

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

1.	Introduction	3
2.	Contexte.....	4
3.	Objectifs.....	6
4.	Contenu de l'enquête	7
4.1	Processus de consultation.....	7
4.2	Contenu commun	7
4.3	Contenu optionnel	8
5.	Plan d'échantillonnage	10
5.1	Population cible.....	10
5.2	Régions socio-sanitaires.....	10
5.3	Taille et répartition de l'échantillon.....	11
5.4	Bases de sondage et stratégies d'échantillonnage des ménages.....	11
5.4.1	Échantillonnage des ménages à partir de la base aréolaire	11
5.4.2	Échantillonnage des ménages à partir de la base de sondage à CA de numéros de téléphone.....	15
5.4.3	Échantillonnage des ménages à partir de la base liste de numéros de téléphone	15
5.5	Échantillonnage des personnes interviewées	16
5.6	Répartition de l'échantillon sur la période de collecte des données	17
5.7	Achat d'unités d'échantillonnage supplémentaires à l'Île-du-Prince-Édouard	17
6.	Collecte des données	19
6.1	Développement du questionnaire et méthode de collecte des données.....	19
6.2	Supervision et contrôle.....	19
6.3	Essais sur le terrain	20
6.4	Techniques d'interview	20
6.5	Réduire la non-réponse	20
6.6	Circonstances spéciales dans le cadre des opérations de collecte des données du cycle 1.1 de l' ESCC.....	21
7.	Traitement des données.....	23
7.1	Vérification	23
7.2	Codage	23
7.3	Création de variables dérivées et groupées	23
7.4	Imputation	23
7.5	Pondération	25
7.6	Élimination des renseignements confidentiels	26
8.	Pondération.....	27
8.1	Pondération du fichier.....	28
8.1.1	Pondération de l'échantillon provenant de la base aréolaire.....	28
8.1.2	Pondération de l'échantillon provenant de la base téléphonique	32
8.1.3	Intégration des bases aréolaire et téléphonique (I1).....	35
8.1.4	Effet saisonnier (I2)	36
8.1.5	Poststratification (I3).....	37
8.1.6	Particularités de la pondération pour les trois territoires	37
9.	Qualité des données	39
9.1	Taux de réponse	39

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

9.2	Erreurs dans les enquêtes	48
9.2.1	Erreurs non dues à l'échantillonnage	48
9.2.2	Erreurs dues à l'échantillonnage	49
10.	Lignes directrices pour la totalisation, l'analyse et la diffusion	50
10.1	Lignes directrices pour l'arrondissement	50
10.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation	51
10.2.1	Définitions des catégories d'estimations : de type nominal par opposition à quantitatives	51
10.2.2	Totalisation d'estimations de type nominal	52
10.2.3	Totalisation d'estimations quantitatives	53
10.3	Lignes directrices pour l'analyse statistique	53
10.4	Lignes directrices pour la diffusion	54
11.	Tableaux de la variabilité d'échantillonnage approximative	56
11.1	Comment utiliser les tableaux de CV pour les estimations de type nominal	61
11.2	Exemples d'utilisation des tableaux de CV pour des estimations de type nominal	63
11.3	Comment utiliser les tableaux de CV pour calculer les limites de confiance	66
11.4	Exemple d'utilisation de tableaux de CV pour obtenir des limites de confiance	67
11.5	Comment utiliser les tableaux de CV pour effectuer un test Z	68
11.6	Exemple d'utilisation des tableaux de CV pour effectuer un test Z	68
11.7	Variances ou coefficients de variation exacts	68
11.8	Seuils pour la diffusion des estimations relatives à l'ESCC	70
12.	Utilisation du fichier	76
12.1	Utilisation des facteurs de pondération	76
12.2	Convention appliquée pour nommer les variables	76
12.2.1	Structure élémentaire des noms des variables du cycle 1.1 de l' ESCC	76
12.2.2	Positions 1 à 3 : Nom de la variable/section du questionnaire	77
12.2.3	Position 4 : Cycle	78
12.2.4	Position 5 : Type de variable	78
12.2.5	Positions 6 à 8 : Nom de la variable	78
12.3	Accès aux fichiers maîtres	79

Liste des annexes

Annexe A : [Questionnaire](#)

Annexe B : [Cliché d'enregistrement](#)

Annexe C : [Dictionnaire des données](#)

Annexe D : [Variables dérivées et groupées](#)

Annexe E : [Tableaux de CV](#)

1. Introduction

L'Enquête sur la santé dans les collectivités canadiennes (l'ESCC) est une enquête transversale qui vise à recueillir des renseignements sur l'état de santé, l'utilisation des services de santé et les déterminants de la santé de la population canadienne. Le cycle de collecte des données de l'ESCC s'étend sur deux années. La première année du cycle, indiquée par la notation « .1 », correspond à une enquête générale sur la santé de la population réalisée auprès d'un grand échantillon et conçue pour fournir des estimations fiables au niveau de la région socio-sanitaire. La deuxième année du cycle, représentée par la notation « .2 », correspond à une enquête de moins grande portée conçue pour fournir des données au niveau provincial sur des sujets particuliers ayant trait à la santé.

Le présent fichier de microdonnées contient les données du cycle 1.1 de l'ESCC, c'est-à-dire la première année du cycle. Les renseignements ont été recueillis de septembre 2000 à novembre 2001 pour 136 régions socio-sanitaires couvrant les dix provinces et les trois territoires. Les données du cycle 1.1 de l'ESCC sont recueillies auprès des personnes de 12 ans et plus vivant dans des logements privés. Sont exclus de la base de sondage les habitants des réserves indiennes et des terres de la Couronne, les résidents des établissements, les membres à temps plein des Forces armées canadiennes et les habitants de 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 transversales du cycle 1.1 de l'ESCC qui sont décrits en détail dans le texte et les annexes qui suivent.

Pour toute question concernant les ensembles de données ou leur utilisation, s'adresser à :

- Soutien technique ou renseignements généraux sur les données :
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 : 1-613-951-1746
Courriel : hd-ds@statcan.ca
- Renseignements sur le télé-accès : 1-613-951-1653
Courriel : cchs-escc@statcan.ca
Télécopieur : 1-613-951-4198

2. Contexte

En 1991, le Groupe de travail national sur l'information en matière de santé a relevé plusieurs 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 national de l'information sur la santé.

L'Initiative du Carnet de route a été lancée en réponse directe aux préoccupations et aux souhaits exprimés par plus de 500 personnes représentant un large éventail d'organismes et de groupes d'intérêt. Au début de 1999, les trois organismes nationaux susmentionnés ont mené une consultation nationale à grande échelle sur les besoins d'information en matière de santé. Les participants ont insisté sur le fait que les organismes nationaux doivent collaborer en vue de renforcer le système canadien d'information sur la santé et mettre à profit les investissements et les compétences considérables aux niveaux local, régional et provincial/territorial².

Le Carnet de route représente une contribution importante à l'édification d'un système national complet d'information sur la santé et de l'infrastructure requise pour donner aux Canadiens l'information dont ils ont besoin pour entretenir et améliorer le système de santé et la santé de la population du Canada³. Un plan d'action coordonné est requis. Le gouvernement seul ou une seule organisation ne peut pas lutter contre les problèmes mentionnés plus haut. La collaboration à tous les niveaux — organismes de santé nationaux, provinciaux, territoriaux, régionaux et locaux — est une condition préalable au succès⁴.

Notre système d'information sur la santé devrait nous fournir l'information pour répondre aux questions cruciales ci-dessous⁵ :

1. À quel point le système de santé est-il sain?
2. À quel point les Canadiens sont-ils en santé?

La première question englobe l'efficacité, l'efficience et la réceptivité du système de santé. En règle générale, un système de santé efficace, efficient et réceptif est un système qui offre aux Canadiens les soins de qualité auxquels ils s'attendent⁶.

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

² 1999. Ibid. p. 1.

³ 1999. Ibid. p. 1.

⁴ 1999. Ibid. p. 3.

⁵ 1999. Ibid. p. 3.

⁶ 1999. Ibid. p. 3.

La deuxième question est plus générale et traite des objectifs de base du système : la santé des Canadiens s'améliore-t-elle? Afin de répondre à cette question et à d'autres aussi importantes, nous avons besoin d'un système solide d'information sur la santé⁷. Ce système doit posséder six grandes caractéristiques⁸. Il doit être :

- sécuritaire et respecter le droit des Canadiens à la vie privée;
- cohérent;
- pertinent;
- intégrable;
- flexible;
- convivial et accessible.

Ce nouveau système d'information sur la santé doit être à jour, fournir des renseignements orientés vers la personne et s'appuyer sur des normes de données communes à d'autres enquêtes sur la santé de la population canadienne, telles que l'Enquête nationale sur la santé de la population (ENSP). Il doit également fournir de nouveaux ensembles de données ou des ensembles de données étoffées, des données sur les services de santé, des données sur les résultats relatifs à la santé, l'état de santé et les déterminants non médicaux de la santé, des données sur les résultats d'interventions particulières, des études spéciales portant sur des questions prioritaires, des données sur les coûts selon le service, des protocoles d'échange de données, une plus grande capacité d'analyse des données, ainsi que des rapports publics sur le système de santé⁹.

L'Enquête sur la santé dans les collectivités canadiennes (ESCC) a été conçue compte tenu de ce mandat. Le format, le contenu et les objectifs de cette enquête ont été définis après avoir mené des consultations approfondies auprès de spécialistes et d'intervenants fédéraux, provinciaux et communautaires en vue de déterminer leurs exigences en matière de données¹⁰.

Le présent Guide du fichier de microdonnées à grande diffusion est publié en réponse à l'exigence de recueillir des données fiables et pertinentes sur les services de santé, l'état de santé et les questions relatives à la santé revêtant une importance pour la population canadienne — aux niveaux régional, provincial et national — et de diffuser cette information au public.

⁷ 1999. Ibid. p. 5.

⁸ Ces caractéristiques sont décrites en détail dans le document intitulé Carnet de route de l'information sur la santé : Répondre aux besoins, 1999, Institut canadien d'information sur la santé. ISBN 1-895581-30-3. (<http://www.cihi.ca>)

⁹ 1999. Ibid. p. 11-14.

¹⁰ 1999. Initiative du carnet de route ... Lancer le processus. Institut canadien d'information sur la santé/Statistique Canada. ISBN 1-895581-70-2. p. 19.

3. Objectifs

Les objectifs principaux du cycle 1.1 de l'ESCC sont les suivants :

- fournir des estimations transversales à jour et fiables des déterminants de la santé, de l'état de santé et de l'utilisation des services de santé à travers le Canada;
- recueillir des données au niveau infraprovincial;
- créer un instrument d'enquête souple permettant :
 - de combler des lacunes statistiques particulières au niveau de la région socio-sanitaire;
 - d'élaborer un contenu d'enquête thématique en vue de recueillir des données importantes;
 - de répondre aux nouvelles questions ayant trait à la santé et aux services de santé à mesure qu'elles surviennent.

L'ESCC, en tant que composante importante du Programme des enquêtes sur la santé, permet de combler des besoins d'information accrus en matière de santé. Il s'agit de

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

4. Contenu de l'enquête

La première sous-section du présent section donne une description générale du processus de consultation mis en place pour élaborer le contenu de l'enquête et un résumé du contenu final sélectionné pour le cycle 1.1 de l'enquête. La deuxième sous-section décrit le contenu commun en détail. La troisième sous-section illustre le contenu des modules optionnels du cycle 1.1.

4.1 Processus de consultation

L'un des principaux objectifs de l'ESCC est de combler les lacunes statistiques ayant trait aux déterminants de la santé, à l'état de santé et à l'utilisation des services de santé au niveau de la région socio-sanitaire. Les modules formant le contenu du cycle 1.1 de l'ESCC ont été sélectionnés après avoir consulté longuement les représentants des autorités régionales, provinciales et fédérales et les chercheurs. En juin 1999, on a lancé auprès des principaux utilisateurs de l'information sur la santé un processus de consultation de portée nationale basé sur une série d'ateliers d'un jour sur un système de fiches de travail pour le classement des sujets par ordre de priorité. Ont participé aux ateliers des utilisateurs de données représentants les régions socio-sanitaires, les ministères provinciaux de la Santé et les organismes fédéraux, ainsi que des chercheurs travaillant en milieu universitaire ou hospitalier et des groupes d'intérêts spéciaux. En tout, plus de 17 ateliers ont eu lieu auxquels ont participé plus de 225 personnes.

Les consultations ont montré que les intervenants des divers paliers de représentation s'entendaient en grande partie sur les priorités relatives au contenu et aux besoins de données, mais ont aussi révélé certaines différences. En dernière analyse, tous les sujets et sous-sujets possibles d'enquête ont été classés comme ayant une priorité forte, moyenne ou faible d'après les discussions avec les participants aux ateliers, les données des fiches de travail et d'autres discussions avec des chercheurs spécialisés dans le domaine des enquêtes sur la santé.

Toutes ces consultations ont abouti à un questionnaire comprenant un contenu commun et un contenu optionnel. Le contenu commun englobe les sujets considérés comme étant hautement prioritaire par toutes les régions socio-sanitaires. Le volet du contenu optionnel du questionnaire regroupe les sujets correspondant à des besoins locaux de données qui ont été choisis individuellement par chaque région socio-sanitaire. Les sujets communs et optionnels sont décrits de façon plus détaillée aux sous-sections suivantes.

4.2 Contenu commun

Le contenu commun englobe des sujets très divers, allant de la consommation d'alcool et la dépendance à l'égard de l'alcool à l'activité physique et l'incapacité au cours des deux dernières semaines. Le tableau qui suit donne la liste des modules communs de l'enquête réalisée la première année du premier cycle de l'ESCC (cycle 1.1), telle

qu'établie d'après les consultations sur le contenu tenues en juin 1999 dans les diverses régions du Canada.

Les questions portant sur les sujets formant le contenu commun ont été posées à tous les répondants dans toutes les régions socio-sanitaires. Les réponses à ces questions fournissent une base nationale de renseignements sur la santé qui, pondérés, permettent de réaliser une étude transversale des questions ayant trait à la santé au Canada.

Tableau 4.1 : Modules du contenu commun

<ul style="list-style-type: none"> • Consommation d'alcool • Dépendance à l'égard de l'alcool • Tension artérielle • Allaitement • Problèmes de santé chroniques • Consultation des spécialistes de la santé mentale • Exposition à la fumée des autres • Insécurité alimentaire • Consommation de fruits et de légumes • État de santé général • Utilisation des soins de santé • Indice de l'état de santé (HUI) • Taille et poids • Blessures 	<ul style="list-style-type: none"> • Mammographie • Test Papanicolaou • Activités physiques • Test de l'antigène spécifique prostatique • Limitation des activités • Usage du tabac • Variantes du tabagisme • Incapacité au cours des deux dernières semaines • Composition du ménage et conditions de logement • Revenu • Population active • Caractéristiques sociodémographiques • Administration
--	--

4.3 Contenu optionnel

Les sujets formant les modules de contenu optionnel ont également été relevés durant le processus de consultation (voir le tableau 4.2). Cependant, on les a considérés comme étant optionnels, pour que toute région ayant besoin de données sur un sujet particulier puissent sélectionner le module pertinent afin de l'inclure dans le questionnaire du cycle 1.1 de l'ESCC la concernant. L'avantage de cette démarche est que les régions socio-sanitaires peuvent étendre la couverture des sujets ayant trait à la santé en fonction des caractéristiques de la région. L'inconvénient tient au fait que, contrairement aux données provenant des modules du contenu commun, celles résultant des modules du contenu optionnel ne peuvent être généralisées facilement à l'ensemble du Canada. Par conséquent, la taille et les caractéristiques des régions où sont utilisés les modules du contenu optionnel limitent la comparaison interrégionale.

