pour 2007. Finalement, les cas de non-réponse totale ont été considérés lors de la pondération en corrigeant les poids des personnes qui ont répondu à l'enquête afin de compenser pour ceux qui n'ont pas répondu. Voir la section 8 pour avoir de plus amples détails sur la correction de la pondération pour la non-réponse.

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 de l'ESCC 2007 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 des fichiers de microdonnées 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 ces fichiers de microdonnées (Maitre, Partager, ou FMGD) correspondent à celles produites par Statistique Canada, il est vivement conseillé à l'utilisateur de les arrondir en se conformant aux lignes directrices suivantes.

- a) Les estimations qui figurent dans le corps d'un tableau statistique doivent être arrondies à la centaine près par la méthode d'arrondissement classique. Selon cette méthode, si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre retenu ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente d'une unité (1) la valeur du dernier chiffre retenu. Par exemple, si l'on veut arrondir à la centaine près de la façon classique une estimation dont les deux derniers chiffres sont compris entre 00 et 49, il faut les remplacer par 00 et ne pas modifier le chiffre précédent (le chiffre des centaines). Si les deux derniers chiffres sont compris entre 50 et 99, il faut les remplacer par 00 et augmenter d'une unité (1) le chiffre précédent.
- b) Les totaux partiels de marge et les totaux de marge des tableaux statistiques doivent être calculés à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la centaine près selon la méthode d'arrondissement 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 fichier de microdonnées ne pourront être considérées comme représentatives de la population observée et ne correspondront pas à celles de Statistique Canada.

L'utilisateur ne doit pas non plus perdre de vue qu'en raison du traitement réservé au champ du poids, certains 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 fichier de microdonnées.

Estimations de type nominal :

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

Exemple de question de type nominal :

Actuellement, est-ce que ... fume(z) des cigarettes tous les jours, à l'occasion ou jamais?
(SMK_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? (SMK_204)

□□ Nombre de cigarettes

10.2.2 Totalisation d'estimations de type nominal

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

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

10.2.3 Totalisation d'estimations quantitatives

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

- a) multiplier la valeur de la variable étudiée par le poids final, puis faire la somme de cette quantité pour tous les enregistrements visés pour obtenir le numérateur (\hat{X});
- b) 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 **SMK_204** et le poids **WTS_M**. Ensuite additionnez cette valeur pour les enregistrements pour lesquels la valeur de la variable **SMK_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 **SMK_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.

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) pour s'assurer qu'il y a assez d'unités pour calculer un estimation de qualité. Pour les utilisateurs de FMGD, 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 utilisateurs des fichiers maître ou de partage, on devrait avoir au moins 10 observations. Pour les estimations pondérées basées sur des échantillons de 10 ou plus (30 ou plus pour le FMGD), 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 30, 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).

$$\hat{X} = 0,777$$

$$z = 2$$

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

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

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

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 FICHIERS DE MICRODONNÉES : DESCRIPTION, ACCÈS ET UTILISATION

L'ESCC produit trois types de fichiers de microdonnées : les fichiers maîtres, les fichiers de partage et les fichiers de microdonnées à grande diffusion (FMGD).

12.1 Fichiers maîtres

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

12.1.1 Centre de données de recherche

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

12.1.2 Totalisations personnalisées

Une autre méthode d'accès aux fichiers maîtres consiste à offrir à tous les utilisateurs de faire appel au personnel du Service à la clientèle de la Division de la statistique de la santé pour produire des totalisations personnalisées. Ce service est offert moyennant le recouvrement des coûts. Il permet aux utilisateurs qui ne savent pas se servir de logiciels de totalisation d'obtenir des résultats personnalisés. Les résultats sont filtrés pour s'assurer qu'ils sont conformes aux normes de confidentialité et de fiabilité avant d'être diffusés. Pour plus de renseignements, communiquez avec le Service à la clientèle (613) 951-1653 ou par courriel à hd-ds@statcan.gc.ca.

12.1.3 Télé-accès

En dernier lieu, le service de télé-accès aux fichiers maîtres de l'enquête est un moyen d'accéder à ces données s'il est impossible de passer par un Centre de recherche en données. On peut fournir à l'acheteur d'un produit de microdonnées un fichier maître de données synthétique ou fichier « fictif » et le cliché d'enregistrement correspondant. Grâce à ces outils, le chercheur peut mettre au point son propre ensemble de programmes analytiques. Il ne lui reste plus qu'à envoyer le code pour les totalisations personnalisées par courrier électronique à cchs-escs@statcan.gc.ca. Le code est transmis au réseau interne protégé de Statistique Canada et traité en regard du fichier maître approprié de données 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 qu'ils sont conformes aux normes de confidentialité et de fiabilité, puis, les données de sortie sont renvoyées au client. Ce service est gratuit.

12.2 Fichiers de partage

Les fichiers de partage contiennent toutes les variables et tous les enregistrements des répondants de l'ESCC qui ont accepté de partager leurs données avec les partenaires de partage de Statistique Canada, soit les ministères de la santé des provinces et territoires, Santé Canada et l'Agence de santé de la fonction publique. Statistique Canada demande également aux répondants résidant au Québec leur permission de partager leurs données avec l'Institut de la statistique du Québec. Statistique Canada ne fournit le fichier de partage qu'à ces organisations. Les identificateurs personnels sont retirés des fichiers de partage pour préserver la confidentialité des répondants. Les utilisateurs de ces fichiers doivent au préalable avoir porté serment qu'ils ne divulgueront en aucun temps toute information susceptible d'identifier un répondant à l'enquête.

12.3 Fichiers de microdonnées à grande diffusion

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

Les variables les plus susceptibles de permettre l'identification d'une personne sont supprimées du fichier ou agrégées en catégories moins détaillées.

Les FMGD contiennent les données collectées au cours de deux années. Ils comprennent les questions qui ont été posées au cours des deux années. À moins d'exception, ces questions sont habituellement celles comprises dans la composante du contenu de base, du contenu thématique collecté sur deux années et le contenu optionnel choisi pour deux années.

Les FMGD sont accessibles gratuitement dans les établissements d'enseignement post-secondaires font partie de l'Initiative de démocratisation des données. Ils sont aussi disponibles gratuitement sur demande auprès du Service à la clientèle.

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

Depuis les diffusions des données de 2008 et de 2007-2008, les utilisateurs qui ont accès aux fichiers de partage ou aux fichiers maîtres ont le choix d'utiliser les fichiers d'un an ou de deux ans. Les décisions concernant la période à utiliser pour une analyse donnée devraient être guidées par le niveau de détail et de qualité requis. Dans le cas d'un fichier d'un an, les estimations ne sont pas toujours disponibles en raison de la qualité liée à la taille limitée de l'échantillon.

Dans le cadre de l'ESCC, il est recommandé d'utiliser un coefficient de variation de moins de 33 % et de compter au moins dix répondants qui affichent les caractéristiques dans le domaine, avant de publier une estimation. Cela ne sera pas possible dans le cas des caractéristiques rares et des

domaines détaillés pour les fichiers d'un an. Les utilisateurs devront plutôt se fier aux fichiers de deux ans ou aux fichiers pluriannuels.

