



Guide de l'utilisateur des microdonnées

L'Enquête canadienne sur l'expérience des soins de santé primaires

2006-2007



Statistique
Canada

Statistics
Canada

Canada

Table des matières

1.0	Introduction	5
2.0	Contexte	7
3.0	Objectifs	9
4.0	Concepts et définitions	11
5.0	Méthodologie de l'enquête	13
5.1	Population visée par l'Enquête sur la santé dans les collectivités canadiennes	13
5.2	Plan de sondage de l'Enquête sur la santé dans les collectivités canadiennes	13
5.3	Population visée par l'Enquête canadienne sur l'expérience des soins de santé primaires	14
5.4	Stratégie d'échantillonnage des personnes et taille de l'échantillon	14
6.0	Collecte des données	17
6.1	Conception du questionnaire	17
6.2	Supervision et contrôle de qualité	18
7.0	Traitement des données	19
7.1	Saisie des données	19
7.2	Vérification	19
7.3	Codage des questions ouvertes	19
7.4	Imputation	20
7.5	Création de variables dérivées	20
7.6	Pondération	20
7.7	Suppression de renseignements confidentiels	20
8.0	Qualité des données	23
8.1	Taux de réponse	23
8.2	Erreurs relatives à l'enquête	23
8.2.1	Base de sondage	24
8.2.2	Collecte des données	24
8.2.3	Non-réponse	24
8.2.4	Mesure de l'erreur d'échantillonnage	25
9.0	Lignes directrices pour la totalisation, l'analyse et la diffusion de données	27
9.1	Lignes directrices pour l'arrondissement d'estimations	27
9.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation	28
9.3	Définitions de types d'estimations : catégoriques et quantitatives	28
9.3.1	Estimations catégoriques	28
9.3.2	Estimations quantitatives	28
9.3.3	Totalisation d'estimations catégoriques	29
9.3.4	Totalisation d'estimations quantitatives	29
9.4	Lignes directrices pour l'analyse statistique	30
9.5	Lignes directrices pour la diffusion de coefficients de variation	31
9.6	Seuils pour la diffusion des estimations pour l'Enquête canadienne sur l'expérience des soins de santé primaires	33

10.0	Tables de variabilité d'échantillonnage approximative	35
10.1	Comment utiliser les tables de coefficients de variation pour des estimations catégoriques.....	36
10.1.1	Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques	37
10.2	Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance	41
10.2.1	Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance	42
10.3	Comment utiliser les tables de coefficients de variation pour effectuer un test t.....	42
10.3.1	Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t	43
10.4	Coefficients de variation pour des estimations quantitatives.....	43
10.5	Tables des coefficients de variation.....	44
11.0	Pondération	45
11.1	Procédures de pondération pour l'Enquête sur la santé dans les collectivités canadiennes.....	45
11.2	Procédures de pondération pour l'Enquête canadienne sur l'expérience des soins de santé primaires.....	45
12.0	Questionnaire	47
13.0	Cliché d'enregistrement à valeurs univariées	49

1.0 Introduction

Statistique Canada a mené l'Enquête canadienne sur l'expérience des soins de santé primaires (ECE-SSP) en janvier et février 2007 avec l'appui et la collaboration du Conseil canadien de la santé. Ce manuel a été produit pour faciliter la manipulation du fichier de microdonnées portant sur les résultats de l'enquête.

Toutes les questions concernant l'ensemble de données ou son utilisation devraient être adressées à :

Statistique Canada

Services à la clientèle
Division des enquêtes spéciales
Téléphone : 613-951-3321 ou appelez sans frais : 1-800-461-9050
Télécopieur : 613-951-4527
Courriel : des@statcan.ca

Conseil canadien de la santé

Kira Leeb
90, av. Eglinton Est, Suite 900
Toronto (Ontario) M4P 2Y3
Téléphone : (416) 453-8934
Télécopieur : (416) 481-1381
Courriel : kleeb@healthcouncilcanada.ca

2.0 Contexte

Le Conseil canadien de la santé (CCS) a été créé dans le cadre de l'Accord des premiers ministres sur le renouvellement des soins de santé signé en 2003. Son rôle est de rendre compte au public l'avancement du renouvellement des soins de santé au Canada. Le but du Conseil est de fournir une perspective à l'échelle du système sur la réforme des soins de santé à la population canadienne, et ce, en se concentrant particulièrement sur les questions liées à l'évaluation de la responsabilité et de la transparence.

Au cours de l'été 2006, on a demandé à la Division des enquêtes spéciales de mener cette enquête pour le compte du CCS. Le Conseil a déterminé les secteurs clés en matière d'accès et d'utilisation des soins de santé (en particulier pour les personnes atteintes d'une maladie chronique) pour lesquels des données nationales étaient nécessaires. Le questionnaire a été développé à l'automne 2006 et la collecte de données a débuté en janvier 2007.

3.0 Objectifs

Les objectifs principaux de l'enquête sont de recueillir des données sur les problèmes touchant les Canadiens relativement à leur expérience en matière de soins de santé. Plus particulièrement, de dresser le portrait de l'accès aux soins primaires et de leur utilisation ainsi que de fournir de l'information sur les questions qui concernent tout particulièrement les Canadiens atteints d'une maladie chronique et leur expérience du système de soins de santé. Par la suite, les données pourront fournir des renseignements aux fins de l'élaboration de politiques et de stratégies efficaces visant à améliorer les soins de santé pour tous les Canadiens.