Tableau 4.2 : Modules du contenu optionnel

<ul style="list-style-type: none">• Examen des seins• Auto-examen des seins• Changements pour améliorer la santé• Facteurs de stress chez les enfants et les adultes (traumatismes)• Visites chez le dentiste• Dépression• Détresse• Conduite automobile et consommation d'alcool• Consommation de médicaments• Examen de la vue• Vaccins contre la grippe• Soins à domicile• Contrôle	<ul style="list-style-type: none">• Humeur• Problèmes courants• Examen général• Événements récents• Activités sédentaires• Estime de soi• Comportement sexuel• Outils pour cesser de fumer• Soutien social• Spiritualité• Pensées suicidaires et tentatives de suicide• Utilisation de précautions (équipement de protection)• Stress au travail
--	--

5. Plan d'échantillonnage

5.1 Population cible

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

5.2 Régions socio-sanitaires

À des fins administratives, chaque province est divisée en plusieurs régions socio-sanitaires (RSS) et chaque territoire est considéré comme formant une RSS unique (tableau 5.1). En collaboration avec les provinces, Statistique Canada a modifié légèrement les limites de certaines RSS afin qu'elles correspondent aux données géographiques du Recensement de 1996. Durant le cycle 1.1 de l'ESCC, des données ont été recueillies pour 133 RSS dans les 10 provinces, ainsi que pour une RSS par territoire, soit, en tout, pour 136 RSS.

Tableau 5.1. Nombre de régions socio-sanitaires et tailles visées d'échantillon selon la province/territoire

Province	Nombre de RSS	Taille totale de l'échantillon (visée)
Terre-Neuve	6	4 010
Île-du-Prince-Édouard	2	2 000
Nouvelle-Écosse	6	5 040
Nouveau-Brunswick	7	5 150
Québec	16	24 280
Ontario	37	42 260
Manitoba	11	8 000
Saskatchewan	11	7 720
Alberta	17	14 200
Colombie-Britannique	20	18 090
Yukon	1	850
Territoires du Nord-Ouest	1	900
Nunavut	<u>1</u>	<u>800</u>
Canada	136	133 300

5.3 Taille et répartition de l'échantillon

Afin de produire des estimations fiables pour les 136 RSS et compte tenu du budget accordé pour le cycle 1.1 de l'ESCC, il a été établi que cette composante de l'enquête devrait être réalisée auprès d'un échantillon de 133 300 personnes. La production d'estimations fiables au niveau de la RSS était l'objectif primordial, mais la qualité des estimations pour certaines caractéristiques importantes au niveau provincial a été jugée importante également. Par conséquent, la stratégie de répartition de l'échantillon, qui comporte trois étapes, accorde une importance plus ou moins égale aux RSS et aux provinces. Lors des deux premières étapes, l'échantillon a été réparti entre les provinces en fonction de leur compte de population et du nombre de RSS qu'elles contiennent (tableau 5.1). À la troisième étape, chaque échantillon provincial a été réparti entre les RSS proportionnellement à la racine carrée de la population estimée de la RSS.

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

Il convient de souligner que les trois territoires, qui ont été traités séparément, n'étaient pas visés par la stratégie susmentionnée de répartition de l'échantillon. En tout, 850 unités d'échantillonnage ont été attribuées au Yukon, 900 aux Territoires du Nord-Ouest et 800 au Nunavut.

5.4 Bases de sondage et stratégies d'échantillonnage des ménages

L'échantillon de ménages du cycle 1.1 de l'ESCC a été sélectionné à partir de trois bases de sondage. La majorité de l'échantillon (83 %) provient d'une base de sondage aréolaire. Pour certaines RSS, on a utilisé une base de sondage à composition aléatoire (CA) et/ou une base liste de numéros de téléphone. Environ 7 % de l'échantillon de ménages provenait de la base de sondage à CA et presque 10 %, de la base liste de numéros de téléphone.

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'échantillonnage¹¹. À la première étape, on a

¹¹ Statistique Canada (1998). *Méthodologie de l'Enquête sur la population active du Canada*. Statistique Canada, numéro 71-526-XPB au catalogue.

formé des strates homogènes et sélectionné un échantillon indépendant de grappes, dans chaque strate. À la deuxième étape, on a dressé la liste des logements pour chaque grappe, puis on a sélectionné les logements, ou les ménages, d'après les listes.

Pour les besoins du plan d'échantillonnage, chaque province est répartie en trois catégories de région, à savoir les grands centres urbains, les villes et les régions rurales. Des strates géographiques ou socioéconomiques sont formées à l'intérieur de chaque grand centre urbain. Dans les strates, des grappes sont formées par regroupement de 150 à 250 logements. Dans certains centres urbains, des strates distinctes sont créées pour les immeubles à appartements ou les secteurs de dénombrement (SD) pour lesquels le revenu moyen du ménage est élevé. Dans chaque strate, on sélectionne six grappes ou immeubles résidentiels (pouvant compter de 12 à 18 appartements) 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, on sélectionne six grappes (habituellement des SD) par la méthode PPT. Pour les strates où la densité de population est faible, on suit un plan en trois étapes en vertu duquel on sélectionne deux ou trois unités primaires d'échantillonnage (UPE), qui correspondent normalement à des groupes de SD, puis on les répartit en grappes dont six sont sélectionnées pour faire partie de l'échantillon. La sélection est réalisée à chaque étape selon la méthode PPT.

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

Tableau 5.2. Unité primaire d'échantillonnage, taille et rendement

Région	Unité primaire d'échantillonnage (UPE)	Taille (ménages par UPE)	Rendement (ménages échantillonnés)
Toronto, Montréal, Vancouver	Grappe	200 à 250	6
Autres villes	Grappe	150 à 200	8
Base des immeubles à appartements	Appartement	Varie	5
La plupart des régions rurales/petits centres urbains	Secteur de dénombrement	300	10

Afin de répondre aux exigences particulières à l'ESCC, certaines modifications ont dû être apportées à cette stratégie d'échantillonnage¹². Pour obtenir un échantillon de base de 97 000 ménages pour l'ESCC, il faut sélectionner 123 000 logements d'après la base aréolaire (pour tenir compte des logements vacants et des ménages non-répondants). Chaque mois, le plan d'échantillonnage de l'EPA fournit environ 68 000 logements répartis entre les diverses régions économiques du Canada, alors que, pour l'ESCC, il faut obtenir un échantillon total de 123 000 logements répartis entre les RSS, dont les limites géographiques diffèrent de celles des régions économiques de l'EPA. Globalement, l'ESCC nécessite la sélection d'un nombre de logements presque deux fois plus élevé que celui que produit le mécanisme de sélection de l'EPA, ce qui correspond à un *facteur de redressement* de 1,8 (123 000/68 000). Toutefois, comme ce facteur de redressement varie de 0,6 à 6,0 au niveau de la RSS, certains ajustements sont nécessaires.

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

¹² Morano M., Lessard, S. et Béland, Y. (2000). Creation of a dual frame for the Canadian Community Health Survey, 2000 *Proceedings of the Survey Methods Section*, Ottawa: La Société statistique du Canada, 249-254.

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

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

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

Il convient de mentionner que la base de sondage du cycle 1.1 de l'ESCC couvre 90 % des ménages privés du Yukon, 97 % de ceux des Territoires du Nord-Ouest et 90 % de ceux du Nunavut.

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

Dans certaines RSS, on a utilisé pour certains mois de collecte, un échantillon de numéros de téléphone provenant de la base de sondage à composition aléatoire (CA) en plus de la base aréolaire. 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 générale¹³. 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 socio-sanitaires. À l'intérieur de chaque strate CA, on a choisi au hasard une banque de cent numéros et généré aléatoirement un numéro compris entre 00 et 99 afin de créer un numéro de téléphone complet à 10 chiffres. Cette méthode a été répétée jusqu'à ce que l'on ait atteint le nombre requis de numéros de téléphone pour la strate CA. Comme, fréquemment, le numéro obtenu n'est pas en service ou est hors du champ d'observation, il faut générer un grand nombre de numéros supplémentaires pour atteindre la taille visée d'échantillon. Ce taux de réussite diffère selon la région. Dans le cas du cycle 1.1 de l'ESCC, il variait de 15 % à 61 % au niveau régional.

La base de sondage à CA a été utilisée à un degré variable dans plus de 60 RSS pour compléter la base aréolaire et a été la seule base de sondage utilisée pour cinq RSS.

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

À l'instar de la base de sondage à CA, on a utilisé une base liste de numéros de téléphone pour compléter la base aréolaire pour certaines RSS pour certains mois de collecte. À cette fin, on a couplé l'annuaire téléphonique du Canada, un disque compact disponible sur le marché contenant les noms, les adresses et les numéros de téléphone répertoriés dans les annuaires de téléphone du Canada, à des fichiers

¹³ 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.

internes administratifs de conversion afin d'obtenir les codes postaux que l'on a fait correspondre aux RSS afin de créer des listes stratifiées de numéros de téléphone, à raison d'une liste par RSS. Dans chaque strate, on a sélectionné le nombre requis de numéros de téléphone d'après la base liste par échantillonnage aléatoire simple. Comme pour la base de sondage à CA, on a sélectionné des numéros de téléphone supplémentaires pour tenir compte des numéros hors service ou hors du champ d'observation. Le taux de réussite enregistré pour la sélection d'après la liste de numéros de téléphone est nettement plus élevé que celui observé pour la liste établie par CA, variant de 50 % à 65 %.

Il importe de souligner que la couverture de la base liste de numéros de téléphone est moins importante que celle de la base de sondage à CA, car les numéros non publiés n'ont aucune chance d'être sélectionnés. Néanmoins, comme la liste des numéros de téléphone n'a été utilisée que pour les RSS pour lesquelles la base aréolaire était la source principale de l'échantillon, l'effet du sous-dénombrement dû à l'utilisation de la base liste de numéros de téléphone a été minimal et a pu être corrigé par pondération.

À des degrés divers, la base liste de numéros de téléphone a été utilisée pour plus de 50 RSS pour compléter la base aréolaire.

5.5 Échantillonnage des personnes interviewées

La sélection des répondants a été conçue de façon à ce que les jeunes (de 12 à 19 ans) et les personnes âgées (65 ans et plus) soient surreprésentés dans l'échantillon. La stratégie d'échantillonnage adoptée tient compte des besoins des utilisateurs de données, du coût, de l'efficacité du plan d'échantillonnage, du fardeau de réponse et des contraintes opérationnelles¹⁴.

Pour 82 % des ménages échantillonnés à partir de la base aréolaire, on a sélectionné au hasard dans chaque ménage une personne de 12 ans et plus, et pour les 18 % restants, on a sélectionné au hasard dans chaque ménage deux personnes de 12 ans et plus. La règle de sélection des personnes dans les ménages faisant partie de la base aréolaire était basée sur la composition du ménage. Le tableau 5.3 décrit la règle de sélection des personnes dans les ménages échantillonnés à partir de la base aréolaire.

¹⁴ Béland, Y., Bailie, L., Catlin, G. et Singh, M.P. CCHS and NPHS — An Improved Health Survey Program at Statistics Canada, *2000 Proceedings of the American Statistical Association Meeting, Survey Research Methods Section*, Indianapolis: American Statistical Association, 677-682.

Tableau 5.3. Stratégie de sélection fondée sur la composition du ménage — échantillon provenant de la base aréolaire

Nombre de personnes de 12 à 19 ans	Nombre de personnes de 20 ans et plus					
	0	1	2	3	4	5+
0	-	A	A	A	A	B
1	A	A	C	C	C	B
2	A	C	C	C	C	C
3+	A	C	C	C	C	C

A : sélection aléatoire d'une personne de 12 ans et plus

B : sélection aléatoire de deux personnes de 12 ans et plus

C : sélection aléatoire d'une personne dans le groupe des 12 à 19 ans **et** sélection aléatoire d'une personne de 20 ans et plus

Pour tous les ménages sélectionnés à partir des bases de sondage téléphoniques, on a sélectionné au hasard dans chaque ménage une seule personne de 12 ans et plus parmi l'ensemble des membres du ménage.

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

Afin d'équilibrer la charge de travail des intervieweurs et de réduire au minimum les effets saisonniers éventuels sur des caractéristiques importantes telle que l'activité physique, dans chaque RSS, l'échantillon initial de logements/numéros de téléphone a été réparti au hasard, de façon égale, sur les 12 mois de la collecte des données. Pour commencer, chaque UPE sélectionnée au premier degré de l'échantillonnage à partir de la base aréolaire a été affectée au hasard à un trimestre de collecte des données (Q1 : septembre à novembre 2000, Q2 : décembre 2000 à février 2001, Q3 : mars à mai 2001 et Q4 : juin à août 2001). Pour chaque trimestre de collecte, les logements sélectionnés ont alors été attribués au hasard à un mois de collecte. Pour les listes des numéros de téléphone, des échantillons indépendants ont été sélectionnés chaque mois. Cette stratégie a permis d'assurer que chaque échantillon trimestriel soit représentatif de la population canadienne faisant partie du champ d'observation de l'enquête.

5.7 Achat d'unités d'échantillonnage supplémentaires à l'Île-du-Prince-Édouard

Au cours de la collecte des données, le gouvernement de l'Île-du-Prince-Édouard a fourni des fonds supplémentaires afin que l'on puisse sélectionner un échantillon plus important de logements. Cet achat d'unités d'échantillonnage supplémentaires avait pour objet d'obtenir un échantillon de taille suffisante pour produire des estimations fiables pour cinq régions infraprovinciales. Le plan d'échantillonnage original du cycle 1.1 de l'ESCC ne tenait compte que de deux régions infraprovinciales. Les unités d'échantillonnage supplémentaires ont été regroupées à l'échantillon principal

afin de produire un grand fichier de données. Pour des raisons de confidentialité, seules les deux régions initiales (urbaine et rurale) sont considérées sur ce fichier microdonnées.

L'échantillon d'unités supplémentaires a été sélectionné entièrement à partir de la base liste de numéros de téléphone. Pour cela, on a couplé l'annuaire du téléphone du Canada à des fichiers administratifs internes afin de stratifier les numéros de téléphone publiés dans les cinq régions infraprovinciales (West Prince, East Prince, Queens, Southern Kings et Eastern Kings). La taille de l'échantillon sélectionné par région infraprovinciale a été établie d'après les fonds disponibles et les exigences de la province quant à l'obtention d'estimations fiables selon la région infraprovinciale. En tout, 1 300 unités d'échantillonnage supplémentaires ont été ajoutées à l'échantillon de 2 000 unités prévues pour l'Île-du-Prince-Édouard. Les 3 300 unités d'échantillonnage ont été réparties entre les régions infraprovinciales par la méthode de la racine carrée du compte de population. Cette méthode bien connue de répartition permet de réaliser un juste équilibre entre les exigences de fiabilité visant les données provinciales et infraprovinciales. Le tableau 5.4 donne la répartition de l'échantillon selon la région infraprovinciale. Les données concernant ces unités d'échantillonnage supplémentaires ont été recueillies de mai à octobre 2001.

Tableau 5.4. Répartition finale de l'échantillon, y compris les unités d'échantillonnage supplémentaires, à l'Île-du-Prince-Édouard

Région infraprovinciale	Taille de l'échantillon
West Prince	525
East Prince	780
Queens	1 055
Southern Kings	520
Eastern Kings	420
Total	3 300

6. Collecte des données

6.1 Développement du questionnaire et méthode de collecte des données

Le questionnaire du cycle 1.1 de l' ESCC a fait l'objet d'interviews assistées par ordinateur (IAO). Des unités d'échantillonnage sélectionnées à partir de la base aréolaire ont répondu aux questions suivant la méthode d'interview en personne assistée par ordinateur (IPAO) tandis que les autres unités, sélectionnées à partir des bases de sondage téléphoniques, ont répondu aux questions suivant la méthode de l'interview téléphonique assistée par ordinateur (ITAO).

L'IAO procure un certain nombre d'avantages quant à la qualité des données par rapport aux autres méthodes de collecte. Premièrement, le libellé des questions, comprenant les périodes de référence et les pronoms, est personnalisé automatiquement en fonction de facteurs comme l'âge et le sexe du répondant, de la date de l'interview et des réponses aux questions précédentes.

En second lieu, on applique des mesures de contrôle qui isolent les réponses incohérentes ou hors normes, et des prompts apparaissent à l'écran lorsqu'une entrée incorrecte est enregistrée. Le répondant reçoit une rétroaction immédiate et l'intervieweur peut corriger toute incohérence.

Troisièmement, le processus permet de sauter automatiquement les questions qui ne concernent pas le répondant.

6.2 Supervision et contrôle

Les intervieweurs IPAO travaillaient séparément depuis leur domicile à l'aide d'un ordinateur portable et étaient supervisés à distance par des intervieweurs principaux. Les interviews complétées étaient envoyées quotidiennement de leur domicile au bureau central de Statistique Canada, et ce, par transmission téléphonique protégée.

Les intervieweurs ITAO travaillaient dans des bureaux centralisés sous la supervision d'un intervieweur principal. Il incombait au surveillant de projet du bureau régional, à l'intervieweur principal et à l'équipe d'assistance technique de transmettre les dossiers de chacun des cinq bureaux ITAO au bureau central.