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

12.5 Utilisation de la variable de pondération

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

Le module **Sécurité alimentaire** mesure des concepts qui s'appliquent non seulement à la situation du répondant, mais à celle de l'ensemble de son ménage. L'analyse des variables peut selon le niveau d'analyse nécessiter le recours à un poids calculé de manière à représenter le nombre de ménages au Canada, plutôt que le nombre de personnes. Cette variable de pondération **WTS_HH** se trouve dans un fichier distinct (HS_HHWT.txt). Elle peut être utilisée en remplacement de la variable **WTS_M** pour des analyses au niveau des ménages à l'échelle nationale et provinciale. La variable de pondération **WTS_M** doit être utilisée pour des analyses régionales.

12.6 Convention appliquée pour nommer les variables à partir de 2007

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

La convention appliquée pour nommer les variables de l'ESCC a été modifiée à partir de 2007. Ainsi, la lettre correspondant à l'édition de l'enquête (par exemple, A = cycle 1.1, C = cycle 2.1 et E = 3.1) n'est plus utilisée dans les noms de variables. Une nouvelle variable (REFPER, format = AAAAMM-AAAAMM) a été ajoutée aux fichiers de microdonnées afin d'identifier le début et la fin de la période de référence au cours de laquelle les données ont été recueillies. Celle-ci s'avérera utile en particulier pour les utilisateurs désireux d'utiliser les données de plusieurs périodes de collecte à la fois. Ainsi, les noms des variables correspondant à des modules ou à des questions identiques entre une année de collecte et une autre (par ex., 2007 et 2008) seront les mêmes.

La convention appliquée pour nommer les variables à compter de l'ESCC 2007 utilise jusqu'à huit caractères. Les noms de variables sont structurés de la manière suivante :

Positions 1 à 3 : Nom du module/de la section du questionnaire
Position 4 : Type de variable (sous-tiret, C, D, F ou G)
Positions 5 à 8 : Numéro de la question et option de réponse (s'il s'agit d'une question à réponse multiple)

L'exemple 1 présente la structure du nom de la variable correspondant à la question 202 du module Usage du tabac, c'est-à-dire SMK_202 :

Positions 1 à 3 : SMK Module sur l'usage du tabac
Position 4 : _ (sous-tiret = données recueillies)
Position 5 à 8 : 202 numéro de la question

L'exemple 2 présente la structure du nom de la variable correspondant à la question 2 du module Utilisation des soins de santé (HCU_02A) qui constitue une question à réponse multiple :

Positions 1 à 3 : HCU Module sur l'utilisation des soins de santé
Position 4 : _ (sous-tiret = données recueillies)
Position 5 à 8 : 02AA numéro de la question correspondant et option de réponse

Dans les positions 1 à 3, on retrouve l'acronyme de chacun des modules. Ces acronymes apparaissent à côté des noms de modules qui sont tous présentés dans la figure de l'annexe A.

La position 4 désigne le type de variable selon qu'il s'agit d'une variable collectée directement à partir d'une question du questionnaire (« _ »), d'une variable codée (« C »), dérivée (« D »), groupée (« G ») ou d'une variable indicatrice (« F »)¹⁴.

En général, les quatre dernières positions (5 à 8) correspondent à la numérotation de la variable qui figure sur le questionnaire. On supprime la lettre « Q » utilisée pour représenter le mot « question » et on présente tous les numéros de question au moyen d'un groupe de deux ou trois chiffres. Par exemple, la question Q01A du questionnaire devient simplement 01A et la question Q15, simplement 15.

¹⁴ Du mot anglais *flag*.

Tableau 12.1 Désignation des codes utilisés à la 4e position du nom des variables de l'ESCC

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

Parfois, certaines questions comportent plusieurs réponses alors la position finale dans la séquence du nom de la variable est représentée par une lettre. Pour ce genre de questions, de nouvelles variables sont créées dans le but de différencier un « oui » d'un « non » pour chaque possibilité de réponse. Par exemple, si la question 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.

12.7 Convention appliquée pour nommer les variables avant 2007

Tel que mentionné précédemment, la convention appliquée pour nommer les variables a été modifiée en 2007. On a enlevé l'indicateur du cycle au cours duquel les variables avaient été collectées. Cet indicateur se trouvait à la 4^e position pour les cycles 1.1 à 3.1.

Voici la liste des lettres utilisées dans les fichiers de microdonnées de l'ESCC entre les cycles 1.1 et 3.1 et leur cycle correspondant.

Lettre	Cycle et nom du cycle
A	Cycle 1.1 : Enquête sur la santé dans les collectivités canadiennes
B	Cycle 1.2 : Enquête sur la santé dans les collectivités canadiennes, santé mentale et bien-être
C	Cycle 2.1 : Enquête sur la santé dans les collectivités canadiennes
D	Cycle 2.2 : Enquête sur la santé dans les collectivités canadiennes, nutrition
E	Cycle 3.1 : Enquête sur la santé dans les collectivités canadiennes

Annexe A

Annexe A- Enquête sur la santé dans les collectivités canadiennes (2007-2008)

Contenu de base (toutes les régions)				
<ul style="list-style-type: none"> • Âge du répondant (ANC) • Consommation d'alcool (ALC) • Problèmes de santé chroniques (CCC) • Exposition à la fumée secondaire (ETS) • Vaccins contre la grippe (FLU) • État de santé général (GEN) • Utilisation des soins de santé (HCU) • Douleurs et malaise (HUP) 		<ul style="list-style-type: none"> • Taille et poids – autodéclarés (HWT) • Expériences maternelles – Allaitement (MEX) • Consommation de fruits et de légumes (FVC) • Activités physiques (PAC) • Limitation des activités (RAC) • Usage du tabac (SMK) 		<p>Administration et renseignements sociodémographiques</p> <ul style="list-style-type: none"> • Renseignements administratifs (ADM) • Renseignement sur le logement (DWL) • Éducation (EDU) • Revenu (INC) • Population active (LBS) <p>Renseignements sociodémographiques</p>
Contenu thématique (toutes les régions)				
<i>Thème pour 2007-2008: Mode de vie sain</i>		<i>Thème pour 2007: Accès aux services de soins de santé au Canadaⁱ</i>		<i>Thème pour 2008: Prévention des maladies chroniques</i>
<ul style="list-style-type: none"> • Changements faits pour améliorer la santé (CIH) • Sécurité alimentaire (FSC) • Santé bucco-dentaire 1 (OH1) 	<ul style="list-style-type: none"> • Activités physiques – Installations au travail (PAF) • Activités sédentaires (SAC) 	<ul style="list-style-type: none"> • Accès aux services de soins de santé (ACCZ) • Satisfaction des patients à l'égard des services de soins de santé (PASZ) • Temps d'attente (WTMZ) 	<ul style="list-style-type: none"> • Analyse de sang (BLT) • Test pap (PAP) • Mammographie (MAM) 	<ul style="list-style-type: none"> • Spirométrie(SPI) • Examen général (PCU) • Dépistage du cancer du côlon et du rectum (CCS)