Les résultats de l'enquête fourniront une perspective globale de l'expérience des Canadiens en matière de soins de santé et mettront en lumière les problèmes qui touchent les personnes atteintes d'une maladie chronique et permettront d'accroître la sensibilisation à cet égard. Finalement, un des buts ultimes est de fournir des informations qui aideront à faciliter la prise de décisions en ce qui concerne les ressources en fournissant des données de référence qui permettront de suivre l'évolution au fil du temps.

4.0 Concepts et définitions

Étant donné que l'Enquête canadienne sur l'expérience des soins de santé primaires a été réalisée par téléphone, on a fait appel à une terminologie facile à comprendre pour éviter les longues explications des concepts utilisés dans le cadre de l'enquête. L'analyse et l'interprétation des données exigent néanmoins le recours à des concepts et définitions standards. Les questions de l'enquête ont été conçues à la lumière de ces définitions.

Soins de santé primaires : la source principale de soins préventifs, ainsi que les soins continus ou essentiels que les gens reçoivent dans leur collectivité. Ils comprennent les médecins réguliers et les cliniques familiales. Souvent, il s'agit du premier contact du patient avec le système de soins de santé.

Fournisseur de soins primaires : un médecin régulier ou un endroit où les répondants vont pour obtenir des soins de santé.

Médecins réguliers : tous les médecins ayant un diplôme médical d'une université, payés par le régime d'assurance-maladie provincial. On exclut les médecins qui ne sont pas en titre et ceux qui ne participent pas au régime d'assurance-maladie provincial.

5.0 Méthodologie de l'enquête

L'Enquête canadienne sur l'expérience des soins de santé primaires (ECE-SSP) a été menée du 16 janvier au 21 février 2007 auprès d'un sous-échantillon des personnes incluses dans l'échantillon de l'Enquête sur la santé dans les collectivités canadiennes (ESCC), cycle 3.1. Son plan de sondage est donc étroitement lié à celui de l'ESCC. Le plan de l'ESCC 3.1 est décrit brièvement à l'intérieur des sections 5.1 et 5.2¹. Les sections 5.3 et 5.4 décrivent comment l'ECE-SSP s'est écartée du plan de base de l'ESCC 3.1.

5.1 Population visée par l'Enquête sur la santé dans les collectivités canadiennes

Les données de l'ESCC sont recueillies auprès des personnes de 12 ans et plus vivant dans des logements privés des dix provinces et trois territoires. Sont exclus de la base de sondage les habitants des réserves indiennes et des terres de la Couronne, les personnes vivant en institution, 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.

5.2 Plan de sondage de l'Enquête sur la santé dans les collectivités canadiennes

Afin de produire des estimations fiables pour les 122 régions sociosanitaires (RSS) et compte tenu du budget accordé pour le cycle 3.1 de l'ESCC, il a été établi que cette composante de l'enquête devrait être réalisée auprès d'un échantillon de 130 000 personnes. 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. 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.

L'ESCC a utilisé trois bases de sondage pour sélectionner les échantillons de ménages : 50 % de l'échantillon des ménages est venu de la base aréolaire, 49 % est venu d'une base liste de numéros de téléphone et le 1 % restant est venu d'une base d'échantillonnage à composition aléatoire (CA). Pour la plupart des régions sociosanitaires, 50 % de l'échantillon a été sélectionné à partir d'une base aréolaire et 50 % d'une base liste de numéros de téléphone.

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

Pour les besoins du plan d'échantillonnage, 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. Dans certains centres urbains, des strates distinctes sont créées pour les immeubles à appartements ou les secteurs de

¹ Pour plus de détails sur le plan de sondage de l'ESCC 3.1, voir le Guide pour le fichier de microdonnées à grande diffusion, N° 82M0013GPF au catalogue.

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. On obtient l'échantillon final par échantillonnage systématique des logements.

5.3 Population visée par l'Enquête canadienne sur l'expérience des soins de santé primaires

La population visée par l'ECE-SSP est définie de la même façon que pour l'ESCC 3.1, à l'exception du fait qu'elle se limite aux personnes âgées de 18 ans et plus au 16 janvier 2007. Aussi, parce que l'ECE-SSP vise à représenter la population canadienne au début de 2007 mais est sélectionnée à partir des répondants de l'ESCC 3.1, dont la collecte a été réalisée de janvier à décembre 2005, les personnes qui se sont jointes à la population visée entre les deux enquêtes sont exclues. Cette exclusion ne touche pas les gens qui n'avaient pas encore 18 ans au moment de l'ESCC 3.1 puisque celle-ci englobe les personnes âgées de 12 ans et plus.