On ne pouvait compter, pour la collecte par ITAO, sur un ordonnanceur d'appels automatique, c'est-à-dire un système central qui optimise l'agencement des rappels et l'ordonnancement des rendez-vous. Au lieu, un lot de cas était distribué au début de chaque mois à chaque ordinateur de tous les bureaux ITAO. Le nombre de cas attribués à chaque ordinateur faisait alors l'objet d'un traitement manuel. À cause du nombre relativement restreint de cas réservés à l'ITAO, cette démarche s'est révélée raisonnablement efficace et l'absence d'un ordonnanceur d'appels ne semble pas avoir eu d'effet nuisible sur la qualité des données.

6.3 Essais sur le terrain

À la fin du printemps et au début de l'été de 2000, on a procédé à des essais sur le terrain étudiant séparément les méthodes IPAO et ITAO. Les essais s'étendaient aux cinq bureaux régionaux de Statistique Canada.

Les principaux objectifs des essais sur la méthode IPAO étaient d'évaluer les réactions des répondants aux questions et d'obtenir des estimations quant au temps requis pour remplir les diverses sections du questionnaire. On a aussi évalué les procédures des opérations sur le terrain, la formation des intervieweurs et l'application IAO.

Les tests sur la méthode ITAO visaient des objectifs similaires. On a également évalué l'infrastructure technique des bureaux ITAO de même que les procédés d'interview propres à l'ITAO.

6.4 Techniques d'interview

Dans tous les logements choisis, on demandait à un membre du ménage bien informé de fournir l'information démographique de base sur tous les occupants du logement. Selon la composition du ménage et selon que la méthode d'interview choisie était l'IPAO ou l'ITAO, on choisissait d'interviewer une ou deux personnes plus en profondeur.

Les intervieweurs IPAO étaient formés pour effectuer un premier contact personnel avec les occupants de chacun des logements échantillonnés. Dans les cas où la première visite entraînait une non-réponse, il était permis de faire un suivi par téléphone.

Dans les cas où le répondant choisi était absent pour une période de temps prolongée ou incapable de compléter l'interview, un autre membre informé du ménage a fourni l'information requise au sujet du répondant choisi. Il s'agit là d'une interview par procuration. Quoique les interviewés étaient 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 de ces interviews par procuration sont demeurées sans réponse. Il fallait donc tout tenter pour réduire au minimum le nombre d'interviews de ce genre.

6.5 Réduire la non-réponse

Avant même que l'intervieweur n'effectue un premier contact, les occupants du logement retenu avaient reçu une lettre de présentation et une brochure. Ces documents expliquaient l'importance de l'enquête et fournissaient des exemples sur la façon dont les données du cycle 1.1 de l'ESCC allaient être utilisées.

Les intervieweurs ont reçu instructions de faire toutes les tentatives raisonnables pour obtenir les interviews nécessaires à l'ESCC. Lorsque la visite de l'intervieweur tombait au mauvais moment, il prenait un rendez-vous à un moment plus convenable. S'il n'y avait personne à la maison, il effectuait de nombreuses visites de rappel. Aux personnes qui refusaient dès le premier contact de participer à l'ESCC, le bureau régional envoyait une lettre insistant sur l'importance de l'enquête et de la collaboration du ménage. Suivait un second appel (ou visite), d'un intervieweur principal, d'un surveillant de projet ou un d'autre intervieweur qui tentait de convaincre les répondants de l'importance de participer à l'enquête. Au cours des derniers mois de la collecte des données, on a revu les cas de non-réponse et contacté à nouveau les personnes choisies qui avaient précédemment refusé pour les encourager à participer à l'enquête. Cette diligence à assurer le contact a peut-être contribué à obtenir de meilleurs résultats en maximisant le taux de réponse.

Pour pallier le problème de langue susceptible de nuire aux interviews, tous les bureaux régionaux de Statistique Canada ont embauché des intervieweurs qui parlaient plusieurs langues. Lorsqu'il le fallait, des cas étaient transférés à un intervieweur capable de remplir le questionnaire dans la langue voulue. De plus, les questions de l'enquête étaient traduites dans les langues suivantes : le mandarin, le panjabi et l'inuktitut.

6.6 Circonstances spéciales dans le cadre des opérations de collecte des données du cycle 1.1 de l'ESCC

Le plan original prévoyait que la collecte des données s'effectuerait entre septembre 2000 et le début d'octobre 2001, donc sur une période de 13 mois. Ce plan avait fait l'objet d'une élaboration soignée qui assurait la réalisation des objectifs de qualité des données de l'enquête. Pour répartir équitablement la charge de travail des intervieweurs et éliminer tout effet saisonnier, l'échantillon final a été divisé au hasard en 12, de manière à le distribuer judicieusement entre les mois, sur chaque RSS. On a ajouté un 13^e mois de collecte afin de donner aux intervieweurs l'occasion de faire une dernière tentative de résoudre les cas de non-réponse.

Dans la plupart des cas, les opérations de collecte lors des enquêtes auprès des ménages se déroulent rondement et dans les limites des paramètres établis. Pour l'ESCC, la charge de travail imposée par la taille de l'échantillon a constitué un défi de taille pour l'infrastructure de la collecte de données en place. Pour assurer le succès des opérations de collecte, il a fallu modifier un certain nombre de procédés établis, certains plus que d'autres.¹⁵ L'ajout d'un 14^e mois de collecte, le transfert de cas d'un bureau ITAO à un autre et l'introduction d'une nouvelle méthode d'échantillonnage pour choisir les numéros de téléphone figurent parmi les changements de procédés les

¹⁵ Béland, Y. , Dufour, J. and Hamel, M. (2001). Faire échec à la non-réponse dans le cadre de l'Enquête sur la santé dans les collectivités canadiennes, *Recueil des présentations effectuées au Symposium 2001 organisé par Statistique Canada*, Statistique Canada.

plus importants. À la fin de la collecte des données, le taux de réponse à l'échelle nationale s'est élevé à 84,7 %. Le lecteur trouvera tous les détails concernant les taux de réponse au section 9.

7. Traitement des données

7.1 Vérification

L'application IAO a effectué la plus grande partie de la vérification des données au moment de l'interview. Les intervieweurs ne pouvaient pas taper 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.

Il est possible de donner une réponse ouverte à plusieurs questions du questionnaire du cycle 1.1 de l'ESCC. Les réponses ouvertes à certaines de ces questions étaient codées à l'intérieur de l'une des catégories inscrites sur la liste à partir du moment où l'information écrite renvoyait à une catégorie sur la liste.

7.3 Création de variables dérivées et groupées

Pour faciliter l'analyse des données, on a dérivé un certain nombre de variables à partir des éléments trouvés dans le questionnaire du cycle 1.1 de l'ESCC. Le cinquième caractère du nom des variables dérivées est en général un « D » ou un « G ». 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. L'annexe D fournit des détails sur la façon de dériver ces variables plus complexes.

7.4 Imputation

À cause de leur caractère privé ou délicat, de nombreuses questions ou modules du questionnaire du cycle 1.1 de l'ESCC ne se prêtaient qu'à des réponses personnelles et étaient sautées lorsque l'interview était effectuée par procuration. Au cours de la collecte des données, une plus grande proportion que prévue d'interviews a été faite par procuration. Il n'était permis de recourir à ces interviews par procuration qu'après confirmation que le répondant sélectionné ne serait pas disponible au cours de la période de collecte, en cas d'incapacité mentale ou physique, ou lorsqu'il y avait barrière de langue. À la fin de la collecte des données, 6,3 % des interviews avaient été

complétées par procuration; le taux variait entre 2 % et 23 % dans les régions socio-sanitaires. Si bien qu'il manquait des renseignements importants sur les personnes représentées, ce qui valait pour le tiers, environ, du questionnaire. On a dû sauter dix modules communs du questionnaire entièrement et deux, partiellement, de même que 21 des modules optionnels.

Par conséquent, des valeurs ont été imputées par la méthode du « plus proche voisin » à l'étape du traitement des données pour remplacer les renseignements non recueillis durant les interviews par procuration¹⁶. L'imputation n'a été utilisée que pour compléter les renseignements recueillis sur les personnes pour lesquelles l'interview avait été réalisée par procuration. Elle n'a pas été utilisée en cas de non-réponse totale ou partielle lors des interviews réalisées auprès de la personne sélectionnée proprement dite. En vertu de cette méthode, on a recherché, dans des catégories d'imputation prédéfinies, un répondant « donneur » dont les caractéristiques étaient semblables (le plus proche voisin) à celles de la personne pour laquelle les renseignements avaient été recueillis par procuration; puis, on a imputé à cette dernière les valeurs recueillies pour le « donneur ». On a trouvé le plus proche voisin à l'aide d'une fonction de distance précise qui utilisait des renseignements pertinents disponibles et sur le répondant par procuration et sur le répondant « donneur ». Là où on ne pouvait assurer la qualité des données par imputation, on a attribué le code manquant aux réponses.

Voici les modules entièrement imputés :

- La tension artérielle
- Les séances dentaires
- Les examens de la vue
- Le contact avec des professionnels de la santé mentale
- La dépendance à l'alcool
- La conduite en état d'ébriété
- Le soutien social
- La dépression
- Les pensées suicidaires et les tentatives de suicide
- Les comportements sexuels
- La consommation de fruits et de légumes

Voici les modules partiellement imputés :

- Le test de Papanicolaou (PAPA_020 seulement)
- Le test de dépistage de l'antigène prostatique (PSAA_170 seulement)

¹⁶ St-Pierre, M. et Béland, Y. (2002). *Imputation of proxy respondents in the Canadian Community Health Survey* – Recueil des présentations de la section des méthodes d'enquête, Société statistique du Canada; à paraître.

- La mammographie (MAMA_30, MAMA_37 et MAMA_38)
- La vaccination anti-grippale (FLUA_160 seulement)
- L'examen des seins (BRXA_110 seulement)
- L'auto-examen des seins (BSXA_120 seulement)
- La taille et le poids (HWTa_4 seulement)

Voici les modules, sautés durant l'interview par procuration, qui n'ont pas été imputés :

- L'examen médical complet
- Les aides au renoncement au tabac
- L'état général de santé
- L'estime de soi
- La maîtrise de la situation
- La spiritualité
- L'humeur
- La détresse
- Le stress au travail
- Les activités physiques
- Les activités sédentaires
- L'utilisation d'équipement de protection
- Les changements effectués pour améliorer la santé
- L'allaitement naturel
- La satisfaction du patient

7.5 Pondération

Le principe de base de l'estimation dans un échantillon aléatoire comme celui du cycle 1.1 de l'ESCC repose sur le fait que chaque personne représente, en plus d'elle-même, plusieurs autres personnes qui ne font pas partie de l'échantillon. Par exemple, dans un simple échantillon aléatoire 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 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 au section 8.

7.6 Élimination des renseignements confidentiels

Il convient de souligner que le fichier de microdonnées à grande diffusion décrit plus haut diffère, sous un nombre important d'aspects, du fichier maître de l'enquête tenu par Statistique Canada. Ces différences découlent des mesures prises pour protéger l'anonymat des répondants. La protection des répondants est assurée grâce à la suppression des valeurs individuelles, au regroupement et à l'établissement des valeurs extrêmes des variables. Les utilisateurs qui demandent l'accès à l'information non comprise sur le fichier de microdonnées à grande diffusion, ont trois choix: acheter des tableaux personnalisés, utiliser un des Centres de données de recherche¹⁷, ou utiliser le service de télé-accès. (Voir Section 12.3)

¹⁷ L'information la plus récente sur les Centres de données de recherche se retrouver à www.statcan.ca

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 comprise dans l'échantillon final, c'est-à-dire, dans l'échantillon de personnes ayant répondu à l'enquête. Ce poids correspond au nombre de personnes représentées par le répondant dans l'ensemble de la population.

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

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

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

Diagramme A: Sommaire de la stratégie de pondération

Base aréolaire	Base téléphonique
A0 - Poids initial	T0 - Poids initial
A1 - Accroissement de l'échantillon	T1 - Couverture de la base liste
A2 - Stabilisation	T2 - Nombre de mois
A3 - Retrait des unités hors champ	T3 - Retrait des unités hors champ
A4 - Non-réponse ménage	T4 - Combinaison des bases CA et liste
A5 - Création du poids-personne	T5 - Non-réponse ménage
A6 - Non-réponse personne	T6 - Ménages sans téléphone
Poids final de la base aréolaire	T7 - Création du poids-personne
↗	T8 - Non-réponse personne
	T9 - Lignes multiples
	Poids final de la base téléphonique
	↘
I1 - Intégration	
I2 - Effet saisonnier	
I3 - Poststratification	
Poids final du cycle 1.1 de l' ESCC	

8.1 Pondération du fichier

Tel que mentionné auparavant, les unités des bases aréolaire et téléphonique sont traitées séparément jusqu'à l'étape d'intégration (I1). La sous-section 8.1.1 fournit les détails de la stratégie de pondération pour la base aréolaire, puis la sous-section 8.1.2, ceux pour la base téléphonique. L'intégration des deux bases est traitée en 8.1.3, puis suivent les deux étapes finales de la pondération, c'est-à-dire l'ajustement pour contrôler la saisonnalité des données puis la poststratification, qui sont expliquées dans les sous-sections 8.1.4 et 8.1.5 respectivement.

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

8.1.1 Pondération de l'échantillon provenant de la base aréolaire

A0 – Poids initial

Puisque le mécanisme utilisé pour sélectionner l'échantillon de la base aréolaire a été celui établi pour l'Enquête sur la population active (EPA), le poids initial a dû être calculé selon les particularités de cette enquête. D'abord, à l'intérieur de chacune des strates définies par l'EPA, des grappes (unités primaires) sont sélectionnées avec probabilités proportionnelles à la taille (selon les comptes de recensement de 1991). À

l'intérieur de chacune des grappes sélectionnées, un échantillon de logements est ensuite choisi à l'aide d'un échantillonnage systématique. Le produit des probabilités de chacune de ces sélections représente la probabilité de sélection du logement et son inverse représente le poids initial du logement. Pour de détails sur le mécanisme de sélection, de même qu'une définition plus complète des strates et des grappes, se référer à Statistique Canada (1998)¹⁸.

A1 – Accroissement de l'échantillon

Certaines modifications ont dû être faites au mécanisme standard de l'EPA lors de la sélection de l'échantillon pour le cycle 1.1 de l'ESCC. Le plan de l'EPA peut fournir un échantillon d'environ 68 000 logements à l'échelle du pays alors que les besoins du cycle 1.1 de l'ESCC sont presque le double. Les modifications apportées afin de pouvoir obtenir la taille d'échantillon désirée ont été, en résumé, de répéter le même processus de sélection une deuxième fois à l'intérieur de toutes les grappes sélectionnées dans la RSS où le besoin en échantillon se faisait sentir. Ceci a eu l'effet d'accroître l'échantillon et on a dû en tenir compte dans la pondération afin de correctement représenter la probabilité de sélection. Un facteur d'ajustement représentant le taux d'accroissement de l'échantillon a donc été calculé. Le poids initial calculé en A0 est multiplié par ce facteur d'ajustement, ce qui permet d'obtenir le poids A1.

A2 – Stabilisation

Dans certaines RSS, l'accroissement de l'échantillon tel que décrit au paragraphe précédent résultait en un échantillon beaucoup plus grand que nécessaire. Une stabilisation a donc été instaurée afin de ramener la taille de l'échantillon au niveau désiré. Le processus de stabilisation consistait à sous-échantillonner des logements aléatoirement à l'intérieur de la RSS. Un facteur d'ajustement représentant l'effet de la stabilisation a donc été calculé afin 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. Ces derniers conservent donc le même poids qu'à l'étape précédente que l'on appelle maintenant poids A3.

¹⁸ Statistique Canada (1998). *Méthodologie de l'enquête sur la population active du Canada*. Statistique Canada., numéro 71-526-XPB au catalogue.