Contenu optionnel (certaines régions)			
<ul style="list-style-type: none"> • Accès aux services de soins de santé (ACC)ⁱⁱ • Consommation d'alcool – Dépendance (ALD) • Consommation d'alcool – Anciens buveurs (ALN) • Consommation d'alcool au cours de la dernière semaine (ALW) • Tension artérielle – Vérification (BPC) • Examen des seins (BRX) • Auto-examen des seins (BSX) • Stress - Facteurs de stress durant l'enfance et la vie adulte (CST) • Consultations au sujet de la santé mentale (CMH) • Dépistage du cancer du côlon et du rectum (CCS) • Visites chez le dentiste (DEN) • Dépression (DEP) • Soins pour le diabète (DIA) • Usage de suppléments alimentaires - Vitamines et minéraux (DSU) • Détresse (DIS) • Conduite et sécurité (DRV) 	<ul style="list-style-type: none"> • Examens de la vue (EYX) • Choix alimentaires (FDC) • Indice de l'état de santé (HUI) • Satisfaction à l'égard du système de soins de santé (HCS) • État de santé (SF-36) (SFR) • Services de soins de santé à domicile (HMC) • Mesures de sécurité à la maison (HMS) • Usage de drogues illicites (DRG) • Blessures (INJ) • Couverture d'assurance (INS) • Mammographie (MAM) • Contrôle de soi (MAS) • Expériences maternelles – Consommation d'alcool au cours de la grossesse (MXA) • Expériences maternelles – Usage du tabac au cours de la grossesse (MXS) • Usage du tabac - Dépendance à la nicotine (NDE) 	<ul style="list-style-type: none"> • Santé bucco-dentaire 2 (OH2) • Satisfaction des patients à l'égard des services de soins de santé (PAS)ⁱⁱ • Satisfaction des patients à l'égard des soins de santé communautaires (PSC) • Jeu excessif (CPG) • Dépistage du cancer de la prostate (PSA) • Bien-être psychologique (PWB) • Stress - Événements récents (RLE) • Satisfaction à l'égard de la vie (SWL) • Estime de soi (SFE) • Comportements sexuels (SXB) • Sommeil (SLP) • Test pap (PAP) • Usage du tabac - Méthodes pour cesser de fumer (SCA) • Usage du tabac – Consultation d'un médecin (SPC) 	<ul style="list-style-type: none"> • Usage du tabac – Les étapes du changement (SCH) • Soutien social – Disponibilité (SSA) • Soutien social – Utilisation (SSU) • Valeurs spirituelles (SPR) • Stress - Faire face au stress (STC) • Stress – Sources (STS) • Pensées suicidaires et tentatives de suicide (SUI) • Protection contre le soleil (SSB) • Usage du tabac - Autres produits du tabac (TAL) • Utilisation de l'équipement protecteur (UPE) • Organismes à but non lucratif – Participation (ORG) • Stress au travail (WST) • Usage du tabac chez les jeunes (YSM)

i Demandé à un sous-échantillon de répondants. Le thème de 2007 (Accès aux services de soins de santé au Canada) n'est pas demandé aux répondants des territoires.

ii Ces modules de contenu thématique de 2007 ont aussi été sélectionnés comme contenu optionnel par certaines régions

Annexe B

Annexe B - Sélection du contenu optionnel selon la province ou le territoire (2007-2008)

Modules optionnels	T.-N.-L.	Î.-P.-É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sask.	Alb.	C.-B.	Yn	T.N.-O.	Nt
Accès aux services de soins de santé				*									
Consommation d'alcool - Dépendance								*		*			
Consommation d'alcool – Anciens buveurs	*									*			
Consommation d'alcool au cours de la dernière semaine	*		*			*				*			*
Tension artérielle - Vérification									*			*	
Examen des seins				*								*	
Auto-examen des seins	*			*									
Stress - Facteurs de stress durant l'enfance et la vie adulte												*	
Dépistage du cancer du côlon et du rectum	*	*				*		*				*	*
Consultations au sujet de la santé mentale		*	*	*		*			*			*	*
Visites chez le dentiste	*								*			*	
Dépression			*	*	*				*		*		*
Soins pour le diabète	*	*	*	*		*				*	*	*	
Usage de suppléments alimentaires - Vitamines et minéraux											*	*	
Détresse			*		*								
Conduite et sécurité							*		*			*	
Examens de la vue						*							
Choix alimentaires		*					*		*	*		*	
Satisfaction à l'égard du système de soins de santé						*					*	*	
Indice de l'état de santé					*								
Services de soins de santé à domicile						*							
Mesures de sécurité à la maison												*	
Usage de drogues illicites			*							*			
Blessures			*							*			
Couverture d'assurance				*									

Modules optionnels	T.-N.-L.	Î.-P.-É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sask.	Alb.	C.-B.	Yn	T.N.-O.	Nt
Mammographie	*		*	*		*		*				*	
Contrôle de soi							*						
Expériences maternelles – Consommation d'alcool au cours de la grossesse						*				*		*	
Expériences maternelles – Usage du tabac au cours de la grossesse						*				*		*	*
Usage du tabac - Dépendance à la nicotine									*			*	*
Santé bucco-dentaire 2	*												
Test pap				*		*		*				*	
Satisfaction des patients à l'égard des services de soins de santé											*	*	
Satisfaction des patients à l'égard des soins de santé communautaires				*				*			*	*	
Jeu excessif					*	*		*					*
Dépistage du cancer de la prostate	*	*										*	
Bien-être psychologique				*									
Stress - Événements récents											*		
Satisfaction à l'égard de la vie					*				*				*
Estime de soi													*
Comportements sexuels			*	*		*		*				*	*
État de santé (SF-36)	*	*					*						
Sommeil			*								*		
Usage du tabac – Consultation d'un médecin									*				
Usage du tabac – Les étapes du changement						*							
Usage du tabac - Méthodes pour cesser de fumer						*							
Soutien social - Disponibilité			*		*					*	*		*
Soutien social - Utilisation										*	*		*
Valeurs spirituelles								*					*
Stress - Faire face au stress											*		*
Stress - Sources											*		*
Pensées suicidaires et tentatives de suicide						*			*	*		*	

Modules optionnels	T.-N.-L.	Î.-P.-É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sask.	Alb.	C.-B.	Yn	T.N.-O.	Nt
Protection contre le soleil		*											
Usage du tabac - Autres produits du tabac						*			*				
Utilisation de l'équipement protecteur			*					*	*				
Organismes à but non lucratif - Participation			*									*	
Stress au travail									*				
Usage du tabac chez les jeunes										*			