5.4 Stratégie d'échantillonnage des personnes et taille de l'échantillon

Puisque l'ECE-SSP ne nécessite qu'une petite partie de l'échantillon de l'ESCC 3.1, il a été décidé d'échantillonner seulement parmi les répondants provenant de la base aréolaire de l'ESCC. Les principales raisons étaient que cette base fournit une couverture quasi-complète de la population et que l'utilisation d'une seule base facilite le calcul des poids de sondage. Aussi, l'ECE-SSP ne vise à produire que des estimations au niveau national. Pour cette raison, l'échantillon de la base aréolaire de l'ESCC 3.1 a été stratifié uniquement selon les 122 RSS et l'échantillon a été réparti à chaque RSS proportionnellement à son poids démographique chez les 18 ans et plus. Cette façon de faire a pour but de diminuer l'effet de plan observé pour les estimations provenant de la base aréolaire de l'ESCC où, étant donné les objectifs d'estimation, la fraction de sondage est beaucoup plus élevée dans certaines petites RSS. À l'intérieur de chaque RSS, les répondants de l'ESCC ont été ordonnés selon les autres variables de stratification et de sélection de l'ESCC (et par le fait même de l'EPA) et un échantillon aléatoire systématique a été sélectionné.

La taille de l'échantillon total de l'ECE-SSP est de 3 800 personnes. Le tableau qui suit montre le nombre de personnes échantillonnées dans chaque province et territoire.

Province et territoire	Taille de l'échantillon
Terre-Neuve-et-Labrador	65
Île-du-Prince-Édouard	20
Nouvelle-Écosse	113
Nouveau-Brunswick	92
Québec	903
Ontario	1 463
Manitoba	137
Saskatchewan	114
Alberta	372
Colombie-Britannique	506
Territoires	15
Canada	3 800

6.0 Collecte des données

Une lettre de présentation et un dépliant renfermant de l'information au sujet du Conseil canadien de la santé ont été postés aux répondants environ une semaine avant le début de la collecte des données de l'Enquête canadienne sur l'expérience des soins de santé primaires (ECE-SSP), qui a eu lieu de la mi-janvier jusqu'à la dernière semaine de février 2007. Les données ont été recueillies au moyen d'interviews téléphoniques utilisant papier et crayon, ce qui signifie que les intervieweurs lisaient les questions par téléphone aux répondants et inscrivaient les réponses aux questions à la main sur les questionnaires papier.

Le module frontal du questionnaire comprenait une série de codes de réponse normalisés pour faire face à tous les résultats possibles des appels, ainsi que les scripts connexes que devaient lire les intervieweurs. Une approche normalisée a été mise au point afin de présenter l'organisme, le nom et l'objet de l'enquête, les parrains de l'enquête, la façon dont les résultats de l'enquête seront utilisés et la durée de l'interview. Nous avons expliqué aux répondants comment ils ont été sélectionnés pour participer à l'enquête, que leur participation à l'enquête était volontaire et que les renseignements qu'ils nous fourniraient demeureraient strictement confidentiels.

Le gestionnaire de l'enquête a rencontré les cadres supérieurs responsables de la collecte pour discuter des problèmes et des questions avant le début de la séance de formation. Le Manuel de l'intervieweur comportait une description du contexte et les objectifs, ainsi qu'une description détaillée des concepts et des définitions propres à l'ECE-SSP. Il renfermait également un glossaire et une série de questions et réponses.

Les intervieweurs ont reçu leur formation sur le contenu de l'enquête dans le cadre d'une séance de formation en classe. De plus, ils ont effectué une série d'interviews simulées pour se familiariser avec l'enquête et ses concepts et définitions. La documentation sur les questions et réponses a été fournie aux intervieweurs afin de les aider à répondre aux questions fréquentes des répondants.

La collecte des données a été réalisée par le personnel spécialisé du bureau central de Statistique Canada à Ottawa. On a eu recours à un système d'ordonnancement automatisé d'interview téléphonique assistée par ordinateur (BLAISE) pour gérer et faciliter le processus de collecte, de manière à ce que tous les cas soient attribués de façon aléatoire aux intervieweurs et que les répondants soient contactés à divers moments de la journée, et à divers jours de la semaine, en vue d'augmenter la probabilité de prise de contact.

La durée moyenne d'une interview a été estimée à 22 minutes. Toutefois, la durée des interviews fluctuait selon les circonstances des répondants. Par exemple, la durée moyenne de l'interview était estimée à 30 minutes pour les répondants ayant des états chroniques et à 12 minutes lorsque ce n'était pas le cas. La moyenne globale se rapprochait davantage de 30 minutes que des 22 minutes qu'on avait estimées au départ.

Les questionnaires remplis ont été envoyés à l'imagerie et à la saisie des données.

6.1 Conception du questionnaire

Le Conseil canadien de la santé avait créé un questionnaire provisoire renfermant plus de 120 questions. Ce questionnaire provisoire était un amalgame des questions de diverses enquêtes menées aux États-Unis, au Canada et dans d'autres pays du Commonwealth. La version provisoire a été remaniée afin d'harmoniser les concepts, les définitions et les périodes de référence. La nouvelle version a été créée pour refléter les buts et les objectifs de recherche du Conseil canadien de la santé et renfermait de nouvelles questions qui ne faisaient pas partie de la version provisoire. Le questionnaire a été grandement écourté, et l'enchaînement des questions amélioré. Le questionnaire remanié a été traduit par la Division des langues officielles et de la traduction, puis mis à l'essai avec le concours du Centre d'information sur la conception

des questionnaires (CICQ) au moyen d'interviews sur place dans les deux langues officielles. L'essai a été effectué auprès de répondants de divers groupes d'âges et aux antécédents ethniques variés. Une partie du groupe d'essai était composée de personnes ayant reçu un diagnostic d'état chronique. D'autres changements ont été apportés au questionnaire en fonction des résultats du processus d'essai du questionnaire.