A4 – Non-réponse ménage

Lors de la collecte, une certaine proportion des ménages interviewés a inévitablement résulté en non-réponse. Ceci survient habituellement lorsque le ménage refuse de participer à l'enquête, fournit des données inutilisables, ou encore, ne peut être rejoint pour réaliser l'interview. Le poids des ménages non-répondants est redistribué aux répondants à l'aide de classes de réponse. L'algorithme CHAID (Chi-Square Automatic Interaction Detector), disponible dans Knowledge Seeker¹⁹, permet d'identifier les caractéristiques qui divisent le mieux l'échantillon en groupes selon leurs propensions à répondre. Noter que ces groupes sont formés indépendamment à l'intérieur de chaque RSS. Puisque l'information disponible auprès des non-répondants est très limitée, seules quelques caractéristiques telles que la période de collecte et un indicateur du milieu rural/urbain ont pu être utilisées pour la création des classes. Il s'est en fait révélé que seule la caractéristique période de collecte (avec 5 périodes; sept. à nov. 2000 / déc. 2000 à fév. 2001 / mars à mai 2001 / juin à août 2001 / sept. à oct. 2001) était significative pour la création des classes pour chacune des RSS. Un facteur d'ajustement a donc été calculé à l'intérieur de chaque classe de la façon suivante:

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

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

A5 – Création du poids-personne

Puisque l'unité d'échantillonnage finale pour l'ESCC est la personne, le poids-ménage calculé jusqu'ici doit être converti en un poids-personne. Le facteur calculé à cette étape incorpore également l'ajustement qui doit être fait pour tenir compte qu'une ou deux personnes ont pu être sélectionnées dans chaque ménage. Pour ce faire, le facteur d'ajustement est basé sur le nombre de personnes présentes dans le ménage (cette information est recueillie lors du listage des membres du ménage), la distribution de ce nombre dans les groupes d'âge 12-19 et 20+, puis le nombre de personnes sélectionnées parmi ceux-ci. Consultez le tableau 5.3 pour obtenir plus de détails sur l'algorithme utilisé par l'ESCC pour déterminer le nombre de personnes à sélectionner à l'intérieur d'un ménage.

Pour les personnes sélectionnées provenant de ménages où seulement une personne était sélectionnée, le facteur d'ajustement consiste simplement au nombre de personnes dans le ménage. Pour les cas où une personne dans le groupe d'âge 12-19,

¹⁹ ANGROSS Software (1995). Knowledge Seeker IV for Windows - User's Guide. ANGROSS Software International Limited.

et une personne dans le groupe d'âge 20+ étaient sélectionnées, les ajustements sont respectivement le nombre de personnes dans le ménage dans le groupe d'âge 12-19 et le groupe d'âge 20+. Finalement, pour les cas où deux personnes étaient sélectionnées sans aucune contrainte sur l'âge (mais toujours 12+), le facteur d'ajustement de chaque personne est la moitié du nombre de personnes de 12 ans et plus dans le ménage. Le poids-ménage obtenu en A4 est alors multiplié par le facteur d'ajustement dérivé en A5 pour produire le poids-personne A5.

A6 – Non-réponse personne

Dans le cadre du cycle 1.1 de l'ESCC, une interview peut être vue comme un processus en deux étapes. Dans un premier temps, l'intervieweur obtient la liste complète des personnes vivant dans le ménage, puis par la suite interviewe la(les) personne(s) sélectionnée(s) dans le ménage. Dans certains cas, les intervieweurs ne réussissent qu'à compléter la première étape, soit parce qu'ils ne peuvent entrer en contact avec la personne sélectionnée, ou encore parce que la personne sélectionnée refuse d'être interviewée. De tels cas sont définis comme étant des non-réponses à l'échelle de la personne, et un facteur d'ajustement doit être appliqué aux poids des personnes répondantes pour compenser cette non-réponse. Tout comme pour la non-réponse à l'échelle du ménage, l'ajustement est appliqué à l'intérieur de classes définies à partir des caractéristiques disponibles pour les répondants et non-répondants. Toutes les caractéristiques recueillies lors du listage des membres du ménage étaient en fait disponibles pour créer ces classes. L'algorithme CHAID a encore une fois été utilisé pour obtenir la définition des classes et le résultat final présente quelques variations dans la définition des classes d'une RSS à l'autre. Selon la RSS, les caractéristiques suivantes ont pu être utilisées pour former les classes d'ajustement : le sexe, le groupe d'âge, l'indicateur de milieu rural/urbain, l'éducation, l'état matrimonial et le nombre de personnes sélectionnées dans le ménage. Un facteur d'ajustement est calculé à l'intérieur de chaque classe de la façon suivante:

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

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

Étant donné que cet ajustement est le dernier nécessaire pour l'échantillon provenant de la base aréolaire, le poids A6 représente donc le **poids final de la base aréolaire**. Ce poids sera plus tard intégré au poids final de la base téléphonique pour créer le poids final du cycle 1.1 de l'ESCC.

8.1.2 Pondération de l'échantillon provenant de la base téléphonique

Tel que mentionné précédemment, la base téléphonique est en fait composée de deux bases : la base de composition aléatoire (CA), puis une base liste de numéros de téléphone. Les unités provenant de ces deux bases sont toutefois traitées ensemble et sont donc toutes soumises aux mêmes ajustements. Il existe toutefois deux exceptions; d'abord, puisque la probabilité de sélection est relative à la base utilisée pour faire la sélection, cette probabilité sera légèrement différente dépendamment que l'unité provienne de la base CA ou de la base liste. L'autre exception implique l'ajustement T1. Les détails de ces deux exceptions sont donnés dans les sous-sections réservées aux deux ajustements impliqués.

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

T0 – Poids initial

Le poids initial est calculé quelque peu différemment selon que l'échantillon provienne de la base CA ou de la base liste. Dans les deux cas, le poids initial est défini comme étant l'inverse de la probabilité de sélection, mais puisque les méthodes de sélection diffèrent, les probabilités diffèrent aussi. Pour la base CA, la sélection des numéros est faite à l'intérieur de chaque strate CA. Une strate CA représente un agrégat d'indicatifs régionaux et préfixes (IRP; les six premiers chiffres du numéro à 10 chiffres), contenant chacune des banques valides de cent numéros (voir Norris et Paton²⁰ 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. L'inverse de ces probabilités de sélection représente le poids initial T0.

²⁰ 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 – Couverture de la base liste

Puisque la base liste ne couvre pas certains numéros de téléphone qui sont toutefois couverts par la base CA, un ajustement doit être apporté au poids initial des unités de la base liste pour que les deux bases soient comparables en ce qui a trait à la couverture. Cet ajustement consiste à gonfler le poids des unités de la base liste proportionnellement au taux de couverture dans chaque RSS. L'estimation de ce taux de couverture a été la tâche la plus ardue, et a pu être faite à l'aide des données recueillies auprès de l'échantillon de la base aréolaire. En effet, le questionnaire utilisé pour l'interview des personnes sélectionnées par la base aréolaire incluait un ensemble de questions vérifiant la présence d'un téléphone dans le logement du répondant, le nombre de lignes utilisées à des fins personnelles, puis le numéro pour chacune de ces lignes. Pour dériver le taux de couverture désiré, on a simplement calculé le pourcentage des numéros de téléphone recueillis étant présents sur la base liste. L'inverse de ce taux représente le facteur utilisé pour cet ajustement. Le facteur, une fois multiplié par le poids initial T0, produit le poids T1.

T2 – Nombre de mois

Contrairement à la base aréolaire pour laquelle l'échantillon a été sélectionné entièrement au début du processus d'échantillonnage, des échantillons ont été tirés à chaque mois pour la base téléphonique. À chacun de ces échantillons mensuels correspond un poids initial faisant en sorte que chaque échantillon soit représentatif de la RSS. Toutefois, pour que l'échantillon total ne représente qu'une seule fois la population, un facteur d'ajustement a dû être appliqué pour réduire les poids de chaque échantillon mensuel. Le facteur d'ajustement appliqué à chaque échantillon mensuel était égal à la proportion que représentait cet échantillon mensuel parmi l'échantillon total. Noter que cet ajustement a été fait séparément pour les bases CA et liste, ce qui fait en sorte que l'échantillon provenant de chacune de ces deux bases représente la population totale. Pour remédier à cette situation, les échantillons CA et liste sont plus tard combinés (à l'étape T4) de façon à ce que l'échantillon total de la base téléphonique ne représente qu'une seule fois la population totale. Le poids T2 est donc obtenu en multipliant le poids T1 par le facteur d'ajustement défini ci-dessus.

T3 - 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 T3.

T4 - Combinaison des bases CA et liste

Tel que mentionné à l'ajustement T2, les bases CA et liste représentent jusqu'à cette étape, chacune la population totale dans les RSS où elles sont utilisées. Afin d'unir les unités de ces deux bases pour qu'elles représentent ensemble la population totale, on applique un facteur d'ajustement. Ce facteur est uniquement basé sur les tailles d'échantillon observées dans chaque base. Pour les unités de la base CA, le facteur représente la proportion de l'échantillon total de la base téléphonique provenant de la base CA. Le complément de cette proportion représente le facteur utilisé pour les unités de la base liste. Ces facteurs sont calculés et appliqués indépendamment dans chaque RSS où les deux bases téléphoniques ont été utilisées. Ainsi, le poids T4 est obtenu en multipliant le poids T3 par le facteur de combinaison.

T5 - Non-réponse ménage

L'ajustement fait ici pour compenser l'effet de la non-réponse ménage est identique à celui appliqué pour la base aréolaire (ajustement A4). Comme c'était le cas pour A4, la seule caractéristique significative pour expliquer la non-réponse était la période de collecte, qui a donc 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:

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

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

T6 - Ménages sans téléphone

Une certaine proportion de la population canadienne n'a pas accès à une ligne téléphonique résidentielle privée. Tel qu'expliqué à l'étape T1, de l'information concernant la présence d'un téléphone dans le logement du répondant est recueillie auprès de l'échantillon de la base aréolaire. Cette information a été utilisée pour estimer la proportion de ménages n'ayant pas le téléphone à l'échelle de chaque RSS. Tout comme pour T1, cette proportion est ensuite utilisée pour gonfler le poids des unités de la base téléphonique, ajustant ainsi pour la sous-représentation de la base due à cette sous-population non couverte. Le facteur utilisé pour cet ajustement correspond à l'inverse de la proportion estimée, et une fois multiplié par le poids T5, procure le poids T6.

T7 – Création du poids-personne

Tout comme l'ajustement A5, cet ajustement permet de convertir ce qui était jusqu'à cette étape-ci un poids-ménage en un poids-personne. Puisque contrairement à la base aréolaire, une seule personne est choisie par ménage pour la base téléphonique, le

facteur d'ajustement est relativement simple; il représente simplement le nombre de personnes dans le champ de l'enquête à l'intérieur du ménage de la personne sélectionnée. Ce facteur multiplié par le poids T6 donne le poids T7.

T8 - Non-réponse personne

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

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

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

T9 - Lignes multiples

Le fait que certains ménages possèdent plus d'une ligne téléphonique résidentielle a un impact sur la pondération; plus le ménage a de lignes, meilleure est sa probabilité d'être sélectionné. Conséquemment, les poids doivent être ajustés pour tenir compte du nombre de lignes résidentielles que le ménage possède. Noter que cette information est recueillie dans la portion initiale de l'interview auprès de la personne sélectionnée. Le facteur d'ajustement représente donc l'inverse du nombre de lignes. Le poids T9 est donc obtenu en multipliant ce facteur par le poids T8.

Puisque cet ajustement est le dernier nécessaire pour l'échantillon provenant de la base téléphonique, le poids T9 représente donc le **poids final de la base téléphonique**. Ce poids sera par la suite, à l'étape I1, intégré au poids final de la base aréolaire pour finalement créer le poids final du cycle 1.1 de l'ESCC.

8.1.3 Intégration des bases aréolaire et téléphonique (I1)

Cette étape consiste à intégrer les poids finaux des échantillons aréolaire et téléphonique créés jusqu'à maintenant, en un seul poids en appliquant une méthode d'intégration²¹. Un facteur d'ajustement, compris entre 0 et 1, est déterminé de façon à représenter l'importance relative de chaque échantillon dans l'échantillon total. Cette importance relative est mesurée en termes de taille d'échantillon et d'effet de plan.

²¹ Skinner, C.J. and Rao, J.N.K. (1996). Estimation in Dual Frame Surveys with Complex Designs. *Journal of the American Statistical Association*, 91, 433, 349-356.

Plus la proportion d'échantillon qu'une base représente dans l'échantillon total est grande, plus grande sera son importance relative dans l'échantillon totale. Pour ce qui de l'effet de plan, l'importance relative sera plus grande pour les unités provenant de la base dont l'effet de plan est plus petit. Pour obtenir le facteur d'ajustement d'intégration, on calcule d'abord un facteur α , obtenu de la façon suivante:

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

où n_A et n_T représentent respectivement les tailles d'échantillon des bases aréolaire et téléphonique, alors que R représente le rapport médian des effets de plan observés pour chacune des deux bases. Le poids des unités de la base aréolaire est multiplié par ce facteur α , alors que le poids des unités de la base téléphonique est multiplié par $1-\alpha$. Noter que dans les cas où une RSS n'est couverte que par une seule base, le facteur d'ajustement est égal à 1. Le produit du facteur d'ajustement dérivé ici, par le poids final calculé auparavant (A_6 ou T_9 dépendant de quelle base provient l'unité), procure le poids intégré I_1 .

8.1.4 Effet saisonnier (I_2)

L'ESCC avait initialement planifié répartir la collecte des données également sur les douze mois de l'année de référence de l'enquête afin de contrôler entre autres l'effet saisonnier des données recueillies. Certains événements ont toutefois affecté ce plan, de sorte qu'un ajustement additionnel a dû être ajouté pour assurer qu'il n'y ait d'effet saisonnier dans les estimations produites à l'aide des données du cycle 1.1 de l'ESCC²². L'ajustement appliqué en I_2 a été fait de façon à ce que la somme des poids des unités interviewées lors d'une des quatre saisons, représente exactement 25 % de la somme des poids de l'échantillon total. Bref, après l'application de cet ajustement, la portion d'échantillon interviewée à chaque saison représente 25 % de la population totale de chaque RSS.

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

$$\frac{\text{Somme des poids } I_1 \text{ pour l'échantillon total}}{4 \times \text{somme des poids } I_1 \text{ de l'échantillon interviewé lors de la saison } S}$$

Cet ajustement saisonnier appliqué au poids I_1 permet d'obtenir le poids I_2 .

²² Béland, Y. , Dufour, J. and Hamel, M. (2001). Faire échec à la non-réponse dans le cadre de l'Enquête sur la santé dans les collectivités canadiennes, *Recueil des présentations effectuées au Symposium 2001 organisé par Statistique Canada*, Statistique Canada.

8.1.5 Poststratification (I3)

La dernière étape nécessaire afin d'obtenir le poids final du cycle 1.1 de l'ESCC est la poststratification. La poststratification est appliquée afin d'assurer que la somme des poids finaux corresponde aux estimations de populations définies à l'échelle des RSS, pour chacun des 10 groupes d'âge-sexe d'intérêt, c'est-à-dire les cinq groupes d'âges 12-19, 20-29, 30-44, 45-64, 65+, pour chacun des deux sexes. Noter que pour la Colombie-Britannique, la post stratification a été faite en utilisant une géographie révisée contenant seize régions au lieu des vingt utilisés initialement lors de la planification et du déroulement de l'enquête.

Les estimations de population utilisées sont basées sur les comptes du Recensement de 1996, de même que sur les estimations des comptes de naissance, décès, immigration et émigration. La moyenne des estimations mensuelles pour chacun des croisements RSS-âge-sexe a été retenue pour calculer la poststratification. Le poids I2 a donc été ajusté afin d'obtenir le poids final I3, à l'aide du facteur d'ajustement I3 défini comme suit:

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

Le poids I3 correspond au **poids final du cycle 1.1 de l'ESCC** que l'on retrouve dans le fichier de données portant le nom de variable WTSAM.

8.1.6 Particularités de la pondération pour les trois territoires

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

D'abord pour la base aréolaire, tel que mentionné à la sous-section 5.4.1, une étape additionnelle de sélection a été ajoutée pour les territoires. Chaque territoire était initialement stratifié selon des regroupements de communautés à l'intérieur desquels on a sélectionné aléatoirement une communauté. Noter que les capitales de chaque territoire formaient une strate à elles-seules, et étaient donc toutes trois sélectionnées automatiquement à cette première sélection. Cette particularité n'a eu d'effet que dans le calcul de la probabilité de sélection, et donc dans la valeur du poids initial (A0). Une fois ce poids initial calculé, la même série d'ajustements (A1 à A6) a été appliquée aux unités de la base aréolaire. Les classes d'ajustement pour les non-réponses ménage et personne ont été construites à l'aide du même ensemble de variables disponibles pour les provinces. Seule la définition des périodes de collecte a été modifiée pour mieux refléter le déroulement de la collecte qui a débuté dans les territoires en novembre 2000. Les quatre périodes utilisées étaient définies comme

étant novembre 2000 à février 2001, mars à mai 2001, juin à août 2001, puis septembre à octobre 2001.