Annexe C

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

0	Canada	
10	Terre-Neuve-et-Labrador	
1011-C		Eastern Regional Integrated Health Authority
1012-I		Central Regional Integrated Health Authority
1013-I		Western Regional Integrated Health Authority
1014-H		Labrador-Grenfell Regional Integrated Health Authority
11	Île-du-Prince-Édouard	
1101-D		Kings County
1102-A		Queens County
1103-C		Prince County
12	Nouvelle-Écosse	
1201-C		Zone 1
1202-C		Zone 2
1203-C		Zone 3
1204-C		Zone 4
1205-I		Zone 5
1206-A		Zone 6
13	Nouveau-Brunswick	
1301-C		Région 1
1302-C		Région 2
1303-C		Région 3
1304-C		Région 4
1305-I		Région 5
1306-I		Région 6
1307-I		Région 7
24	Québec	
2401-C		Région du Bas-Saint-Laurent
2402-C		Région du Saguenay - Lac-Saint-Jean
2403-A		Région de la Capitale-Nationale
2404-C		Région de la Mauricie et du Centre-du-Québec
2405-C		Région de l'Estrie
2406-G		Région de Montréal
2407-A		Région de l'Outaouais
2408-C		Région de l'Abitibi-Témiscamingue
2409-H		Région de la Côte-Nord
2410-H		Région du Nord-du-Québec
2411-I		Région de la Gaspésie - Îles-de-la-Madeleine
2412-E		Région de la Chaudière-Appalaches
2413-A		Région de Laval
2414-E		Région de Lanaudière
2415-E		Région des Laurentides
2416-A		Région de la Montérégie
35	Ontario	Réseau local d'intégration des services de santé

3501	Réseau d'intégration des services de santé de Érié St. Clair
3502	Réseau d'intégration des services de santé du Sud-Ouest
3503	Réseau d'intégration des services de santé de Waterloo Wellington
3504	Réseau d'intégration des services de santé de Hamilton Niagara Haldimand Brant
3505	Réseau d'intégration des services de santé du Centre-Ouest
3506	Réseau d'intégration des services de santé de Mississauga Halton
3507	Réseau d'intégration des services de santé de Toronto-Centre
3508	Réseau d'intégration des services de santé du Centre
3509	Réseau d'intégration des services de santé du Centre-Est
3510	Réseau d'intégration des services de santé du Sud-Est
3511	Réseau d'intégration des services de santé de Champlain
3512	Réseau d'intégration des services de santé de Simcoe-Nord Muskoka
3513	Réseau d'intégration des services de santé du Nord-Est
3514	Réseau d'intégration des services de santé du Nord-Ouest
35	Ontario par circonscription sanitaire
3526-C	Circonscription sanitaire du district d'Algoma
3527-A	Circonscription sanitaire du comté de Brant
3530-B	Circonscription sanitaire régionale de Durham
3531-E	Circonscription sanitaire d'Elgin-St. Thomas
3533-E	Circonscription sanitaire de Grey Bruce
3534-E	Circonscription sanitaire de Haldimand-Norfolk
3535-E	Circonscription sanitaire du district de Haliburton, Kawartha et Pine Ridge
3536-B	Circonscription sanitaire régionale de Halton
3537-A	Circonscription sanitaire de la cité de Hamilton
3538-A	Circonscription sanitaire des comtés de Hastings et Prince Edward
3539-E	Circonscription sanitaire du comté de Huron
3540-A	Circonscription sanitaire de Chatham-Kent
3541-A	Circonscription sanitaire de Kingston, Frontenac et Lennox et Addington
3542-A	Circonscription sanitaire de Lambton
3543-E	Circonscription sanitaire de Leeds, Grenville et Lanark
3544-A	Circonscription sanitaire de Middlesex-London
3546-A	Circonscription sanitaire régionale de Niagara
3547-C	Circonscription sanitaire du district de North Bay Parry Sound
3549-H	Circonscription sanitaire du Nord-Ouest
3551-B	Circonscription sanitaire de la cité d'Ottawa
3552-E	Circonscription sanitaire du comté d'Oxford
3553-B	Circonscription sanitaire régionale de Peel
3554-E	Circonscription sanitaire du district de Perth
3555-A	Circonscription sanitaire du comté et de la cité de Peterborough
3556-H	Circonscription sanitaire de Porcupine
3557-E	Circonscription sanitaire du comté et du district de Renfrew
3558-E	Circonscription sanitaire de l'Est de l'Ontario
3560-E	Circonscription sanitaire du district de Simcoe Muskoka
3561-C	Circonscription sanitaire de Sudbury et son district
3562-C	Circonscription sanitaire du district de Thunder Bay
3563-C	Circonscription sanitaire de Timiskaming
3565-B	Circonscription sanitaire de Waterloo
3566-B	Circonscription sanitaire de Wellington-Dufferin-Guelph
3568-B	Circonscription sanitaire de Windsor-Comté d'Essex
3570-B	Circonscription sanitaire de York

3595-G 46	Manitoba	Circonscription sanitaire de la cité de Toronto
4610-A		Winnipeg Regional Health Authority
4615-A		Brandon Regional Health Authority
4620-E		North Eastman Regional Health Authority
4625-E		South Eastman Regional Health Authority
4630-E		Interlake Regional Health Authority
4640-D		Central Regional Health Authority
4645-D		Assiniboine Regional Health Authority
4660-D		Parkland Regional Health Authority
4670-H		Norman Regional Health Authority
4685-F 47	Saskatchewan	Burntwood/Churchill
4701-D		Sun Country Regional Health Authority
4702-D		Five Hills Regional Health Authority
4703-D		Cypress Regional Health Authority
4704-A		Regina Qu'Appelle Regional Health Authority
4705-D		Sunrise Regional Health Authority
4706-A		Saskatoon Regional Health Authority
4707-D		Heartland Regional Health Authority
4708-D		Kelsey Trail Regional Health Authority
4709-C		Prince Albert Parkland Regional Health Authority
4710-H		Prairie North Regional Health Authority
4714-F 48	Alberta	Mamawetan/Keewatin/Athabasca
4821-E		Chinook Regional Health Authority
4822-E		Palliser Health Region
4823-B		Calgary Health Region
4824-E		David Thompson Regional Health Authority
4825-E		East Central Health
4826-E		Capital Health
4827-E		Aspen Regional Health Authority
4828-E		Peace Country Health
4829-H 59	Colombie- Britannique	Northern Lights Health Region
5911-E		East Kootenay Health Service Delivery Area
5912-C		Kootenay-Boundary Health Service Delivery Area
5913-A		Okanagan Health Service Delivery Area
5914-C		Thompson/Cariboo Health Service Delivery Area
5921-A		Fraser East Health Service Delivery Area
5922-B		Fraser North Health Service Delivery Area
5923-B		Fraser South Health Service Delivery Area
5931-B		Richmond Health Service Delivery Area
5932-G		Vancouver Health Service Delivery Area
5933-B		North Shore/Coast Garibaldi Health Service Delivery Area
5941-A		South Vancouver Island Health Service Delivery Area
5942-A		Central Vancouver Island Health Service Delivery Area
5943-C		North Vancouver Island Health Service Delivery Area
5951-H		Northwest Health Service Delivery Area

5952-H		Northern Interior Health Service Delivery Area
5953-H		Northeast Health Service Delivery Area
60	Yukon	
6001-H		Yukon
	Territoires du	
61	Nord-Ouest	
6101-H		Territoires du Nord-Ouest
62	Nunavut – 10 communautés les plus grandes	
6201-F		Nunavut – 10 communautés les plus grandes
A	Groupe de régions homologues A	
B	Groupe de régions homologues B	
C	Groupe de régions homologues C	
D	Groupe de régions homologues D	
E	Groupe de régions homologues E	
F	Groupe de régions homologues F	
G	Groupe de régions homologues G	
H	Groupe de régions homologues H	
I	Groupe de régions homologues I	

Annexe D

Annexe D - Répartition de l'échantillon par région socio-sanitaire et par base de sondage de l'ESCC 2007-2008