6.2 Supervision et contrôle de qualité

L'équipe des intervieweurs travaillait sous la surveillance des intervieweurs principaux responsables de veiller à ce que tout le monde soit familier avec les concepts et les procédures de l'enquête. L'observation périodique des intervieweurs et l'examen des documents remplis ont été effectués conformément au protocole de collecte.

7.0 Traitement des données

Le principal produit de L'Enquête canadienne sur l'expérience des soins de santé primaires (ECE-SSP) est un fichier de microdonnées « épuré ». Ce chapitre présente un bref résumé des phases de traitement inhérentes à la production de ce fichier.

Une première phase de détection des erreurs s'est déroulée durant la collecte des données. Les superviseurs des intervieweurs ont examiné à ce stade les questionnaires remplis. Ils ont discuté des incohérences qu'ils ont observées sur ces derniers avec les gens ayant fait passer les interviews et les répondants ont été rappelés au besoin.

Une seconde phase de détection des erreurs s'est déroulée durant le traitement des données, qui comprend bien des étapes. La première étape consistait à procéder à l'épuration générale des données afin :

- 1) d'éliminer les enregistrements en double du fichier,
- 2) de vérifier le questionnaire de base en se référant au fichier de l'échantillon,
- 3) de repérer les enregistrements manquants,
- 4) de créer un fichier de réponses.

La validation des données a inclus une vérification descendante afin d'épurer les chemins empruntés par erreur durant l'interview. Une fois cette étape achevée, on a effectué une vérification en analysant des distributions de fréquences pour repérer les anomalies, les catégories manquantes ou invalides ou les fréquences exceptionnelles, par exemple.

7.1 Saisie des données

Les réponses aux questions de l'enquête, lues par téléphone sur un questionnaire papier par les intervieweurs aux répondants, sont saisies directement par l'intervieweur au moment de l'interview. Les données de réponse ont été saisies et traitées au bureau central d'Ottawa.

Une partie de la vérification est effectuée directement au moment de l'interview. Les données de réponse sont soumises à des processus supplémentaires de contrôle et imputation à leur arrivée au centre de traitement du bureau central.

7.2 Vérification

La première étape du traitement d'enquête effectué au bureau central a été de remplacer les valeurs « hors limites » incluses dans le fichier de données par des blancs. Ce processus a été conçu pour faciliter les vérifications ultérieures.

Les erreurs dans le déroulement du questionnaire, où l'on a relevé des questions qui ne s'appliquaient pas au répondant (et auxquelles on n'aurait donc pas dû répondre) et qui renfermaient des réponses, constituaient le premier type d'erreurs traitées. Dans ces cas, une vérification par ordinateur a éliminé automatiquement les données superflues en suivant l'ordre du questionnaire dicté par les réponses à des questions antérieures et subséquentes, parfois.

Le second type d'erreurs traitées avait trait à un manque d'information dans les questions pour lesquelles le répondant aurait dû répondre. Pour ce type d'erreur, un code de non-réponse ou « non déclaré » était attribué au poste.

7.3 Codage des questions ouvertes

Il n'y avait aucune question ouverte dans cette enquête.

7.4 Imputation

L'imputation est le processus qui fournit des valeurs valides concernant les variables qui ont été retenues pour être modifiées, soit en raison de renseignements invalides, soit en raison de renseignements manquants. Les nouvelles valeurs sont établies de façon à préserver la structure sous-jacente des données et à garantir que les enregistrements qui en résultent passeront tous les contrôles requis. En d'autres mots, l'objectif n'est pas de reproduire les véritables valeurs des microdonnées mais plutôt d'établir des enregistrements intrinsèquement cohérents qui permettront de produire de bonnes estimations agrégées.

Nous pouvons faire la distinction entre trois types de non-réponse. La non-réponse complète est lorsque le répondant ne fournit pas le nombre minimal de réponses. Ces enregistrements sont supprimés et seront pris en compte lors du processus de pondération (voir le chapitre 11.0). La non-réponse ponctuelle est lorsque le répondant ne fournit pas une réponse à une question mais qu'il passe à la question suivante. Ce type de non-réponse est habituellement traité en utilisant le code « non déclaré » ou en ayant recours à l'imputation. Enfin, la non-réponse partielle est lorsque le répondant fournit le nombre minimal de réponses mais ne termine pas l'interview. Ces enregistrements peuvent être traités comme des cas de non-réponse complète ou ponctuelle.

Puisque les données recueillies sur cette enquête portent sur les expériences individuelles des répondants avec le système de santé, aucune imputation n'a été effectuée.

7.5 Création de variables dérivées

À quelques reprises, le sexe et l'âge inscrits au fichier n'étaient pas identiques à ce qui figurait dans le fichier de l'Enquête sur la santé dans les collectivités canadiennes (ESCC), cycle 3.1, en raison peut-être d'une erreur non due à l'échantillonnage. Afin de garantir la cohérence et de faciliter l'analyse des données, on a dérivé le sexe et l'âge à l'aide de l'information tirée de l'ESCC 3.1 parce que cette dernière a été jugée plus précise.