Pour ce qui est de la pondération des unités de la base téléphonique, mentionnons tout d'abord que seule la base CA a été utilisée, et ce, uniquement à l'intérieur des capitales. Ceci élimine donc le besoin d'avoir recours aux ajustements T1 (couverture de la base liste) et T4 (combinaison des bases CA et liste). Les autres ajustements ont tous été appliqués. Tout comme pour la partie aréolaire, la définition des périodes de collecte a été modifiée afin de faire les ajustements de non-réponse. Finalement, l'ajustement T6 (ménages sans téléphone) a aussi subi une légère modification puisque la base CA était utilisée uniquement dans les communautés autres que les capitales. Les proportions de ménages sans téléphone ont été dérivées, tout comme pour les provinces, à partir des données de la base aréolaire, mais en excluant toutefois du calcul les données des ménages situés dans les capitales.

Les deux ensembles de poids (aréolaire et téléphonique) ont ensuite été intégrés, puis ajustés pour la saisonnalité et finalement poststratifiés de façon semblable à ce qui a été fait pour les provinces, à l'exception de deux détails. D'abord, l'intégration a été appliquée uniquement pour les unités situées dans les capitales; les autres communautés ayant été couvertes uniquement par la base téléphonique. Le second détail a trait à la saisonnalité. Étant donné qu'une forte concentration des interviews a été menée sur une très courte période de temps dans le territoire du Nunavut, l'ajustement pour la saisonnalité n'a pu être appliqué de façon efficace. Les estimations produites pour le Nunavut à partir de ces poids ne tiendront donc pas compte d'un possible effet saisonnier des données.

9. Qualité des données

9.1 Taux de réponse

Au total et après avoir retiré les unités hors du champ de l'enquête, 136 937 ménages ont été sélectionnés pour participer à l'ESCC. De ce nombre, 125 159 ont accepté de participer à l'enquête ce qui résulte en un taux de réponse à l'échelle du ménage de 91,4 %. Parmi ces ménages répondants, 142 421 personnes ont été sélectionnées pour participer à l'enquête parmi lesquelles 130 827 ont accepté ce qui résulte en un taux de réponse à l'échelle de la personne de 91,9 %. À l'échelle canadienne, un taux de réponse combiné de **84,7 %** a donc été observé à l'ESCC. Il est important de mentionner que le taux de réponse combiné n'est pas obtenu en multipliant les taux de réponse aux échelles du ménage et de la personne car il y a eu une sélection de deux personnes dans certains ménages. Le tableau 9.1 donne les taux de réponse combinés ainsi que l'information pertinente au calcul de ceux-ci pour chaque région socio-sanitaire ou regroupement de régions socio-sanitaires.

La collecte des données du cycle 1.1 de l'ESCC en Colombie-Britannique s'est déroulée en 2000-2001 en utilisant les limites géographiques des 20 régions socio-sanitaires en vigueur à ce moment-là. L'année suivante, le gouvernement de la Colombie-Britannique redéfinissait les limites de leurs régions socio-sanitaires. Les limites pour les 16 nouvelles régions se retrouvent sur ce fichier de microdonnées à grande diffusion. Il n'est toutefois pas approprié de diffuser les taux de réponse pour ces nouvelles régions.

Il est également important de noter qu'il peut exister des différences entre les nombres inscrits au tableau 9,1 et le nombre d'enregistrements se trouvant dans le FMGD pour les autres provinces et territoires. Les taux de réponse présentés sont fondés sur la géographie définie par le plan de sondage et certaines unités pourraient s'être retrouvées dans une autre région au cours du traitement des données.

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

Taux de réponse à l'échelle du ménage

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

Taux de réponse à l'échelle de la personne parmi les ménages où 2 personnes ont été sélectionnées (base aréolaire seulement)

$$\text{PPRR/A2} = \frac{\text{\# de répondants parmi les ménages où 2 personnes ont été sélectionnées}}{\text{toutes les personnes faisant partie du champ de l'enquête parmi les ménages où 2 personnes ont été sélectionnées}}$$

Taux de réponse à l'échelle de la personne parmi les ménages où 1 personne a été sélectionnée (base aréolaire et bases téléphoniques)

$$\text{PPRR/A1,PPRR/P1} = \frac{\text{\# de répondants parmi les ménages où 1 personne a été sélectionnée}}{\text{toutes les personnes faisant partie du champ de l'enquête parmi les ménages où 1 personne a été sélectionnée}}$$

Afin de calculer adéquatement un taux de réponse combiné, l'utilisateur devra également calculer les trois rapports suivants (ces rapports représentent "l'importance", à l'échelle du ménage, de chaque composante dans le taux de réponse combiné).

Rapport pour les ménages où 2 personnes ont été sélectionnées (base aréolaire)

$$\text{R/A2} = \frac{\text{\# de ménages répondants parmi ceux où 2 personnes ont été sélectionnées}}{\text{tous les ménages répondants provenant des deux bases}}$$

Note: il est important de mentionner que le " # de ménages répondants parmi ceux où 2 personnes ont été sélectionnées " est obtenu en divisant par 2 le nombre de personnes sélectionnées parmi les ménages de la base aréolaire où 2 personnes ont été sélectionnées du tableau 9.1.

Rapport pour les ménages où 1 personne a été sélectionnée (base aréolaire)

$$\text{R/A1} = \frac{\text{\# de ménages répondants parmi ceux où 1 personne a été sélectionnée}}{\text{tous les ménages répondants provenant des deux bases}}$$

Note: le " # de ménages répondants parmi ceux où 1 personne a été sélectionnée " de la base aréolaire est ni plus ni moins que le nombre de personnes sélectionnées parmi les ménages de la base aréolaire où 1 personne a été sélectionnée du tableau 9.1.

Rapport pour les ménages provenant des bases téléphoniques

$$\text{R/P} = \frac{\text{\# de ménages répondants provenant des bases téléphoniques}}{\text{tous les ménages répondants provenant des deux bases}}$$

Une fois que toutes les composantes ci-haut ont été calculées, un utilisateur peut calculer le taux de réponse combiné en utilisant la formule suivante.

Taux de réponse combiné

$$\text{COMB/RR} = \text{HHRR} * [(R/A2 * \text{PPRR}/A2) + (R/A1 * \text{PPRR}/A1) + (R/P * \text{PPRR}/P1)]$$

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

$$\text{HHRR} = \frac{100\,396 + 24\,763}{109\,315 + 27\,622} = \frac{125\,159}{136\,937} = 0,914$$

$$\text{PPRR}/A2 = \frac{29\,777}{34\,524} = 0,863$$

$$\text{PPRR}/A1 = \frac{78\,129}{83\,134} = 0,940$$

$$\text{PPRR}/P1 = \frac{22\,921}{24\,763} = 0,926$$

$$R/A2 = \frac{(34\,524 \div 2)}{100\,396 + 24\,763} = \frac{17\,262}{125\,159} = 0,138$$

$$R/A1 = \frac{83\,134}{100\,396 + 24\,763} = \frac{83\,134}{125\,159} = 0,664$$

$$R/P = \frac{24\,763}{100\,396 + 24\,763} = \frac{24\,763}{125\,159} = 0,198$$

$$\begin{aligned} \text{alors } \text{COMB/RR} &= 0,914 * [(0,138 * 0,863) + (0,664 * 0,940) + (0,198 * 0,926)] \\ &= 0,914 * [0,1191 + 0,6242 + 0,1833] \\ &= 0,8469 \\ &= \mathbf{84,7 \%}. \end{aligned}$$

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

Table 9.1 Tableau 9.1		Area Frame / Base aréolaire										Phone Frames / Bases téléphoniques						All cases Tous les cas
					2 persons select. 2 personnes sélect.			1 person select. 1 personne sélect.										
Prov.	Health Region	# In Scope HH	# Resp HH	HH Resp. Rates	# Pers. Select.	# Resp. Rates	Pers. Resp. Rates	# Pers. Select.	# Resp. Rates	Pers. Resp. Rates	Resp. Rates	# In Scope HH	# Resp HH	HH Resp. Rates	# Resp. Rates	Pers. Resp. Rates	Resp. Rates	Combined Resp. Rates
Terr.	Région socio- sanitaire	# Mén. cibles	# Mén. rép.	Taux de rép. mén.	# Pers. sélect.	# Rép.	Taux de rép. pers.	# Pers. sélect.	# Rép.	Taux de rép. pers.	Taux de rép.	# Mén. cibles	# Mén. rép.	Taux de rép. mén.	# Rép.	Taux de rép. pers.	Taux de rép.	Taux de rép. Combiné
CA	Total	109315	100396	91.8	34524	29777	86.3	83134	78129	94	85.1	27622	24763	89.6	22921	92.6	83.0	84.7
NL	Total	3453	3267	94.6	1384	1211	87.5	2575	2385	92.6	86.6	307	295	96.1	274	92.9	89.3	86.8
	10901	847	803	94.8	284	250	88	661	642	97.1	90.6	90.6
	10902	754	721	95.6	310	281	90.6	566	529	93.5	88.8	88.8
	10903	689	645	93.6	280	251	89.6	505	458	90.7	84.7	84.7
	10904*	1163	1098	94.4	510	429	84.1	843	756	89.7	83.4	307	295	96.1	274	92.9	89.3	84.7
PE	Total	1689	1581	93.6	552	483	87.5	1305	1239	94.9	87.7	2335	2074	88.8	1929	93	82.6	84.7
	11901	955	874	91.5	242	210	86.8	753	731	97.1	87.5	514	455	88.5	429	94.3	83.5	86.1
	11902	734	707	96.3	310	273	88.1	552	508	92	87.8	1821	1619	88.9	1500	92.6	82.4	83.9
NS	Total	4540	4265	93.9	1482	1313	88.6	3524	3373	95.7	88.8	709	670	94.5	633	94.5	89.3	88.8
	12901	856	809	94.5	270	244	90.4	674	651	96.6	90.3	113	106	93.8	105	99.1	92.9	90.6
	12902	406	393	96.8	122	109	89.3	332	312	94	90.3	306	286	93.5	267	93.4	87.3	89.0
	12903	742	711	95.8	248	229	92.3	587	570	97.1	92.3	92.3
	12904	654	614	93.9	242	214	88.4	493	469	95.1	88.1	3	3	100	3	100	100.0	88.1
	12905	554	526	94.9	194	160	82.5	429	405	94.4	87.5	287	275	95.8	258	93.8	89.9	88.3
	12906	1328	1212	91.3	406	357	87.9	1009	966	95.7	86.2	86.2
NB	Total	4743	4461	94.1	1534	1342	87.5	3694	3521	95.3	88.4	144	141	97.9	133	94.3	92.4	88.5
	13901	949	897	94.5	288	253	87.8	753	728	96.7	90.0	90.0
	13902	890	831	93.4	304	265	87.2	679	649	95.6	87.8	87.8
	13903	907	844	93.1	266	214	80.5	711	659	92.7	84.5	84.5

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

Table 9.1 Tableau 9.1		Area Frame / Base aréolaire										Phone Frames / Bases téléphoniques						All cases Tous les cas
Prov.	Health Region	# In Scope HH	# Resp HH	HH Resp. Rates	# Pers. Select.	# Resp. Rates	Pers. Resp. Rates	# Pers. Select.	# Resp. Rates	Pers. Resp. Rates	Resp. Rates	# In Scope HH	# Resp HH	HH Resp. Rates	# Resp. Rates	Pers. Resp. Rates	Resp. Rates	Combined Resp. Rates
Terr.	Région socio- sanitaire	# Mén. cibles	# Mén. rép.	Taux de rép. mén.	# Pers. sélect.	# Rép.	Taux de rép. pers.	# Pers. sélect.	# Rép.	Taux de rép. pers.	Taux de rép.	# Mén. cibles	# Mén. rép.	Taux de rép. mén.	# Rép.	Taux de rép. pers.	Taux de rép.	Taux de rép. Combiné
	13904*	925	871	94.2	330	309	93.6	706	682	96.6	90.4	81	78	96.3	73	93.6	90.1	90.4
	13905*	1072	1018	95	346	301	87	845	803	95	88.9	63	63	100	60	95.2	95.2	89.3
QC	Total	20999	19316	92	5656	4970	87.9	16488	15512	94.1	85.7	1773	1587	89.5	1477	93.1	83.3	85.5
	24901	1019	997	97.8	304	261	85.9	845	779	92.2	89.3	95	87	91.6	85	97.7	89.5	89.3
	24902	1117	1046	93.6	426	338	79.3	833	776	93.2	84.6	84.6
	24903	1765	1643	93.1	336	279	83	1475	1372	93	85.6	85.6
	24904	1553	1477	95.1	378	352	93.1	1288	1241	96.4	91.2	91.2
	24905	1208	1114	92.2	302	271	89.7	963	903	93.8	86.0	86.0
	24906	3090	2622	84.9	606	540	89.1	2319	2180	94	79.3	79.3
	24907	1204	1085	90.1	320	298	93.1	925	872	94.3	84.8	84.8
	24908	1061	991	93.4	326	305	93.6	828	804	97.1	90.2	172	160	93	151	94.4	87.8	89.8
	24909	875	796	91	278	256	92.1	657	626	95.3	86.2	278	256	92.1	243	94.9	87.4	86.5
	24911	840	809	96.3	256	229	89.5	681	622	91.3	87.7	399	373	93.5	351	94.1	88.0	87.8
	24912	1347	1268	94.1	374	347	92.8	1081	1054	97.5	91.1	10	8	80	8	100	80.0	91.0
	24913	870	773	88.9	244	170	69.7	651	530	81.4	70.7	432	377	87.3	341	90.5	78.9	73.4
	24914	1274	1213	95.2	388	345	88.9	1019	972	95.4	89.8	242	209	86.4	189	90.4	78.1	88.0
	24915	1308	1203	92	414	365	88.2	996	949	95.3	86.5	145	117	80.7	109	93.2	75.2	85.4
	24916	2468	2279	92.3	704	614	87.2	1927	1832	95.1	86.7	86.7
ON	Total	32024	29117	90.9	10398	8678	83.5	23918	22175	92.7	82.8	10593	9151	86.4	8425	92.1	79.5	82.0
	35926	646	607	94	184	163	88.6	515	488	94.8	88.2	224	181	80.8	162	89.5	72.3	84.1
	35927	583	509	87.3	182	147	80.8	418	385	92.1	78.6	282	240	85.1	222	92.5	78.7	78.7