Prov./Terr.	Région socio-sanitaire	Base aréolaire		Bases téléphoniques		/Combiné	
		# répondants espérés	taille d'échantillon brute	# répondants espérés	taille d'échantillon brute	# répondants espérés	taille d'échantillon brute
CA	Total	66507	96086	65567	110846	132074	206932
NL	Total	2006	2838	2006	2975	4012	5813
	1011	810	1136	810	1190	1620	2326
	1012	470	682	470	702	940	1384
	1013	426	546	426	632	852	1178
	1014	300	474	300	451	600	925
PE	Total	1241	1894	1239	2108	2480	4001
	1101	177	312	177	321	354	633
	1102	640	980	648	1088	1288	2068
	1103	424	602	414	699	838	1300
NS	Total	2522	3826	2522	3707	5044	7533
	1201	396	620	396	595	792	1215
	1202	320	457	320	459	640	916
	1203	360	488	360	544	720	1032
	1204	350	634	350	546	700	1180
	1205	420	606	420	611	840	1217
	1206	676	1021	676	952	1352	1973
NB	Total	2578	3784	2578	4433	5156	8217
	1301	500	744	500	873	1000	1617
	1302	486	765	486	862	972	1628
	1303	470	711	470	820	940	1531
	1304	270	394	270	486	540	879
	1305	250	347	250	435	500	782
	1306	346	489	346	513	692	1002
	1307	256	334	256	444	512	778
QC	Total	11742	16659	12542	22259	24284	38918
	2401	600	792	600	934	1200	1726
	2402	628	826	628	931	1256	1757
	2403	927	1261	928	1494	1855	2755
	2404	802	1047	802	1227	1604	2274
	2405	618	951	618	980	1236	1931
	2406	1552	2241	1552	2787	3104	5028
	2407	646	996	646	1047	1292	2043
	2408	600	876	600	883	1200	1759
	2409	600	848	600	1030	1200	1878
	2410	0	0	800	3214	800	3214
	2411	600	839	600	954	1200	1793
	2412	722	973	722	1150	1444	2123
	2413	670	950	670	1099	1340	2049
	2414	718	958	718	1155	1436	2113
	2415	760	1360	758	1389	1518	2749
	2416	1300	1740	1300	1985	2600	3725

Prov./Terr.	Région socio-sanitaire	Base aréolaire		Bases téléphoniques		/Combiné	
		# répondants espérés	taille d'échantillon brute	# répondants espérés	taille d'échantillon brute	# répondants espérés	taille d'échantillon brute
ON	Total	22230	31901	22230	38160	44460	70061
	3526	426	688	426	696	852	1384
	3527	406	576	406	640	812	1216
	3530	816	1182	816	1289	1632	2471
	3531	340	458	340	563	680	1021
	3533	474	705	474	852	948	1557
	3534	386	579	386	680	772	1259
	3535	476	782	476	901	952	1683
	3536	706	921	706	1118	1412	2040
	3537	826	1203	826	1354	1652	2557
	3538	470	669	470	811	940	1480
	3539	296	400	296	505	592	905
	3540	400	487	400	659	800	1146
	3541	506	781	506	852	1012	1632
	3542	436	628	436	746	872	1374
	3543	476	699	476	778	952	1478
	3544	750	1163	750	1256	1500	2420
	3546	766	1009	766	1268	1532	2277
	3547	400	601	400	732	800	1333
	3549	328	535	328	722	656	1257
	3551	1026	1488	1026	1597	2052	3084
	3552	376	516	376	542	752	1058
	3553	1314	1817	1314	2345	2628	4162
	3554	326	445	326	490	652	935
	3555	426	652	426	772	852	1424
	3556	376	558	376	602	752	1159
	3557	376	547	376	659	752	1206
	3558	520	711	520	799	1040	1510
	3560	1142	1748	1142	2065	2284	3813
	3561	540	755	540	924	1080	1679
	3562	717	1006	717	1246	1434	2252
	3563	250	373	250	410	500	782
	3565	766	1049	766	1150	1532	2199
	3566	564	746	564	886	1128	1632
	3568	716	976	716	1193	1432	2169
	3570	942	1213	942	1690	1884	2903
	3595	2169	3233	2169	4371	4338	7604
MB	Total	3754	5188	3754	5658	7508	10846
	4610	1056	1413	1056	1494	2112	2907
	4615	280	391	280	434	560	825
	4620	250	364	250	480	500	844
	4625	300	409	300	442	600	851
	4630	346	566	346	526	692	1092
	4640	400	503	400	544	800	1047
	4645	356	490	356	504	712	994

Prov./Terr.	Région socio-sanitaire	Base aréolaire		Bases téléphoniques		/Combiné	
		# répondants espérés	taille d'échantillon brute	# répondants espérés	taille d'échantillon brute	# répondants espérés	taille d'échantillon brute
	4660	266	377	266	386	532	763
	4670	250	355	250	448	500	803
	4685	250	319	250	400	500	719
SK	Total	3612	5161	4112	7780	7724	12941
	4701	300	387	300	440	600	827
	4702	300	423	300	456	600	879
	4703	266	388	266	396	532	784
	4704	620	871	620	872	1240	1743
	4705	310	446	310	458	620	904
	4706	660	889	660	948	1320	1837
	4707	270	456	270	376	540	832
	4708	260	389	260	360	520	749
	4709	326	522	326	518	652	1040
	4710	300	391	300	454	600	845
	4714	0	0	500	2502	500	2502
AB	Total	6104	8750	6104	9284	12208	18034
	4820	510	699	510	762	1020	1461
	4821	416	509	416	600	832	1109
	4822	1396	2048	1396	2072	2792	4120
	4823	700	1024	700	1058	1400	2082
	4824	446	595	446	666	892	1261
	4825	1310	1901	1310	1926	2620	3827
	4826	540	830	540	890	1080	1720
	4827	466	669	466	752	932	1421
	4828	320	475	320	558	640	1033
BC	Total	8048	11864	8050	13250	16098	25114
	5911	304	422	304	490	608	912
	5912	310	468	310	488	620	956
	5913	590	742	590	958	1180	1700
	5914	500	761	500	816	1000	1577
	5921	520	701	520	824	1040	1525
	5922	760	1018	760	1280	1520	2298
	5923	798	1161	800	1314	1598	2475
	5931	426	595	426	734	852	1329
	5932	800	1292	800	1500	1600	2792
	5933	546	885	546	894	1092	1779
	5941	676	955	676	1068	1352	2023
	5942	526	766	526	798	1052	1564
	5943	264	363	264	400	528	763
	5951	326	587	326	586	652	1173
	5952	426	663	426	646	852	1309
	5903	276	484	276	454	552	938
Terr.	60	950	1459	250	679	1200	2138
	61	1020	1631	180	553	1200	2184
	62	700	1132	0	0	700	1132

**Répartition de l'échantillon par réseau local d'intégration des services de santé
et par base de sondage de l'ESCC 2007-2008 en Ontario**