7.6 Pondération

Le principe qui sous-tend une estimation pour un échantillon probabiliste comme celui de l'ECE-SSP veut que chacune des personnes incluses dans l'échantillon « représente », en plus d'elle-même, plusieurs autres personnes qui en sont exclues. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne incluse dans l'échantillon représente 50 membres de la population.

La phase de la pondération est une étape où l'on calcule ce nombre (ou poids) pour chaque enregistrement. Ce poids, qui figure dans le fichier de microdonnées, **doit** servir à calculer des estimations significatives à partir de l'enquête. Si, par exemple, le nombre de personnes qui recommanderaient sûrement ou probablement leur fournisseur de soins primaires à un ami ou à un parent doit être estimé, cette opération s'effectue en sélectionnant les enregistrements correspondant aux personnes incluses dans l'échantillon qui présentent cette caractéristique et en additionnant les poids inscrits dans ces enregistrements.

Le chapitre 11.0 renferme des détails au sujet de la méthode utilisée pour calculer ces poids.

7.7 Suppression de renseignements confidentiels

Il convient de souligner que l'unique fichier de microdonnées produit dans le cadre de l'ECE-SSP est un fichier partagé. Statistique Canada ne détient aucun fichier de microdonnées à grande diffusion (FMGD) ni fichier maître. Le fichier partagé contient des données sur tous les

répondants qui ont accepté de communiquer leurs données au Conseil canadien de la santé, ainsi que sur ceux qui ont accepté de permettre à Statistique Canada de coupler leurs données d'enquête avec l'ESCC 3.1. Il convient de souligner que les données couplées, conformément aux politiques sur la confidentialité de Statistique Canada, ne font pas partie du fichier partagé. Par conséquent, les données couplées ne sont pas communiquées au Conseil canadien de la santé. Comme le taux de partage/couplage était très élevé, soit plus de 94 %, on a jugé que la création d'un fichier maître n'était pas justifiée. Tous les renseignements personnels susceptibles de révéler l'identité des répondants, y compris les noms, numéros de téléphone, adresses municipales et codes postaux, ont été supprimés du fichier partagé. Les estimations produites seront diffusées à l'utilisateur, à condition de respecter les lignes directrices en matière d'analyse et de diffusion, qui sont décrites au chapitre 9.0 du présent document.

8.0 Qualité des données

8.1 Taux de réponse

Au total, 3 800 personnes ont été sélectionnées pour participer à l'Enquête canadienne sur l'expérience des soins de santé primaire (ECE-SSP). De ce nombre, 24 ne faisaient plus partie de la population cible de l'ECC-SSP (par exemple, des décès). Sur les 3 776 personnes admissibles, 2 194 ont répondu à l'enquête, pour un taux de réponse global de 58,1 %. Le tableau qui suit renferme un résumé des taux de réponse au questionnaire de l'ECC-SSP par province.

Provinces et territoires	Personnes sélectionnées de l'ECC 3.1	Personnes admissibles à l'enquête	Répondants	Taux de Réponse (%)
Terre-Neuve-et-Labrador	65	65	38	58,5
Île-du-Prince-Édouard	20	20	11	55,0
Nouvelle-Écosse	113	111	74	66,7
Nouveau-Brunswick	92	91	56	61,5
Québec	903	899	504	56,1
Ontario	1 463	1 456	897	61,6
Manitoba	137	136	82	60,3
Saskatchewan	114	113	56	49,6
Alberta	372	371	204	55,0
Colombie-Britannique	506	499	267	53,5
Territoires	15	15	5	33,3
Canada	3 800	3 776	2 194	58,1

8.2 Erreurs relatives à l'enquête

Les estimations calculées à partir de cette enquête reposent sur un échantillon de personnes. Des estimations légèrement différentes auraient pu être obtenues si un recensement complet avait été effectué en reprenant le même questionnaire et en faisant appel aux mêmes intervieweurs, superviseurs, méthodes de traitement, etc. que ceux effectivement utilisés dans l'enquête. L'écart entre les estimations découlant de l'échantillon et celles que donnerait un dénombrement complet réalisé dans des conditions semblables est appelé erreur d'échantillonnage de l'estimation.

Des erreurs qui ne sont pas liées à l'échantillonnage peuvent se produire à presque toutes les étapes des opérations d'enquête. Les intervieweurs peuvent avoir mal compris les instructions, les enquêtés peuvent se tromper en répondant aux questions, les réponses peuvent être mal saisies sur le questionnaire et des erreurs peuvent survenir lors du traitement et de la totalisation des données. Ces erreurs sont toutes des exemples d'erreurs non dues à l'échantillonnage.

Sur un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations calculées à partir de l'enquête. Toutefois, les erreurs systématiques contribuent à biaiser les estimations de l'enquête. Énormément de temps et d'efforts ont été consacrés à réduire les erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été prises à chacune des étapes du cycle de collecte et de traitement des données afin de

contrôler la qualité des données. Ces mesures comprenaient le recours à des intervieweurs hautement qualifiés, une formation poussée des intervieweurs concernant les procédures d'enquête et du questionnaire, l'observation des intervieweurs en vue de cerner les problèmes liés à la conception du questionnaire ou à une mauvaise compréhension des instructions, des procédures visant à s'assurer que les erreurs de saisie des données étaient réduites au minimum ainsi que des vérifications ayant pour but d'attester la logique du traitement.