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

Table 9.1 Tableau 9.1		Area Frame / Base aréolaire										Phone Frames / Bases téléphoniques						All cases Tous les cas
Prov.	Health Region	# In Scope HH	# Resp HH	HH Resp. Rates	# Pers. Select.	# Resp. Rates	Pers. Resp. Rates	# Pers. Select.	# Resp.	Pers. Resp. Rates	Resp. Rates	# In Scope HH	# Resp HH	HH Resp. Rates	# Resp.	Pers. Resp. Rates	Resp. Rates	Combined Resp. Rates
Terr.	Région socio- sanitaire	# Mén. cibles	# Mén. rép.	Taux de rép. mén.	# Pers. sélect.	# Rép.	Taux de rép. pers.	# Pers. sélect.	# Rép.	Taux de rép. pers.	Taux de rép.	# Mén. cibles	# Mén. rép.	Taux de rép. mén.	# Rép.	Taux de rép. pers.	Taux de rép.	Taux de rép. Combiné
	35930	1457	1313	90.1	496	388	78.2	1065	987	92.7	81.1	81.1
	35931	430	407	94.7	146	127	87	334	325	97.3	90.3	381	331	86.9	304	91.8	79.8	85.4
	35933	543	518	95.4	158	145	91.8	439	421	95.9	90.9	366	312	85.2	292	93.6	79.8	86.4
	35934	579	538	92.9	196	181	92.3	440	417	94.8	87.7	171	140	81.9	128	91.4	74.9	84.7
	35935	536	484	90.3	160	141	88.1	404	382	94.6	84.4	548	478	87.2	459	96	83.8	84.1
	35936	913	861	94.3	326	234	71.8	698	618	88.5	80.5	513	452	88.1	409	90.5	79.7	80.2
	35937	1478	1251	84.6	448	364	81.3	1027	935	91	75.6	42	28	66.7	27	96.4	64.3	75.3
	35938	307	287	93.5	82	65	79.3	246	229	93.1	85.2	702	633	90.2	583	92.1	83.0	83.7
	35939*	613	587	95.8	212	171	80.7	481	447	92.9	86.9	786	690	87.8	643	93.2	81.8	84.0
	35940	682	664	97.4	236	212	89.8	546	532	97.4	93.5	388	344	88.7	318	92.4	82.0	89.3
	35941	586	519	88.6	190	173	91.1	424	402	94.8	83.4	465	401	86.2	367	91.5	78.9	81.4
	35942	660	612	92.7	234	201	85.9	495	471	95.2	86.6	229	204	89.1	185	90.7	80.8	85.1
	35943	696	608	87.4	180	156	86.7	518	495	95.6	82.3	334	282	84.4	256	90.8	76.6	80.5
	35944	1413	1185	83.9	412	346	84	979	919	93.9	77.3	77.3
	35945	396	367	92.7	116	96	82.8	309	293	94.8	86.1	474	415	87.6	387	93.3	81.6	83.7
	35946	1364	1206	88.4	376	326	86.7	1018	946	92.9	81.3	81.3
	35947*	990	923	93.2	252	209	82.9	797	746	93.6	85.9	632	560	88.6	531	94.8	84.0	85.2
	35949	288	257	89.2	74	62	83.8	220	212	96.4	84.4	523	480	91.8	445	92.7	85.1	84.8
	35951	1878	1741	92.7	596	543	91.1	1443	1382	95.8	88.0	88.0
	35952	698	624	89.4	236	194	82.2	506	473	93.5	81.7	81	54	66.7	50	92.6	61.7	79.6
	35953	1579	1453	92	750	535	71.3	1078	980	90.9	79.0	425	366	86.1	316	86.3	74.4	78.0

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

Table 9.1 Tableau 9.1		Area Frame / Base aréolaire										Phone Frames / Bases téléphoniques						All cases Tous les cas
Prov.	Health Region	# In Scope HH	# Resp HH	HH Resp. Rates	# Pers. Select.	# Resp. Rates	Pers. Resp. Rates	# Pers. Select.	# Resp.	Pers. Resp. Rates	Resp. Rates	# In Scope HH	# Resp HH	HH Resp. Rates	# Resp.	Pers. Resp. Rates	Resp. Rates	Combined Resp. Rates
Terr.	Région socio- sanitaire	# Mén. cibles	# Mén. rép.	Taux de rép. mén.	# Pers. sélect.	# Rép.	Taux de rép. pers.	# Pers. sélect.	# Rép.	Taux de rép. pers.	Taux de rép.	# Mén. cibles	# Mén. rép.	Taux de rép. mén.	# Rép.	Taux de rép. pers.	Taux de rép.	Taux de rép. Combiné
	35955	229	199	86.9	44	31	70.5	177	159	89.8	76.2	801	708	88.4	660	93.2	82.4	81.0
	35956	363	333	91.7	128	99	77.3	269	249	92.6	82.2	468	421	90	397	94.3	84.8	83.7
	35957	656	615	93.8	188	172	91.5	521	504	96.7	89.9	72	43	59.7	40	93	55.6	86.5
	35958	631	580	91.9	214	197	92.1	473	453	95.8	87.4	460	379	82.4	345	91	75.0	82.2
	35960	1215	1145	94.2	438	378	86.3	926	851	91.9	85.6	146	111	76	100	90.1	68.5	83.8
	35961	957	883	92.3	268	229	85.4	749	701	93.6	85.2	71	50	70.4	47	94	66.2	83.9
	35962	858	823	95.9	270	253	93.7	688	659	95.8	91.6	65	49	75.4	44	89.8	67.7	89.9
	35965	1363	1247	91.5	462	366	79.2	1016	936	92.1	82.1	82.1
	35966	994	959	96.5	360	335	93.1	779	736	94.5	90.9	127	102	80.3	95	93.1	74.8	89.1
	35968	1257	1146	91.2	414	354	85.5	939	878	93.5	83.9	26	21	80.8	16	76.2	61.5	83.5
	35970	1044	975	93.4	542	481	88.7	704	660	93.8	86.3	791	676	85.5	597	88.3	75.5	81.6
	35995	3142	2691	85.6	828	604	72.9	2277	1904	83.6	70.2	70.2
MB	Total	7379	6960	94.3	2322	2111	90.9	5799	5586	96.3	90.0	909	824	90.6	773	93.8	85.0	89.5
	46910	2095	1923	91.8	552	498	90.2	1647	1567	95.1	86.7	86.7
	46915*	1475	1401	95	446	409	91.7	1178	1152	97.8	92.0	350	317	90.6	308	97.2	88.0	91.2
	46920*	897	874	97.4	354	321	90.7	697	674	96.7	93.0	339	310	91.4	283	91.3	83.5	90.4
	46930	685	632	92.3	252	222	88.1	506	490	96.8	87.7	57	53	93	47	88.7	82.5	87.3
	46940	772	743	96.2	268	237	88.4	609	585	96.1	91.1	91.1
	46960*	1455	1387	95.3	450	424	94.2	1162	1118	96.2	91.4	163	144	88.3	135	93.8	82.8	90.5
SK	Total	6885	6379	92.7	2200	1912	86.9	5279	5037	95.4	87.0	1242	1154	92.9	1060	91.9	85.3	86.8
	47901*	1638	1514	92.4	506	452	89.3	1261	1182	93.7	86.0	249	227	91.2	213	93.8	85.5	85.9

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

Table 9.1 Tableau 9.1		Area Frame / Base aréolaire										Phone Frames / Bases téléphoniques						All cases Tous les cas
Prov.	Health Region	# In Scope HH	# Resp HH	HH Resp. Rates	# Pers. Select.	# Resp. Rates	Pers. Resp. Rates	# Pers. Select.	# Resp.	Pers. Resp. Rates	Resp. Rates	# In Scope HH	# Resp HH	HH Resp. Rates	# Resp.	Pers. Resp. Rates	Resp. Rates	Combined Resp. Rates
Terr.	Région socio- sanitaire	# Mén. cibles	# Mén. rép.	Taux de rép. mén.	# Pers. sélect.	# Rép.	Taux de rép. pers.	# Pers. sélect.	# Rép.	Taux de rép. pers.	Taux de rép.	# Mén. cibles	# Mén. rép.	Taux de rép. mén.	# Rép.	Taux de rép. pers.	Taux de rép.	Taux de rép. Combiné
	47904	1187	1080	91	382	328	85.9	889	841	94.6	84.7	84.7
	47905*	1158	1094	94.5	362	335	92.5	913	899	98.5	92.1	182	165	90.7	157	95.2	86.3	91.3
	47906	1261	1163	92.2	408	352	86.3	959	919	95.8	86.8	86.8
	47907*	995	934	93.9	354	286	80.8	757	706	93.3	85.3	279	260	93.2	246	94.6	88.2	85.9
	47909*	646	594	92	188	159	84.6	500	490	98	88.2	532	502	94.4	444	88.4	83.5	86.0
AB	Total	10961	10201	93.1	3776	3200	84.7	8313	7735	93	85.2	4148	3844	92.7	3521	91.6	84.9	85.1
	48901	858	807	94.1	242	215	88.8	686	663	96.6	89.8	12	11	91.7	10	90.9	83.3	89.7
	48902	539	510	94.6	196	163	83.2	412	397	96.4	88.8	202	179	88.6	166	92.7	82.2	87.0
	48903*	1017	950	93.4	348	297	85.3	776	728	93.8	86.2	376	338	89.9	307	90.8	81.6	85.0
	48904	2147	1956	91.1	680	579	85.1	1616	1505	93.1	83.6	83.6
	48906	926	882	95.2	288	265	92	738	706	95.7	90.6	90.6
	48907	691	650	94.1	264	227	86	518	471	90.9	84.6	114	110	96.5	105	95.5	92.1	85.7
	48908*	815	759	93.1	266	231	86.8	626	539	86.1	80.3	557	515	92.5	469	91.1	84.2	81.9
	48910	2097	1932	92.1	666	574	86.2	1599	1503	94	85.4	85.4
	48911	386	363	94	178	147	82.6	274	258	94.2	85.9	525	485	92.4	440	90.7	83.8	84.7
	48912	483	457	94.6	216	156	72.2	349	294	84.2	77.0	438	411	93.8	377	91.7	86.1	81.3
	48913	472	446	94.5	184	146	79.3	354	339	95.8	87.3	368	344	93.5	317	92.2	86.1	86.8
	48914*	530	489	92.3	248	200	80.6	365	332	91	81.5	1556	1451	93.3	1330	91.7	85.5	84.5
BC	Total	14329	12872	89.8	4254	3749	88.1	10745	10150	94.5	83.9	5077	4716	92.9	4403	93.4	86.7	84.7
	59901	295	288	97.6	94	86	91.5	241	239	99.2	95.6	382	354	92.7	331	93.5	86.6	90.5
	59902	304	286	94.1	78	73	93.6	247	236	95.5	89.6	458	427	93.2	400	93.7	87.3	88.3

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

Table 9.1 Tableau 9.1		Area Frame / Base aréolaire										Phone Frames / Bases téléphoniques						All cases Tous les cas
Prov.	Health Region	# In Scope HH	# Resp HH	HH Resp. Rates	# Pers. Select.	# Resp. Rates	Pers. Resp. Rates	# Pers. Select.	# Resp. Rates	Pers. Resp. Rates	Resp. Rates	# In Scope HH	# Resp HH	HH Resp. Rates	# Resp. Rates	Pers. Resp. Rates	Resp. Rates	Combined Resp. Rates
Terr.	Région socio- sanitaire	# Mén. cibles	# Mén. rép.	Taux de rép. mén.	# Pers. sélect.	# Rép.	Taux de rép. pers.	# Pers. sélect.	# Rép.	Taux de rép. pers.	Taux de rép.	# Mén. cibles	# Mén. rép.	Taux de rép. mén.	# Rép.	Taux de rép. pers.	Taux de rép.	Taux de rép. Combiné
	59903	469	436	93	88	80	90.9	392	365	93.1	86.4	508	474	93.3	446	94.1	87.8	87.1
	59904	986	923	93.6	300	273	91	773	740	95.7	88.9	49	47	95.9	45	95.7	91.8	89.0
	59905	487	452	92.8	164	155	94.5	370	353	95.4	88.4	545	510	93.6	485	95.1	89.0	88.7
	59906	1059	986	93.1	380	349	91.8	796	766	96.2	88.8	88.8
	59907	1424	1273	89.4	552	482	87.3	997	948	95.1	83.5	83.5
	59908	1062	957	90.1	316	257	81.3	799	732	91.6	81.0	199	182	91.5	170	93.4	85.4	81.7
	59909	201	178	88.6	38	35	92.1	159	148	93.1	82.3	513	482	94	443	91.9	86.4	85.2
	59910	987	894	90.6	270	248	91.9	759	730	96.2	86.5	106	104	98.1	96	92.3	90.6	86.9
	59911	470	431	91.7	128	115	89.8	367	354	96.5	87.6	320	299	93.4	283	94.6	88.4	87.9
	59912	308	285	92.5	108	96	88.9	231	223	96.5	88.0	394	364	92.4	350	96.2	88.8	88.5
	59913*	1092	982	89.9	388	344	88.7	788	742	94.2	83.7	197	187	94.9	173	92.5	87.8	84.3
	59915	662	581	87.8	176	139	79	493	463	93.9	80.4	284	265	93.3	255	96.2	89.8	83.2
	59916	1453	1244	85.6	282	222	78.7	1103	998	90.5	76.3	44	41	93.2	40	97.6	90.9	76.8
	59917	674	587	87.1	204	166	81.4	485	444	91.5	78.2	348	320	92	289	90.3	83.0	79.8
	59918	612	500	81.7	170	161	94.7	415	404	97.3	79.2	340	298	87.6	279	93.6	82.1	80.2
	59919	497	443	89.1	224	202	90.2	331	314	94.9	83.5	390	362	92.8	318	87.8	81.5	82.6
	59920	1287	1146	89	294	266	90.5	999	951	95.2	84.2	84.2
Terr.	60901*	2313	1977	85.5	966	808	83.6	1494	1416	94.8	78.7	385	307	79.7	293	95.4	76.1	78.3

* = régions socio-sanitaires regroupées

9.2 Erreurs dans les enquêtes

Une 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 répondant au 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. L'ampleur de la non-réponse varie de non-réponse partielle (le fait de ne pas répondre à une ou plusieurs questions) à une non-réponse totale. Dans le cas du cycle 1.1 de l'ESCC, il n'y a presque pas eu de non-réponse partielle car une fois le questionnaire débuté les répondants avaient tendance à le terminer. Il y a eu non-réponse totale lorsque la personne sélectionnée pour participer à l'enquête a refusé de le faire ou que l'intervieweur a été incapable d'entrer en contact avec elle. On a traité les cas de non-réponse totale en corrigeant les poids des personnes qui ont répondu à l'enquête afin de compenser pour ceux qui n'ont pas répondu. Voir la section 8 pour avoir de plus amples détails sur la correction de la pondération pour la non-réponse.

9.2.2 Erreurs dues à l'échantillonnage

Étant donné que les estimations d'une enquête par sondage comportent inévitablement une erreur due à l'échantillonnage, de bonnes méthodes statistiques exigent que les chercheurs fournissent aux utilisateurs une certaine indication de l'ampleur de cette erreur due à l'échantillonnage. La mesure de l'importance éventuelle des erreurs due 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 elle se rapporte. La mesure résultante, appelée coefficient de variation (CV), s'obtient en divisant l'écart type de l'estimation par l'estimation elle-même et on l'exprime en pourcentage de l'estimation.

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

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

Statistique Canada utilise fréquemment les résultats du CV pour l'analyse des données et conseille vivement aux utilisateurs produisant des estimations à partir des fichiers de données du cycle 1.1 de l'ESCC de faire de même. Pour plus d'information sur le calcul des CVs, 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 en vue d'être publiées ou diffusées de toute autre façon correspondent à celles produites par Statistique Canada, il est vivement conseillé à l'utilisateur de les arrondir en se conformant aux lignes directrices suivantes.

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

Statistique Canada, il est vivement conseillé à l'utilisateur d'indiquer la raison de ces divergences dans le ou les documents à publier ou à diffuser.

- f) Des estimations non arrondies ne doivent être publiées ou diffusées de toute autre façon en aucune circonstance. Des estimations non arrondies donnent l'impression d'être beaucoup plus précises qu'elles ne le sont en réalité.

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

Le plan d'échantillonnage utilisé pour l'ESCC 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 des fichiers 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 à la zone 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 du cycle 1.1 de l'ESCC, il est bon de décrire les deux grandes catégories d'estimations ponctuelles des caractéristiques de la population qui peuvent être produites d'après le fichier de microdonnées de l'enquête.

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 :

SMKA_202 : Actuellement, est-ce que ... fume(z) des cigarettes tous les jours, à l'occasion ou jamais?

- ☐ 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 :

SMKA_204 : Actuellement, combien de cigarettes est-ce que ... fume(z) chaque jour?

|_|_| Nombre de cigarettes

10.2.2 Totalisation d'estimations de type nominal

On peut obtenir, à partir des fichiers 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 des estimations quantitatives d'après le fichier de microdonnées, on doit :

- 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 multiplie la valeur déclarée pour la variable **SMKA_204**²³ par le poids, **WTSAM**, puis on fait la somme des résultats pour tous les enregistrements pour lesquels la valeur de la **variable SMKA_202** est « tous les jours » pour obtenir le numérateur (\hat{X}). Pour obtenir le dénominateur (\hat{Y}), on additionne les poids finals de tous les enregistrements pour lesquels la valeur de la variable **SMKA_202** est « tous les jours ». Pour obtenir le nombre moyen de cigarettes fumées chaque jour par les personnes qui fument tous les jours, on divise (\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), on peut rendre l'application des progiciels courants plus significative en rééchelonnant les poids qui figurent dans les

²³ Voir la section 12.2 pour les conventions appliquées pour nommer les variables.

enregistrements de façon à ce que le poids moyen soit égal à un (1). Les résultats produits par les logiciels classiques sont ainsi plus raisonnables puisque, même s'ils ne reflètent toujours pas la stratification et la mise en grappes du plan d'échantillonnage, ils tiennent compte de la sélection avec probabilités inégales. On peut effectuer cette transformation en utilisant dans l'analyse un poids égal au poids original divisé par la moyenne des poids originaux pour les unités échantillonnées (personnes) qui contribuent à l'estimation en question.