		Base aréolaire		Bases téléphoniques		Combiné	
Prov./Terr.	RLISS	# répondants espérés	taille d'échantillon brute	# répondants espérés	taille d'échantillon brute	# répondants espérés	taille d'échantillon brute
ON	Total	22230	31900	22230	38161	44460	70061
	3501	1552	2092	1552	2598	3104	4689
	3502	2568	3696	2568	4217	5136	7913
	3503	1230	1659	1230	1886	2460	3545
	3504	2624	3671	2624	4308	5248	7979
	3505	1062	1502	1062	1934	2124	3436
	3506	1136	1534	1136	1945	2272	3480
	3507	1085	1641	1085	2117	2170	3758
	3508	1404	1893	1404	2676	2808	4569
	3509	2076	3136	2076	3694	4152	6829
	3510	1344	2003	1344	2270	2688	4273
	3511	2058	2939	2058	3280	4116	6219
	3512	1058	1626	1058	1914	2116	3540
	3513	1992	2975	1992	3363	3984	6338
	3514	1041	1534	1041	1959	2082	3493

Annexe E

Annexe E - Taux de réponse par région socio-sanitaire et par base de sondage de l'ESCC 2007-2008

		Base aréolaire							Bases téléphoniques							
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
Canada	Total	81837	70879	86.6	70879	66114	93.3	80.8	90872	73623	81.0	73623	65845	89.4	72.5	76.4
NL	Total	2387	2181	91.4	2181	2022	92.7	84.7	2624	2307	87.9	2307	2076	90.0	79.1	81.8
	1011	972	888	91.4	888	827	93.1	85.1	1032	903	87.5	903	807	89.4	78.2	81.5
	1012	540	500	92.6	500	463	92.6	85.7	645	568	88.1	568	512	90.1	79.4	82.3
	1013	455	431	94.7	431	413	95.8	90.8	563	497	88.3	497	450	90.5	79.9	84.8
	1014	420	362	86.2	362	319	88.1	76.0	384	339	88.3	339	307	90.6	79.9	77.9
PE	Total	1454	1264	86.9	1264	1184	93.7	81.4	1571	1314	83.6	1314	1183	90.0	75.3	78.2
	1101	159	134	84.3	134	126	94.0	79.2	125	103	82.4	103	91	88.3	72.8	76.4
	1102	880	749	85.1	749	700	93.5	79.5	957	799	83.5	799	727	91.0	76.0	77.7
	1103	415	381	91.8	381	358	94.0	86.3	489	412	84.3	412	365	88.6	74.6	80.0
NS	Total	3113	2752	88.4	2752	2540	92.3	81.6	3289	2843	86.4	2843	2612	91.9	79.4	80.5
	1201	460	438	95.2	438	415	94.7	90.2	522	452	86.6	452	419	92.7	80.3	84.9
	1202	390	341	87.4	341	316	92.7	81.0	384	336	87.5	336	314	93.5	81.8	81.4
	1203	429	381	88.8	381	359	94.2	83.7	445	374	84.0	374	347	92.8	78.0	80.8
	1204	438	382	87.2	382	341	89.3	77.9	471	409	86.8	409	375	91.7	79.6	78.8
	1205	530	490	92.5	490	459	93.7	86.6	545	467	85.7	467	425	91.0	78.0	82.2
	1206	866	720	83.1	720	650	90.3	75.1	922	805	87.3	805	732	90.9	79.4	77.3
NB	Total	3060	2707	88.5	2707	2549	94.2	83.3	3737	3221	86.2	3221	2960	91.9	79.2	81.1
	1301	623	539	86.5	539	495	91.8	79.5	736	622	84.5	622	582	93.6	79.1	79.2
	1302	575	492	85.6	492	470	95.5	81.7	722	621	86.0	621	568	91.5	78.7	80.0
	1303	536	480	89.6	480	449	93.5	83.8	664	577	86.9	577	523	90.6	78.8	81.0
	1304	343	297	86.6	297	283	95.3	82.5	395	348	88.1	348	320	92.0	81.0	81.7
	1305	282	251	89.0	251	239	95.2	84.8	384	329	85.7	329	299	90.9	77.9	80.8

		Base aérotaire							Bases téléphoniques							
	1306	423	399	94.3	399	376	94.2	88.9	461	400	86.8	400	368	92.0	79.8	84.2
	1307	278	249	89.6	249	237	95.2	85.3	375	324	86.4	324	300	92.6	80.0	82.2
QC	Total	14508	12663	87.3	12663	11958	94.4	82.4	17405	13965	80.2	13965	12485	89.4	71.7	76.6
	2401	681	646	94.9	646	627	97.1	92.1	775	664	85.7	664	606	91.3	78.2	84.7
	2402	721	638	88.5	638	595	93.3	82.5	785	666	84.8	666	597	89.6	76.1	79.2
	2403	1207	1031	85.4	1031	963	93.4	79.8	1299	1039	80.0	1039	930	89.5	71.6	75.5
	2404	934	831	89.0	831	792	95.3	84.8	1130	949	84.0	949	862	90.8	76.3	80.1
	2405	794	657	82.7	657	619	94.2	78.0	803	666	82.9	666	609	91.4	75.8	76.9
	2406	1952	1636	83.8	1636	1526	93.3	78.2	2475	1797	72.6	1797	1549	86.2	62.6	69.5
	2407	855	703	82.2	703	659	93.7	77.1	872	693	79.5	693	626	90.3	71.8	74.4
	2408	769	721	93.8	721	687	95.3	89.3	739	644	87.1	644	595	92.4	80.5	85.0
	2409	724	647	89.4	647	618	95.5	85.4	792	649	81.9	649	580	89.4	73.2	79.0
	2410	1224	1010	82.5	1010	898	88.9	73.4	73.4
	2411	675	625	92.6	625	604	96.6	89.5	779	649	83.3	649	569	87.7	73.0	80.7
	2412	824	764	92.7	764	734	96.1	89.1	922	749	81.2	749	680	90.8	73.8	81.0
	2413	885	739	83.5	739	680	92.0	76.8	1044	790	75.7	790	678	85.8	64.9	70.4
	2414	890	764	85.8	764	719	94.1	80.8	977	774	79.2	774	702	90.7	71.9	76.1
	2415	971	863	88.9	863	803	93.0	82.7	1018	810	79.6	810	725	89.5	71.2	76.8
	2416	1626	1398	86.0	1398	1332	95.3	81.9	1771	1416	80.0	1416	1279	90.3	72.2	76.9
ON	Total	27904	23784	85.2	23784	21914	92.1	78.5	31813	25065	78.8	25065	22044	87.9	69.3	73.6
	3526	609	512	84.1	512	478	93.4	78.5	564	477	84.6	477	417	87.4	73.9	76.3
	3527	534	467	87.5	467	406	86.9	76.0	536	435	81.2	435	391	89.9	72.9	74.5
	3530	1049	879	83.8	879	795	90.4	75.8	1158	928	80.1	928	810	87.3	69.9	72.7
	3531	398	339	85.2	339	313	92.3	78.6	447	356	79.6	356	325	91.3	72.7	75.5
	3533	534	476	89.1	476	459	96.4	86.0	636	515	81.0	515	473	91.8	74.4	79.7
	3534	487	425	87.3	425	376	88.5	77.2	554	432	78.0	432	387	89.6	69.9	73.3
	3535	649	555	85.5	555	525	94.6	80.9	643	540	84.0	540	489	90.6	76.0	78.5
	3536	859	736	85.7	736	668	90.8	77.8	984	792	80.5	792	684	86.4	69.5	73.4
	3537	1055	862	81.7	862	776	90.0	73.6	1153	876	76.0	876	764	87.2	66.3	69.7