8.2.1 Base de sondage

Comme l'ECE-SSP était un suivi au cycle 3.1 de l'Enquête sur la santé dans les collectivités canadiennes (la base aréolaire seulement) et que cette dernière utilisait la base de l'Enquête sur la population active (EPA), la qualité des variables d'échantillon dans la base était très élevée et la couverture était très bonne. La base de sondage de l'EPA exclut environ 2 % de tous les ménages au Canada. La base de sondage de l'ECE-SSP exclut donc la même proportion de ménages dans les mêmes régions géographiques. Il est peu probable que cette exclusion introduise un biais important dans les données de l'enquête.

Il est à noter que l'entrevue de l'ECE-SSP avait lieu de 13 à 24 mois après celle de l'ESCC 3.1. Pour certaines des personnes choisies à l'ECE-SSP, il n'y avait pas de numéro de téléphone indiqué dans la base de sondage et pour d'autre, ce numéro était périmé. Des efforts de dépistage ont été faits pour tenter de trouver un numéro de téléphone pour ces personnes mais le temps et les ressources étaient limitées et 557 des 3 800 personnes choisies n'ont pu être contactées pour cette raison.

8.2.2 Collecte des données

La formation des intervieweurs pour l'ECE-SSP consistait à lire le Manuel de l'intervieweur et à se familiariser avec le matériel. L'information fournie aux intervieweurs comprenait une description du contexte et des objectifs de l'enquête ainsi qu'un glossaire de terminologie et une série de questions et de réponses. La période de collecte originale du 16 janvier au 2 février 2007 a été prolongée jusqu'au 21 février.

Bien que l'ECE-SSP ait été réalisée à l'aide d'un questionnaire papier, la gestion des appels et des rendez-vous se faisait à l'aide du progiciel BLAISE. Tous les intervieweurs étaient déjà familiers avec le système.

8.2.3 Non-réponse

L'effet de la non-réponse sur les résultats de l'enquête constitue une source importante d'erreurs non dues à l'échantillonnage dans les enquêtes. L'ampleur de la non-réponse varie de non-réponse partielle (le fait de ne pas répondre à une ou plusieurs questions) à une non-réponse complète. Dans le cas de l'ECE-SSP, 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 complète 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 complète en corrigeant les poids des personnes qui ont répondu à l'enquête afin de compenser pour ceux qui n'y ont pas répondu. Voir le chapitre 11.0 pour avoir de plus amples détails sur la correction de la pondération pour la non-réponse. Dans les cas de non-réponse partielle, aucune imputation n'a été effectuée.

8.2.4 Mesure de l'erreur d'échantillonnage

Puisqu'il est inévitable que des estimations établies à partir d'une enquête-échantillon (ou par sondage) soient sujettes à une erreur d'échantillonnage, une saine pratique de la statistique exige que les chercheurs fournissent aux utilisateurs une certaine indication de l'importance de cette erreur d'échantillonnage. Cette section de la documentation renferme un aperçu des mesures de l'erreur d'échantillonnage dont Statistique Canada se sert couramment et que le Bureau conseille vivement aux utilisateurs qui produisent des estimations à partir de ce fichier de microdonnées d'employer également.

La base pour mesurer l'importance potentielle des erreurs d'échantillonnage est l'erreur-type des estimations calculées à partir des résultats d'une enquête.

En raison, cependant, de la diversité des estimations pouvant être produites à partir d'une enquête, l'erreur-type d'une estimation est habituellement exprimée en fonction de l'estimation à laquelle elle se rapporte. La mesure résultante, appelée coefficient de variation (CV) d'une estimation, s'obtient en divisant l'erreur-type de l'estimation par l'estimation elle-même et s'exprime en pourcentage de l'estimation.

Par exemple, supposons que d'après les résultats de l'enquête, l'on estime que 38,3% des Canadiennes et des Canadiens ont été diagnostiqués ou traités par un professionnel de la santé pour au moins une des conditions chroniques énumérées dans l'enquête et l'on constate que l'erreur-type de cette estimation est de 0,012. Le coefficient de variation de l'estimation est donc calculé comme suit :

$$\left(\frac{0,012}{0,383} \right) \times 100 \% = 3,1 \%$$

De plus amples renseignements sur le calcul du coefficient de variation, se trouvent au chapitre 10.0.

9.0 Lignes directrices pour la totalisation, l'analyse et la diffusion de données

Ce chapitre de la documentation renferme un aperçu des lignes directrices que doivent respecter les utilisateurs qui totalisent, analysent, publient ou autrement diffusent des données calculées à partir des fichiers de microdonnées de l'enquête. Ces lignes directrices devraient permettre aux utilisateurs de microdonnées de produire les mêmes chiffres que ceux produits par Statistique Canada, tout en étant en mesure d'obtenir des chiffres actuellement inédits de façon conforme à ces lignes directrices établies.