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

10.4 Lignes directrices pour la diffusion

Avant de diffuser et(ou) de publier des estimations tirées des fichiers de microdonnées, l'utilisateur doit déterminer le nombre de personnes qui ont fourni les données entrant dans le calcul de l'estimation. Si ce nombre est inférieur à 30, l'estimation pondérée ne doit pas être diffusée, quelle que soit la valeur de son coefficient de variation. Pour les estimations pondérées basées sur des échantillons d'au moins 30 personnes, l'utilisateur doit calculer le coefficient de variation de l'estimation arrondie et suivre les lignes directrices qui suivent.

Table 10.1 : Lignes directrices relatives à la variabilité d'échantillonnage

Type d'estimation	CV (en %)	Lignes directrices
1. Acceptable	0,0 à 16,5	On peut envisager une diffusion générale non restreinte des estimations. Aucune annotation particulière n'est nécessaire.
2. Marginale	16,6 à 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).
3. Inacceptable	33,3 ou plus	<p>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 :</p> <p>« 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. »</p>

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

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, 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) dans ces tableaux 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 tableaux suivants montrent les effets du plan, la taille des échantillons et les chiffres de population, qui ont servi à produire les tableaux de la variabilité d'échantillonnage approximative.

**Données d'entrée des tableaux de la variabilité d'échantillonnage approximative,
pour les régions socio-sanitaires**

RÉGION SOCIO-SANITAIRE	EFFET DU PLAN	TAILLE DE L'ÉCHANTILLON	POPULATION (12 ans et plus)
Health and Community Services St John's Region (10901)	1,37	892	158 173
Health and Community Services Eastern Region (10902)	1,65	810	103 604
Health and Community Services Central Region (10903)	1,51	711	90 416
Hlth & Com. Serv. Western Reg. / Grenfell Reg. Hlth Serv. Board / Hlth Labrador Corp. (10904)	1,93	1 457	108 911
Région 1 - Île-du-Prince-Édouard (11901)	1,92	1 389	54 439
Région 2 - Île-du-Prince-Édouard (11902)	2,05	2 262	61 887
Zone 1 - Nouvelle-Écosse (12901)	1,58	956	108 292
Zone 2 - Nouvelle-Écosse (12902)	1,55	711	70 499
Zone 3 - Nouvelle-Écosse (12903)	1,52	801	88 937
Zone 4 - Nouvelle-Écosse (12904)	1,63	691	83 250

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

RÉGION SOCIO-SANITAIRE	EFFET DU PLAN	TAILLE DE L'ÉCHANTILLON	POPULATION (12 ans et plus)
Zone 5 - Nouvelle-Écosse (12905)	1,39	820	112 272
Zone 6 - Nouvelle-Écosse (12906)	1,40	1 340	324 722
Région 1 - Nouveau-Brunswick (13901)	1,37	985	157 931
Région 2 - Nouveau-Brunswick (13902)	1,35	915	149 340
Région 3 - Nouveau-Brunswick (13903)	1,49	873	137 485
Région 4 / Région 5 - Nouveau-Brunswick (13904)	1,52	1 061	73 071
Région 6 / Région 7 - Nouveau-Brunswick (13905)	1,52	1 162	116 438
Région du Bas-Saint-Laurent (24901)	1,45	1 127	174 533
Région du Saguenay - Lac-Saint-Jean (24902)	1,51	1 122	241 654
Région de Québec (24903)	1,65	1 653	556 346
Région de la Mauricie-Bois-Francis (24904)	1,90	1 622	408 383
Région de l'Estrie (24905)	2,13	1 180	244 455
Région de Montréal-Centre (24906)	1,44	2 721	1 569 344
Région de l'Outaouais (24907)	1,45	1 185	267 803
Région de l'Abitibi-Témiscaminque (24908)	1,55	1 253	123 875
Région de la Côte-Nord (24909)	2,21	1 098	77 721
Région de la Gaspésie-Îles-de-la-Madeleine (24911)	1,55	1 184	85 308
Région de la Chaudière-Appalaches (24912)	1,69	1 427	330 976
Région de Laval (24913)	1,43	1 045	297 108
Région de Lanaudière (24914)	1,57	1 494	330 528
Région de Laurentides (24915)	1,63	1 440	394 808
Région de la Montérégie (24916)	2,09	2 461	1 113 880
Algoma (35926)	1,48	812	105 357
Brant (35927)	1,54	756	105 865
Durham (35930)	2,11	1 383	427 780
Elgin-St Thomas (35931)	1,70	742	69 966
Bruce-Grey-Owen Sound (35933)	1,64	860	133 720
Haldimand-Norfolk (35934)	1,58	723	93 186
Haliburton (35935)	1,47	967	144 537
Halton (35936)	1,48	1 257	321 355
Hamilton-Wentworth (35937)	1,58	1 326	423 505
Hastings and Prince Edward (35938)	1,51	889	132 650
Huron / Perth (35939)	1,41	1 242	114 057
Kent-Chatham (35940)	1,55	1 059	93 445

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

RÉGION SOCIO-SANITAIRE	EFFET DU PLAN	TAILLE DE L'ÉCHANTILLON	POPULATION (12 ans et plus)
Kingston (35941)	1,56	938	149 440
Lambton (35942)	1,60	866	108 811
Leeds (35943)	1,60	901	137 717
Middlesex-London (35944)	1,50	1 282	348 789
Muskoka-Parry Sound (35945)	1,39	763	72 021
Niagara (35946)	1,48	1 275	362 313
North Bay / Timiskaming (35947)	1,57	1 484	109 218
Northwestern (35949)	1,51	710	55 555
Ottawa Carleton (35951)	1,49	1 936	664 036
Oxford (35952)	1,77	713	85 976
Peel (35953)	1,43	1 837	857 586
Peterborough (35955)	1,42	842	109 112
Porcupine (35956)	1,45	755	74 595
Renfrew (35957)	1,70	722	82 501
Eastern Ontario (35958)	1,50	982	163 078
Simcoe (35960)	1,81	1 338	316 951
Sudbury (35961)	1,51	979	165 778
Thunder Bay (35962)	1,50	959	130 084
Waterloo (35965)	1,42	1 304	378 067
Wellington-Dufferin-Guelph (35966)	1,63	1 170	203 929
Windsor-Essex (35968)	1,49	1 250	324 756
York (35970)	1,56	1 732	634 669
City of Toronto (35995)	1,48	2 524	2 176 887
Winnipeg (46910)	1,39	2 070	535 694
Brandon / Marquette / South Westman (46915)	1,64	1 863	97 676
North Eastman / South Eastman (46920)	1,72	1 271	72 950
Interlake (46930)	1,99	762	60 785
Central (46940)	1,54	827	75 577
Parkland / Norman / Burntwood / Churchill (46960)	2,01	1 677	64 811
Weyburn (A) Service Area / Moose Jaw (B) Service Area / Swift Current (C) Service Area (47901)	1,60	1 855	131 237
Regina (D) Service Area (47904)	1,35	1 171	199 000
Yorkton (E) Service Area / Melfort (H) Service Area (47905)	1,68	1 391	84 013

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

RÉGION SOCIO-SANITAIRE	EFFET DU PLAN	TAILLE DE L'ÉCHANTILLON	POPULATION (12 ans et plus)
Saskatoon (F) Service Area (47906)	1,45	1 274	231 775
Rosetown (G) Service Area / North Battleford (J) Service Area (47907)	1,60	1 236	87 992
Prince Albert (I) Service Area / Northern Health Services Branch (K) Service Area (47909)	1,75	1 082	71 976
Chinook Regional Health Authority (48901)	1,78	890	120 653
Palliser Regional Health Authority (48902)	1,66	726	77 508
Headwaters Regional Health Authority / Regional Health Authority #5 (48903)	1,76	1 324	106 187
Calgary Regional Health Authority (48904)	1,54	2 092	809 818
David Thompson Regional Health Authority (48906)	1,56	973	162 697
East Central Regional Health Authority (48907)	1,63	802	85 574
Westview Regional Health Authority / Crossroads Regional Health Authority (48908)	2,09	1 229	111 313
Capital Health Authority (48910)	1,54	2 111	700 227
Aspen Regional Health Authority (48911)	1,69	761	73 940
Lakeland Regional Health Authority (48912)	1,58	814	84 599
Mistahia Regional Health Authority (48913)	1,44	799	73 189
Peace / Keeweenaw / Northern Lights & Northwestern Reg. Hlth Auth. (48914)	1,73	1 935	75 863
East Kootenay (59911)	1,49	645	67 119
Kootenay-Boundary (59912)	1,39	705	69 608
Okanagan (59913)	1,63	1 671	257 796
Thompson / Cariboo (59914)	1,57	1 668	175 967
Fraser Valley (59921)	1,63	1 125	198 477
Simon Fraser (59922)	1,47	2 036	457 268
South Fraser (59923)	1,53	1 437	496 419
Richmond (59931)	1,53	828	149 581
Vancouver (59932)	1,57	1 285	508 699
North Shore / Coast Garibaldi(59933)	1,55	1 475	222 778
South Vancouver Island (59941)	1,43	1 522	337 941
Central Vancouver Island / North Vancouver Island (59942)	1,69	1 526	247 858
Norhtwest / Northeast (59951)	1,58	1 261	109 769

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

RÉGION SOCIO-SANITAIRE	EFFET DU PLAN	TAILLE DE L'ÉCHANTILLON	POPULATION (12 ans et plus)
Northern Interior (59952)	1,55	1 118	122 391
Yukon / Territoires du Nord-Ouest / Nunavut (60901)	1,12	2 517	76 928

**Données d'entrée des tableaux de la variabilité d'échantillonnage, pour les provinces, les
territoires et le Canada**

PROVINCE	EFFET DU PLAN	TAILLE DE L'ÉCHANTILLON	POPULATION (12 ans et plus)
Terre-Neuve	1,63	3 870	461 104
Île-Du-Prince-Édouard	2,02	3 651	116 326
Nouvelle-Écosse	1,63	5 319	787 972
Nouveau-Brunswick	1,51	4 996	634 264
Québec	2,08	22 012	6 216 722
Ontario	2,31	39 278	9 877 292
Manitoba	2,44	8 470	907 493
Saskatchewan	1,82	8 009	805 993
Alberta	2,39	14 456	2 481 568
Colombie-Britannique	1,77	18 302	3 421 671
Territoires	1,12	2 517	76 928
CANADA	2,34	130 880	25 787 333

Données d'entrée des tableaux de la variabilité d'échantillonnage selon le groupe d'âge

GROUPE D'ÂGE	EFFET DU PLAN	TAILLE DE L'ÉCHANTILLON	POPULATION
12-19	2,13	17 557	3 240 646
20-29	2,20	16 326	4 137 811
30-44	2,17	35 614	7 477 907
45-64	2,44	37 150	7 283 951
65+	2,70	24 233	3 647 018

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 officielles. Les possibilités concernant le calcul d'un coefficient de variation exact sont discutées dans la sous-section 11.7.

Rappel : 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 d'estimations basées 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 variabilité d'échantillonnage correspondant à la région géographique 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. 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). 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 populations ou sous-groupes indépendants, mais elle ne donne que des résultats approximatifs dans les autres cas.

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 5 529 104 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 (5 529 104) ne figure pas dans la colonne de gauche (la colonne «Numérateur du pourcentage»); il faut donc utiliser le nombre qui s'en rapproche le plus, soit 6 000 000.
- 3) Le coefficient de variation d'un agrégat estimé (exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 0,8 %.

- 4) Le coefficient de variation approximatif de l'estimation est donc 0,8 %. Par conséquent, l'estimation selon laquelle 5 529 104 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 $5\,529\,104/6\,677\,374 = 82,8 \%$ 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 (82,8 %) et la partie numérateur du pourcentage (5 529 104) pour déterminer le coefficient de variation.
- 3) Le numérateur (5 529 104) ne figure pas dans la colonne de gauche (la colonne «Numérateur du pourcentage»); il faut donc utiliser le nombre qui s'en rapproche le plus, soit 6 000 000. De même, l'estimation du pourcentage ne figure pas parmi les en-têtes de colonnes; il faut donc utiliser le nombre qui s'en rapproche le plus, soit 90,0 %.
- 4) Le nombre qui se trouve à l'intersection de la ligne et de la colonne utilisées, soit 0,3 %, est le coefficient de variation (exprimé en pourcentage) à employer.
- 5) Le coefficient de variation de l'estimation est donc 0,3 %. Par conséquent, l'estimation selon laquelle 82,8 % 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, $2\,985\,871/12\,697\,959 = 23,5 \%$ fument tous les jours (estimation 1), alors que chez les femmes, ce pourcentage est estimé à $2\,543\,234/13\,089\,375 = 19,4 \%$ (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 1,1 % le CV de l'estimation 1 (exprimé en pourcentage) et à 1,1 % le CV de l'estimation 2 (exprimé en pourcentage).
- 2) Selon la règle 3, l'erreur-type pour une différence ($\hat{d} = \hat{X}_2 - \hat{X}_1$) est :

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

où \hat{X}_1 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 $\hat{d} = (,235 - ,194) = ,041$ est donc :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(,235)(,011)]^2 + [(,194)(,011)]^2} \\ &= ,003\end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = ,003 / ,041 = 0,07$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 7,0 % (exprimé en pourcentage). Par conséquent, cette estimation peut être publiée sans réserve.

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime à 5 529 104 le nombre de personnes qui fument tous les jours et à 1 148 270 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 1 148 270. Le chiffre 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 2,1 %.

- 4) Le dénominateur de cette estimation de rapport est 5 529 104. Le chiffre qui se rapproche le plus de ce nombre est 6 000 000. Le coefficient de variation de cette estimation (exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 0,8 %.
- 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,

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(.021)^2 + (.008)^2} \\ &= 0,022\end{aligned}$$

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 1 148 270/5 529 104, soit 0,21:1. Le coefficient de variation de cette estimation est 2,2 % (exprimé en pourcentage); 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 donne une mesure intuitive plus significative 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.

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

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

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

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

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

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

$$\hat{X} = 0,828$$

$$z = 2$$

$$\alpha_{\hat{X}} = 0,003 \text{ est le coefficient de variation de cette estimation selon les tableaux.}$$

$$CI_X = \{0,828 - (2) (0,828) (0,003), 0,828 + (2) (0,828) (0,003)\}$$

$$CI_X = \{0,823, 0,833\}$$

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 la proportion de femmes qui fument tous les jours. 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 0,003. Par conséquent,

$$z = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0,235 - 0,194}{0,003} = \frac{0,041}{0,003} = 13,7$$

Puisque $z = 13,7$ 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 faisant en sorte que le test soit correct.

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 du cycle 1.1 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 du cycle 1.1 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 via d'autres options, dont celle du télé-accès. Le télé-accès, de même que les autres alternatives possibles pour l'obtention de coefficients de variation exacts, est discuté dans la sous-section 12.3.

Spécifiquement pour le calcul de coefficients de variation, le service de télé-accès permet aux utilisateurs d'avoir accès à 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 celui au niveau d'une région socio-sanitaire (par exemple, au niveau urbain ou rural), les tableaux de CV publiés ne conviennent pas. 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 coefficients obtenus par régression linéaire ou par régression logistique, les tableaux de CV ne fourniront pas les coefficients de variation associés corrects. 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 du cycle 1.1 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

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

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

Les tableaux suivants indiquent les seuils de diffusion des totaux selon les estimations pour les régions socio-sanitaires, les provinces, le Canada, ainsi que pour les différents groupes d'âges (au niveau du Canada seulement). Les estimations inférieures à la valeur indiquée dans la colonne «Marginal» ne peuvent en aucun cas être diffusées.