		Base aéroloaire							Bases téléphoniques							
	3538	585	505	86.3	505	459	90.9	78.5	639	531	83.1	531	465	87.6	72.8	75.5
	3539	392	366	93.4	366	356	97.3	90.8	437	351	80.3	351	310	88.3	70.9	80.3
	3540	434	414	95.4	414	392	94.7	90.3	569	446	78.4	446	394	88.3	69.2	78.4
	3541	655	537	82.0	537	488	90.9	74.5	649	528	81.4	528	477	90.3	73.5	74.0
	3542	484	433	89.5	433	402	92.8	83.1	573	445	77.7	445	389	87.4	67.9	74.8
	3543	562	502	89.3	502	470	93.6	83.6	639	541	84.7	541	494	91.3	77.3	80.3
	3544	1032	796	77.1	796	742	93.2	71.9	1100	886	80.5	886	775	87.5	70.5	71.2
	3546	866	769	88.8	769	718	93.4	82.9	1080	857	79.4	857	758	88.4	70.2	75.8
	3547	499	431	86.4	431	388	90.0	77.8	563	465	82.6	465	419	90.1	74.4	76.0
	3549	397	347	87.4	347	314	90.5	79.1	484	388	80.2	388	339	87.4	70.0	74.1
	3551	1366	1086	79.5	1086	978	90.1	71.6	1387	1123	81.0	1123	989	88.1	71.3	71.4
	3552	473	427	90.3	427	395	92.5	83.5	471	400	84.9	400	353	88.3	74.9	79.2
	3553	1714	1458	85.1	1458	1339	91.8	78.1	2024	1530	75.6	1530	1281	83.7	63.3	70.1
	3554	416	375	90.1	375	348	92.8	83.7	439	366	83.4	366	332	90.7	75.6	79.5
	3555	533	446	83.7	446	411	92.2	77.1	552	451	81.7	451	417	92.5	75.5	76.3
	3556	502	443	88.2	443	380	85.8	75.7	494	406	82.2	406	379	93.3	76.7	76.2
	3557	427	398	93.2	398	362	91.0	84.8	508	409	80.5	409	376	91.9	74.0	78.9
	3558	656	568	86.6	568	537	94.5	81.9	688	576	83.7	576	498	86.5	72.4	77.0
	3560	1507	1276	84.7	1276	1190	93.3	79.0	1641	1353	82.4	1353	1206	89.1	73.5	76.1
	3561	621	533	85.8	533	494	92.7	79.5	743	606	81.6	606	557	91.9	75.0	77.1
	3562	830	704	84.8	704	656	93.2	79.0	989	809	81.8	809	726	89.7	73.4	76.0
	3563	248	224	90.3	224	213	95.1	85.9	336	268	79.8	268	231	86.2	68.8	76.0
	3565	930	778	83.7	778	705	90.6	75.8	1008	811	80.5	811	731	90.1	72.5	74.1
	3566	684	627	91.7	627	584	93.1	85.4	773	637	82.4	637	581	91.2	75.2	80.0
	3568	842	718	85.3	718	679	94.6	80.6	1047	822	78.5	822	703	85.5	67.1	73.2
	3570	1128	970	86.0	970	897	92.5	79.5	1508	1108	73.5	1108	930	83.9	61.7	69.3
	3595	2947	2402	81.5	2402	2221	92.5	75.4	3797	2601	68.5	2601	2194	84.4	57.8	65.5
MB	Total	4452	4000	89.8	4000	3789	94.7	85.1	4674	4042	86.5	4042	3731	92.3	79.8	82.4
	4610	1322	1132	85.6	1132	1073	94.8	81.2	1342	1129	84.1	1129	1039	92.0	77.4	79.3

		Base aérolaire							Bases téléphoniques							
	4615	342	305	89.2	305	281	92.1	82.2	365	312	85.5	312	284	91.0	77.8	79.9
	4620	310	282	91.0	282	271	96.1	87.4	352	312	88.6	312	299	95.8	84.9	86.1
	4625	353	329	93.2	329	314	95.4	89.0	372	336	90.3	336	299	89.0	80.4	84.6
	4630	414	375	90.6	375	351	93.6	84.8	400	347	86.8	347	318	91.6	79.5	82.2
	4640	452	417	92.3	417	394	94.5	87.2	491	434	88.4	434	405	93.3	82.5	84.7
	4645	410	379	92.4	379	362	95.5	88.3	416	371	89.2	371	350	94.3	84.1	86.2
	4660	288	268	93.1	268	264	98.5	91.7	325	277	85.2	277	255	92.1	78.5	84.7
	4670	275	259	94.2	259	240	92.7	87.3	325	278	85.5	278	264	95.0	81.2	84.0
	4685	286	254	88.8	254	239	94.1	83.6	286	246	86.0	246	218	88.6	76.2	79.9
SK	Total	4220	3850	91.2	3850	3653	94.9	86.6	5383	4575	85.0	4575	4166	91.1	77.4	81.4
	4701	341	325	95.3	325	307	94.5	90.0	349	307	88.0	307	278	90.6	79.7	84.8
	4702	339	313	92.3	313	297	94.9	87.6	402	333	82.8	333	303	91.0	75.4	81.0
	4703	308	280	90.9	280	264	94.3	85.7	362	284	78.5	284	264	93.0	72.9	78.8
	4704	786	703	89.4	703	663	94.3	84.4	876	736	84.0	736	667	90.6	76.1	80.0
	4705	305	284	93.1	284	273	96.1	89.5	351	291	82.9	291	276	94.8	78.6	83.7
	4706	822	742	90.3	742	701	94.5	85.3	946	797	84.2	797	723	90.7	76.4	80.5
	4707	259	239	92.3	239	230	96.2	88.8	335	292	87.2	292	262	89.7	78.2	82.8
	4708	306	283	92.5	283	272	96.1	88.9	259	216	83.4	216	199	92.1	76.8	83.4
	4709	405	357	88.1	357	339	95.0	83.7	511	439	85.9	439	403	91.8	78.9	81.0
	4710	349	324	92.8	324	307	94.8	88.0	344	306	89.0	306	279	91.2	81.1	84.6
	4714	648	574	88.6	574	512	89.2	79.0	79.0
AB	Total	7448	6323	84.9	6323	5836	92.3	78.4	8372	6825	81.5	6825	6089	89.2	72.7	75.4
	4821	608	556	91.4	556	520	93.5	85.5	664	542	81.6	542	495	91.3	74.5	79.8
	4822	420	381	90.7	381	359	94.2	85.5	529	450	85.1	450	408	90.7	77.1	80.8
	4823	1818	1503	82.7	1503	1369	91.1	75.3	1856	1528	82.3	1528	1366	89.4	73.6	74.4
	4824	895	755	84.4	755	688	91.1	76.9	951	807	84.9	807	709	87.9	74.6	75.7
	4825	521	474	91.0	474	445	93.9	85.4	604	504	83.4	504	453	89.9	75.0	79.8
	4826	1645	1333	81.0	1333	1240	93.0	75.4	1965	1549	78.8	1549	1374	88.7	69.9	72.4
	4827	613	548	89.4	548	512	93.4	83.5	711	573	80.6	573	499	87.1	70.2	76.4