9.1 Lignes directrices pour l'arrondissement d'estimations

Afin que les estimations qui sont destinées à la publication ou à toute autre forme de diffusion qui sont calculées à partir de ces fichiers de microdonnées correspondent à celles produites par Statistique Canada, nous conseillons vivement aux utilisateurs de respecter les lignes directrices qui suivent en ce qui concerne l'arrondissement de telles estimations :

- a) Les estimations dans le corps principal d'un tableau statistique doivent être arrondies à la centaine près à l'aide de la technique d'arrondissement normale. Selon cette technique, si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, le dernier chiffre à conserver est augmenté de 1. Par exemple, selon la technique d'arrondissement normale à la centaine près, si les deux derniers chiffres se situent entre 00 et 49, ils sont remplacés par 00 et le chiffre précédent (le chiffre des centaines) reste inchangé. Si les derniers chiffres se situent entre 50 et 99, ils sont remplacés par 00 et le chiffre précédent est augmenté de 1.
- b) Les totaux partiels marginaux et les totaux marginaux des tableaux statistiques doivent être calculés à partir de leurs composantes non arrondies correspondantes, puis ensuite être arrondis à leur tour à la centaine près à l'aide de la technique d'arrondissement normale.
- c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir de composantes non arrondies (c'est-à-dire des numérateurs et/ou des dénominateurs), puis être arrondis à leur tour à une décimale à l'aide de la technique d'arrondissement normale. Dans le cas d'un arrondissement normal à un seul chiffre, si le dernier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, le dernier chiffre à conserver est augmenté de 1.
- d) Les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs composantes non arrondies correspondantes, puis être arrondies à leur tour à la centaine près (ou à la décimale près) à l'aide de la technique d'arrondissement normale.
- e) Dans les cas, où, en raison de limitations d'ordre technique ou de toutes autres limites, une technique d'arrondissement autre que la technique normale est utilisée produisant des estimations à être publiées ou autrement diffusées différentes des estimations correspondantes publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces différences dans le ou les documents à publier ou à diffuser.
- f) En aucun cas, les utilisateurs ne doivent publier ou autrement diffuser des estimations non arrondies. Des estimations non arrondies laissent entendre qu'elles sont plus précises qu'elles ne le sont en réalité.

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

Le plan d'échantillonnage utilisé pour l'Enquête canadienne sur l'expérience des soins de santé primaires (ECE-SSP) n'était pas autopondéré. Lorsqu'ils produisent des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent appliquer le poids d'enquête approprié.

Si les poids appropriés ne sont pas utilisés, les estimations calculées à partir des fichiers de microdonnées ne peuvent être considérées comme représentatives de la population visée par l'enquête et ne correspondront pas à celles produites par Statistique Canada.

Les utilisateurs devraient également prendre note que certains progiciels pourraient peut-être ne pas permettre la production d'estimations correspondant exactement à celles qu'offre Statistique Canada, en raison du mode de traitement du poids par ces progiciels.

9.3 Définitions de types d'estimations : catégoriques et quantitatives

Avant de discuter de la façon dont on peut totaliser et analyser les données de l'ECE-SSP, il est utile de décrire les deux principaux types d'estimations ponctuelles des caractéristiques de la population qui peuvent être produites à partir du fichier de microdonnées créé pour l'ECE-SSP.

9.3.1 Estimations catégoriques

Les estimations catégoriques sont des estimations du nombre ou du pourcentage de membres de la population visée par l'enquête possédant certaines caractéristiques ou faisant partie d'une catégorie définie. Le nombre de personnes qui recommanderaient sûrement ou probablement leur fournisseur de soins primaires à un ami ou à un parent ou encore la proportion de gens ayant passé au moins une nuit comme patient à l'hôpital, dans une maison de soins infirmiers ou une maison de convalescence au cours des 12 derniers mois constituent des exemples de telles estimations. Une estimation du nombre de personnes possédant une certaine caractéristique peut aussi être désignée comme une estimation d'un agrégat.

Exemples de questions catégoriques :

Q : En général, diriez-vous que VOTRE santé est :

R : Excellente / Très bonne / Bonne / Passable / Mauvaise

Q : Au cours des 12 derniers mois, avez-vous eu besoin de services de santé de routine ou de suivi ?

R : Oui / Non

9.3.2 Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes et d'autres mesures d'une tendance centrale de quantités reposant sur certains ou sur tous les membres de la population visée par l'enquête. Elles comprennent aussi expressément des estimations de la forme \hat{X} / \hat{Y} où \hat{X} est une estimation de la quantité totale pour la population visée par l'enquête et \hat{Y} , est une estimation du nombre de personnes dans la population visée par l'enquête qui

contribuent à cette quantité totale.

Un exemple d'estimation quantitative est le nombre moyen de nuits passées comme patient à l'hôpital, dans une maison de soins infirmiers ou une maison de convalescence au cours des 12 derniers mois par les personnes ayant passé au moins une nuit dans une telle institution. Le numérateur (\hat{X}) est une estimation du nombre total de nuits passées dans ces institutions dans les 12 derniers mois et son dénominateur (\hat{Y}) est le nombre de personnes ayant déclaré y avoir passé au moins une nuit.

Exemples de questions quantitatives :

Q : Combien de nuits dans les 12 derniers mois ?