Tableau des seuils de diffusion des totaux selon les estimations pour les régions socio-sanitaires

RÉGION SOCIO-SANITAIRE	CV	
	CV ENTRE 0 % À 16,5 %	CV ENTRE 16,5 % ET 33,3 %
	ACCEPTABLE	MARGINAL
Health and Community Services St John's Region (10901)	8 500	2 000
Health and Community Services Eastern Region (10902)	7 000	2 000
Health and Community Services Central Region (10903)	6 500	1 500
Région 2 - Île-du-Prince-Édouard (11902)	2 000	500
Zone 1 - Nouvelle-Écosse (12901)	6 000	1 500

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

RÉGION SOCIO-SANITAIRE	CV	
	CV ENTRE 0 % À 16,5 %	CV ENTRE 16,5 % ET 33,3 %
	ACCEPTABLE	MARGINAL
Zone 3 - Nouvelle-Écosse (12903)	6 000	1 500
Zone 4 - Nouvelle-Écosse (12904)	6 500	1 500
Zone 5 - Nouvelle-Écosse (12905)	6 500	1 500
Zone 6 - Nouvelle-Écosse (12906)	12 000	3 000
Région 1 - Nouveau-Brunswick (13901)	7 500	2 000
Région 2 - Nouveau-Brunswick (13902)	7 500	2 000
Région 3 - Nouveau-Brunswick (13903)	8 000	2 000
Région 4 / Région 5 - Nouveau-Brunswick (13904)	3 500	1 000
Région 6 / Région 7 - Nouveau-Brunswick (13905)	5 500	1 500
Région du Bas-Saint-Laurent (24901)	8 000	2 000
Région du Saguenay - Lac-Saint-Jean (24902)	11 500	3 000
Région de Québec (24903)	19 500	5 000
Région de la Mauricie-Bois-Francis (24904)	17 000	4 500
Région de l'Estrie (24905)	15 000	4 000
Région de Montréal-Centre (24906)	30 000	7 500
Région de l'Outaouais (24907)	11 500	3 000
Région de l'Abitibi-Témiscaminque (24908)	5 500	1 500
Région de la Côte-Nord (24909)	5 500	1 500
Région de la Gaspésie-Îles-de-la-Madeleine (24911)	4 000	1 000
Région de la Chaudière-Appalaches (24912)	14 000	3 500
Région de Laval (24913)	14 000	3 500
Région de Lanaudière (24914)	12 500	3 000
Région de Laurentides (24915)	16 000	4 000
Région de la Montérégie (24916)	33 500	8 500
Algoma (35926)	6 500	1 500
Brant (35927)	7 500	2 000
Durham (35930)	22 500	6 000
Elgin-St Thomas (35931)	5 500	1 500
Bruce-Grey-Owen Sound (35933)	9 000	2 500
Haldimand-Norfolk (35934)	7 000	2 000
Haliburton (35935)	7 500	2 000
Halton (35936)	13 500	3 500
Hamilton-Wentworth (35937)	18 000	4 500
Hastings and Prince Edward (35938)	8 000	2 000

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

RÉGION SOCIO-SANITAIRE	CV	
	CV ENTRE 0 % À 16,5 %	CV ENTRE 16,5 % ET 33,3 %
	ACCEPTABLE	MARGINAL
Huron / Perth (35939)	4 500	1 000
Kent-Chatham (35940)	5 000	1 000
Kingston (35941)	8 500	2 000
Lambton (35942)	7 000	2 000
Leeds (35943)	8 500	2 000
Middlesex-London (35944)	14 500	3 500
Muskoka-Parry Sound (35945)	4 500	1 000
Niagara (35946)	15 000	4 000
North Bay / Timiskaming (35947)	4 000	1 000
Northwestern (35949)	4 000	1 000
Ottawa Carleton (35951)	18 500	4 500
Oxford (35952)	7 000	2 000
Peel (35953)	24 000	6 000
Peterborough (35955)	6 500	1 500
Porcupine (35956)	5 000	1 500
Renfrew (35957)	6 500	1 500
Eastern Ontario (35958)	8 500	2 000
Simcoe (35960)	15 000	4 000
Sudbury (35961)	9 000	2 500
Thunder Bay (35962)	7 000	2 000
Waterloo (35965)	14 500	3 500
Wellington-Dufferin-Guelph (35966)	10 000	2 500
Windsor-Essex (35968)	13 500	3 500
York (35970)	20 500	5 000
City of Toronto (35995)	46 000	11 500
Winnipeg (46910)	13 000	3 000
Brandon / Marquette / South Westman (46915)	3 000	1 000
North Eastman / South Eastman (46920)	3 500	1 000
Interlake (46930)	5 500	1 500
Central (46940)	5 000	1 000
Parkland / Norman / Burntwood / Churchill (46960)	2 500	500
Weyburn (A) Service Area / Moose Jaw (B) Service Area / Swift Current (C) Service Area (47901)	4 000	1 000
Regina (D) Service Area (47904)	8 000	2 000

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

RÉGION SOCIO-SANITAIRE	CV	
	CV ENTRE 0 % À 16,5 %	CV ENTRE 16,5 % ET 33,3 %
	ACCEPTABLE	MARGINAL
Yorkton (E) Service Area / Melfort (H) Service Area (47905)	3 500	1 000
Saskatoon (F) Service Area (47906)	9 500	2 500
Rosetown (G) Service Area / North Battleford (J) Service Area (47907)	4 000	1 000
Prince Albert (I) Service Area / Northern Health Services Branch (K) Service Area (47909)	4 000	1 000
Chinook Regional Health Authority (48901)	8 500	2 000
Palliser Regional Health Authority (48902)	6 000	1 500
Headwaters Regional Health Authority / Regional Health Authority #5 (48903)	5 000	1 500
Calgary Regional Health Authority (48904)	21 500	5 500
David Thompson Regional Health Authority (48906)	9 000	2 500
East Central Regional Health Authority (48907)	6 000	1 500
Westview Regional Health Authority / Crossroads Regional Health Authority (48908)	6 500	1 500
Capital Health Authority (48910)	18 500	4 500
Aspen Regional Health Authority (48911)	5 500	1 500
Lakeland Regional Health Authority (48912)	5 500	1 500
Mistahia Regional Health Authority (48913)	4 500	1 000
Peace / Keeweenaw / Northern Lights & Northwestern Reg. Hlth Auth. (48914)	2 500	500
East Kootenay (59911)	5 000	1 500
Kootenay-Boundary (59912)	4 500	1 000
Okanagan (59913)	9 000	2 000
Thompson / Cariboo (59914)	6 000	1 500
Fraser Valley (59921)	10 000	2 500
Simon Fraser (59922)	12 000	3 000
South Fraser (59923)	18 500	4 500
Richmond (59931)	9 500	2 500
Vancouver (59932)	22 000	5 500
North Shore / Coast Garibaldi(59933)	8 500	2 000
South Vancouver Island (59941)	11 500	3 000
Central Vancouver Island / North Vancouver Island (59942)	9 500	2 500

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

RÉGION SOCIO-SANITAIRE	CV	
	CV ENTRE 0 % À 16,5 %	CV ENTRE 16,5 % ET 33,3 %
	ACCEPTABLE	MARGINAL
Norhtwest / Northeast (59951)	5 000	1 000
Northern Interior (59952)	6 000	1 500
Yukon / Territoires du Nord-Ouest / Nunavut (60901)	1 000	500

Tableau des seuils de diffusion des totaux selon les estimations pour les provinces, les territoires et le Canada

PROVINCE OU TERRITOIRES	CV	
	CV ENTRE 0 % ET 16,5 %	CV ENTRE 16,5 % ET 33,3 %
	ACCEPTABLE	MARGINAL
Terre-Neuve	7 000	1 500
Ile-Du-Prince-Édouard	2 500	500
Nouvelle-Écosse	9 000	2 000
Nouveau-Brunswick	7 000	1 500
Québec	21 500	5 500
Ontario	21 500	5 000
Manitoba	9 500	2 500
Saskatchewan	6 500	1 500
Alberta	15 000	3 500
Colombie-Britannique	12 000	3 000
Territoires	1 000	500
CANADA	17 000	4 000

***GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)***

Tableau des seuils de diffusion des totaux selon les estimations des groupes d'âge, Canada

GROUPE D'ÂGE	CV	
	CV ENTRE 0 % ET 16,5 %	CV ENTRE 16,5 % ET 33,3 %
	ACCEPTABLE	MARGINAL
12-19	14 500	3 500
20-29	20 500	5 000
30-44	16 500	4 000
45-64	17 500	4 500
65+	15 000	3 500

12. Utilisation du fichier

La présente section débute par un examen de la **variable de pondération** et des explications sur la façon de les utiliser quand on effectue des totalisations d'après les fichiers de microdonnées à grande diffusion. Suit une explication de la convention appliquée pour nommer les variables du cycle 1.1 de l'ESCC. Enfin vient la description des diverses méthodes d'accès aux données que peuvent adopter les analystes.

12.1 Utilisation des facteurs de pondération

Un seul poids WTSAM figure dans le fichier de micro données à grande diffusion. Ce poids est applicable à chaque groupe d'âge, région, province ou territoire. TOUTES LES VARIABLES DU FICHIER DEVRAIENT ÊTRE ANALYSÉES À L'AIDE DE CE POIDS.

(Pour une description plus détaillée du calcul de ce poids, consulter le section 8 sur la pondération.)

12.2 Convention appliquée pour nommer les variables

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

12.2.1 Structure élémentaire des noms des variables du cycle 1.1 de l' ESCC

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

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

Position 4 : Cycle de l'enquête

Position 5 : Type de variable

Positions 6 à 8 : Numéro de la question

Par exemple, la structure du nom de la variable correspondant à la question 8B, module de la dépression, cycle 1.1, c'est-à-dire **DPSA_08B**, est la suivante :

Positions 1 à 3 : **DPS** Module de la dépression

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

Position 4 : **A** Cycle 1.1
Position 5 : **_** (_ = données recueillies)
Position 6 à 8 : **08B** numéro de la question et option de réponse

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

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

ADM	Administration	INC	Revenu
ALC	Consommation d'alcool	INJ	Blessures
ALD	Dépendance à l'égard de l'alcool	LBF	Population active
BPC	Tension artérielle	MAM	Mammographie
BRF	Allaitement	MAS	Contrôle
BRX	Examen des seins	MDB	Humeur
BSX	Auto-examen des seins	PAC	Activités physiques
CCC	Problèmes de santé chroniques	PAP	Test Papanicolaou
CIH	Changements pour améliorer la santé	PCU	Examen général
CMH	Consultations des spécialistes de la santé mentale	PSA	Test de l'antigène spécifique prostatique
DEN	Visites chez le dentiste	RAC	Limitation des activités
DHH	Données démographiques et composition du ménage	REP	Mouvement répétitif
DIS	Détresse	SAC	Activités sédentaires
DPS	Dépression	SAM	Identificateurs d'échantillon
DRG	Consommation de médicaments	SAT	Satisfaction des patients
DUI	Conduite automobile et consommation d'alcool	SCA	Outils pour arrêter de fumer
EDU	Niveau de scolarité	SDC	Renseignements sociodémographiques
ETS	Exposition à la fumée des autres	SFE	Estime de soi
EYX	Examen des yeux	SMK	Usage du tabac
FIN	Insécurité alimentaire	SPR	Spiritualité
FLU	Vaccination contre la grippe	SSM	Soutien social
FVC	Consommation de fruits et de légumes	SUI	Pensées suicidaires et tentatives de suicide
GEN	État de santé général	SXB	Comportement sexuel
GEO	Identificateurs géographiques (méthodologie)	TAL	Variantes du tabagisme
HCU	Utilisation des soins de santé	TWD	Incapacité au cours des deux dernières semaines
HMC	Soins à domicile	UPE	Utilisation de protections
HUI	Indice de l'état de santé (HUI)	WTS	Poids de sondage
HWT	Taille et poids	WST	Stress au travail

**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)**

12.2.3 Position 4 : Cycle

- | Cycle | Description |
|--------------|--|
| A | <u>Cycle 1.1 : Enquête sur la santé dans les collectivités canadiennes</u>
: enquête à l'échelle régionale, échantillon stratifié selon la région socio-sanitaire
: contenu commun et contenu optionnel sélectionnés par les régions socio-sanitaires
: estimations à l'échelle régionale (régions socio-sanitaires), (provincial, territorial et national) |
| B | <u>Cycle 1.2 : Enquête sur la santé dans les collectivités canadiennes, santé mentale et bien-être</u>
: enquête à l'échelle provinciale
: contenu thématique et contenu général supplémentaire
: estimations aux niveaux provincial, territorial et national |

12.2.4 Position 5 : Type de variable

_	Variable collectée	Variable qui figure directement sur le questionnaire.
C	Variable codée	Variable codée à partir d'une ou de plusieurs variables collectées (par exemple, code de la Classification type des industries (CTI)).
D	Variable transversale 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., groupe d'âge).

12.2.5 Positions 6 à 8 : Nom de la variable

En général, les trois dernières positions correspondent à la numérotation de la variable qui figure sur le questionnaire. On supprime la lettre « Q » utilisée pour représenter le mot “question” et on présente tous les numéros de question au moyen d'un groupe de

deux chiffres. Par exemple, la question Q01A du questionnaire devient simplement 01A et la question 15 (Q15), simplement 15.

Parfois, certaines questions comportent plusieurs réponses alors la position finale dans la séquence du nom de la variable est représentée par une lettre. Pour ce genre de questions, de nouvelles variables sont créées dans le but de différencier un “oui” d’un “non” pour chaque possibilité de réponse. Par exemple, si la question Q2 a 4 réponses possibles, les nouvelles questions seraient Q2A pour la première possibilité, Q2B pour la deuxième, Q2C pour la troisième et ainsi de suite. Si seulement les options 2 et 3 sont choisies, alors Q2A = Non, Q2B = Oui, Q2C = Oui et Q2D = Non.

12.3 Accès aux fichiers maîtres

Afin de respecter le droit à la vie privée des répondants qui participent à l’enquête, les fichiers de microdonnées doivent répondre à des normes sévères de sécurité et de confidentialité, conformément à la *Loi sur la statistique*. Pour s’assurer du respect de ces normes, chaque fichier de microdonnées est soumis à un processus officiel d’examen destiné à confirmer qu’aucune personne ne pourra être identifiée. Les valeurs rares pour certaines variables susceptibles de permettre l’identification d’une personne sont supprimées du fichier ou agrégées en catégories moins détaillées, de façon à réduire au minimum le risque de divulgation de renseignements personnels. Fréquemment, ces variables sont les plus essentielles à l’analyse complète des données d’enquête. Puisqu’une quantité importante de ressources est investie dans la collecte de ces données, il est important de prendre des mesures pour tirer le plein potentiel analytique des fichiers de microdonnées afin de bien rentabiliser l’investissement statistique.

Le téléaccès aux fichiers maîtres de l’enquête est un moyen d’accéder à ces données. On peut fournir à l’acheteur d’un produit de microdonnées un fichier maître “fictif” d’essai et le cliché d’enregistrement correspondant. Grâce à ces outils, il peut mettre au point son propre ensemble de programmes analytiques en se servant du fichier fictif pour confirmer que les routines fonctionnent convenablement. Il ne lui reste plus qu’à envoyer le code pour les totalisations personnalisées par courrier électronique à cchs-escc@statcan.ca. Le code est transmis au réseau interne protégé de Statistique Canada et traité en regard du fichier maître approprié de données du cycle 1.1 de l’ESCC. L’accès à distance permet aux utilisateurs de soumettre leurs programmes informatiques à Statistique Canada, qui les traite. 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 au section 10 de ce document. Les résultats sont filtrés pour vérifier s’ils sont conformes aux normes de confidentialité et de fiabilité, puis, les données de sortie sont renvoyées au client. Ce service est gratuit.

Une deuxième méthode offerte à tous les clients consiste à demander au personnel spécialisé des Services personnalisés à la clientèle de la Division de la statistique de la santé de produire des totalisations personnalisées. Ce service permet aux utilisateurs

***GUIDE DU FICHIER DE MICRODONNÉES À GRANDE
DIFFUSION DU CYCLE 1.1 DE L'ESCC (2000-2001)***

qui ne savent pas se servir de logiciels de totalisation d'obtenir des résultats personnalisés. Comme dans le cas du télé-accès, les résultats sont filtrés pour s'assurer qu'ils sont conformes aux normes de confidentialité et de fiabilité avant d'être diffusés. Contrairement au télé-accès, ce service est offert contre remboursement des frais.

Enfin, le Programme de recherche permet aux chercheurs de soumettre à Statistique Canada un projet de recherche fondé sur les données des fichiers maîtres. Un ensemble particulier de règles est appliqué afin de décider quels projets seront acceptés. Lorsque le projet est accepté, le chercheur est considéré comme étant "réputé employé" par Statistique Canada pour la durée de l'étude et se voit accorder l'accès au fichier maître de l'enquête dans des locaux désignés de Statistique Canada.