		Base aréolaire							Bases téléphoniques							
	4828	569	494	86.8	494	449	90.9	78.9	665	531	79.8	531	479	90.2	72.0	75.2
	4829	359	279	77.7	279	254	91.0	70.8	427	341	79.9	341	306	89.7	71.7	71.2
BC	Total	9972	8377	84.0	8377	7848	93.7	78.7	11450	8977	78.4	8977	8055	89.7	70.3	74.2
	5911	353	323	91.5	323	307	95.0	87.0	455	354	77.8	354	321	90.7	70.5	77.7
	5912	364	330	90.7	330	316	95.8	86.8	413	322	78.0	322	295	91.6	71.4	78.6
	5913	702	629	89.6	629	607	96.5	86.5	859	681	79.3	681	615	90.3	71.6	78.3
	5914	607	533	87.8	533	503	94.4	82.9	657	542	82.5	542	488	90.0	74.3	78.4
	5921	590	516	87.5	516	478	92.6	81.0	707	585	82.7	585	535	91.5	75.7	78.1
	5922	897	726	80.9	726	688	94.8	76.7	1105	867	78.5	867	767	88.5	69.4	72.7
	5923	997	873	87.6	873	800	91.6	80.2	1165	907	77.9	907	799	88.1	68.6	74.0
	5931	507	441	87.0	441	423	95.9	83.4	639	483	75.6	483	432	89.4	67.6	74.6
	5932	1092	829	75.9	829	793	95.7	72.6	1243	894	71.9	894	790	88.4	63.6	67.8
	5933	639	543	85.0	543	500	92.1	78.2	769	574	74.6	574	515	89.7	67.0	72.1
	5941	865	723	83.6	723	668	92.4	77.2	883	704	79.7	704	644	91.5	72.9	75.1
	5942	638	549	86.1	549	518	94.4	81.2	702	576	82.1	576	529	91.8	75.4	78.1
	5943	314	259	82.5	259	250	96.5	79.6	373	321	86.1	321	294	91.6	78.8	79.2
	5951	460	350	76.1	350	324	92.6	70.4	485	365	75.3	365	322	88.2	66.4	68.4
	5952	607	486	80.1	486	433	89.1	71.3	616	495	80.4	495	438	88.5	71.1	71.2
	5953	340	267	78.5	267	240	89.9	70.6	379	307	81.0	307	271	88.3	71.5	71.1
YT	6001	1211	1073	88.6	1073	1013	94.4	83.6	325	289	88.9	289	262	90.7	80.6	83.0
NT	6101	1219	1100	90.2	1100	1049	95.4	86.1	229	200	87.3	200	182	91.0	79.5	85.0
NU	6201	889	805	90.6	805	759	94.3	85.4	85.4

Taux de réponse par réseau local d'intégration des services de santé (RLISS) et par base de sondage de l'ESCC 2007-2008 en Ontario

		Area frame / Base aréolaire							Phone frames / Bases téléphoniques							
Prov. Terr.	RLISS	# 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
ON	Total	27904	23784	85.2	23784	21914	92.1	78.5	31813	25065	78.8	25065	22044	87.9	69.3	73.6
	3501	1760	1565	88.9	1565	1473	94.1	83.7	2189	1713	78.3	1713	1486	86.7	67.9	74.9
	3502	3244	2781	85.7	2781	2613	94.0	80.5	3504	2847	81.3	2847	2543	89.3	72.6	76.4
	3503	1500	1305	87.0	1305	1202	92.1	80.1	1657	1338	80.7	1338	1220	91.2	73.6	76.7
	3504	3233	2764	85.5	2764	2501	90.5	77.4	3644	2860	78.5	2860	2533	88.6	69.5	73.2
	3505	1393	1178	84.6	1178	1088	92.4	78.1	1655	1252	75.6	1252	1038	82.9	62.7	69.8
	3506	1446	1250	86.4	1250	1127	90.2	77.9	1728	1334	77.2	1334	1122	84.1	64.9	70.9
	3507	1453	1137	78.3	1137	1043	91.7	71.8	1794	1216	67.8	1216	1053	86.6	58.7	64.6
	3508	1753	1497	85.4	1497	1392	93.0	79.4	2410	1721	71.4	1721	1443	83.8	59.9	68.1
	3509	2751	2324	84.5	2324	2149	92.5	78.1	3019	2385	79.0	2385	2099	88.0	69.5	73.6
	3510	1650	1409	85.4	1409	1293	91.8	78.4	1752	1454	83.0	1454	1297	89.2	74.0	76.1
	3511	2624	2209	84.2	2209	2021	91.5	77.0	2785	2277	81.8	2277	2024	88.9	72.7	74.8
	3512	1391	1171	84.2	1171	1089	93.0	78.3	1503	1249	83.1	1249	1118	89.5	74.4	76.3
	3513	2479	2143	86.4	2143	1953	91.1	78.8	2700	2222	82.3	2222	2003	90.1	74.2	76.4
	3514	1227	1051	85.7	1051	970	92.3	79.1	1473	1197	81.3	1197	1065	89.0	72.3	75.4

ANNEXE F : Régions sociosanitaires

Régions sociosanitaires – Diffusion (avant FMGD)

Newfoundland

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

Prince Edward Island

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

Nova Scotia

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

New Brunswick

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

Quebec

- 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 Prince Edward Island

Nouvelle-Écosse

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

Nouveau-Brunswick

- 1301 Region 1
- 1302 Region 2
- 1303 Region 3
- 1304 Region 4/5
- 1306 Region 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
 4660 Parkland RHA
 4670 Norman RHA
 4685 Burntwood RHA/Churchill 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
 4660 Parkland/Norman/Burntwood-Churchill

Régions sociosanitaires – Diffusion (avant FMGD)

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

4821	Chinook Regional Health Authority
4822	Palliser Health Region
4823	Calgary Health Region
4824	David Thompson Regional Health Authority
4825	East Central Health
4826	Capital Health
4827	Aspen Regional Health Authority
4828	Peace Country Health
4829	Northern Lights Health Region

British Columbia

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

Territories

6001	Yukon Territory
6101	Northwest Territories
6201	Nunavut

Régions sociosanitaires dans le FMGD

Saskatchewan

4701	Sun Country/Five Hills/Cypress
4704	Regina Qu'Appelle RHA
4705	Sunrise/Kelsey Trail
4706	Saskatoon RHA
4707	Heartland/Prairie North
4709	Prince Albert Parkland/Mamawetan-Keewatin Yatthé-Athabaska

Alberta

4821	Chinook Regional Health Authority
4822	Palliser Health Region
4823	Calgary Health Region
4824	David Thompson Regional Health Authority
4825	East Central Health
4826	Capital Health
4827	Aspen Regional Health Authority
4828	Peace Country/Northern Lights
4829	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

Territoires

6001	Yukon Territory/Northwest Territories/Nunavut
------	---