R : |_|_| nuits

Q : En vous incluant, combien de personnes vivent habituellement ici ?

R : |_|_| personnes

9.3.3 Totalisation d'estimations catégoriques

On peut obtenir des estimations du nombre de gens possédant une certaine caractéristique à partir du fichier de microdonnées en additionnant les poids finaux de tous les enregistrements possédant la ou les caractéristiques qui nous intéressent. On obtient des proportions et des rapports de la forme \hat{X} / \hat{Y} en :

- additionnant les poids finaux des enregistrements présentant la caractéristique qui nous intéresse pour le numérateur (\hat{X}),
- additionnant les poids finaux des enregistrements présentant la caractéristique qui nous intéresse pour le dénominateur (\hat{Y}), puis en
- divisant l'estimation calculée en a) par celle en b) (\hat{X} / \hat{Y}).

9.3.4 Totalisation d'estimations quantitatives

On peut obtenir des estimations de quantités à partir du fichier de microdonnées en multipliant la valeur de la variable qui nous intéresse par le poids final de chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements qui nous intéressent. Pour obtenir, par exemple, une estimation du nombre total de fois où les répondants ont vu ou consulté un médecin de famille ou un omnipraticien pour des troubles physiques, émotifs ou mentaux au cours des 12 derniers mois, multipliez la valeur déclarée à la question D02 (nombre de consultations avec un médecin de famille ou un omnipraticien) par le poids final de l'enregistrement, puis additionnez cette valeur pour tous les enregistrements où la variable D_SEX =2 (sexe du participant = féminin).

Pour obtenir une moyenne pondérée de la forme \hat{X} / \hat{Y} , le numérateur (\hat{X}) est calculé comme une estimation quantitative et le dénominateur (\hat{Y}) est calculé comme une estimation catégorique. Pour estimer, par exemple, le nombre moyen de fois où des femmes ont vu ou consulté un médecin de famille ou un omnipraticien pour des troubles physiques, émotifs ou mentaux au cours des 12 derniers mois,

- estimez le nombre total de consultations (\hat{X}) tel qu'il est décrit ci-dessus,

- b) estimez le nombre de femmes (\hat{Y}) dans la population cible en additionnant les poids finaux de tous les enregistrements où la variable D_SEX = 2, puis
- c) divisez l'estimation a) par l'estimation b) (\hat{X} / \hat{Y}).

9.4 Lignes directrices pour l'analyse statistique

L'Enquête canadienne sur l'expérience des soins de santé primaires repose sur un plan d'échantillonnage complexe comportant une stratification, de multiples étapes de sélection ainsi que des probabilités inégales de sélection des répondants. L'utilisation des données provenant d'enquêtes aussi complexes présente des problèmes pour les analystes, parce que le plan d'enquête et les probabilités de sélection influent sur les procédures d'estimation et de calcul de la variance qui devraient être utilisées. Il faut utiliser les poids de l'enquête pour que les estimations et les analyses des données de l'enquête soient exemptes de biais.

Bien que de nombreuses procédures d'analyse que l'on trouve à l'intérieur de progiciels statistiques permettent d'utiliser des poids, la signification ou la définition du poids inclus dans ces procédures peut différer de ce qui convient dans le contexte d'une enquête-échantillon, de telle sorte que dans bien des cas, les estimations produites au moyen de ces progiciels sont correctes, mais que les variances calculées sont de piètre qualité. Les variances approximatives pour des estimations simples comme des totaux, des proportions et des rapports (pour des variables qualitatives) peuvent être calculées à partir des tables de variabilité d'échantillonnage approximative qui accompagnent les données.

Pour d'autres techniques d'analyse (de régression linéaire, de régression logistique et de l'analyse de variance, par exemple), il existe une méthode qui peut rendre les variances calculées par les progiciels courants plus significatives, en intégrant les probabilités inégales de sélection. L'application de cette méthode entraîne une remise à l'échelle des poids de façon à ce que le poids moyen soit de 1.

Supposons, par exemple, qu'il faille effectuer l'analyse de tous les répondants de sexe masculin. Les étapes à suivre pour remettre à l'échelle les poids sont les suivantes :

- 1) sélectionner tous les répondants du fichier qui ont déclaré D_SEX = homme;
- 2) calculer le poids MOYEN pour ces enregistrements en additionnant les poids originaux des personnes établis à partir du fichier de microdonnées pour ces enregistrements puis diviser cette somme par le nombre de répondants ayant déclaré D_SEX = homme;
- 3) pour chacun de ces répondants, calculer un poids REMIS À L'ÉCHELLE égal au poids original de la personne divisé par le poids MOYEN;
- 4) effectuer l'analyse portant sur ces répondants en utilisant le poids REMIS À L'ÉCHELLE.

Parce qu'on ne tient toujours compte ni de la stratification ni des grappes du plan d'échantillonnage, les estimations de variance calculées avec cette méthode risquent cependant d'être des sous-estimations.

Il faut connaître les détails du plan d'enquête pour calculer des estimations de variance plus précises. De tels détails ne peuvent être fournis dans le fichier de microdonnées pour des raisons de confidentialité. Statistique Canada peut, contre remboursement des frais, calculer des variances qui tiennent compte du plan complet d'échantillonnage pour beaucoup de statistiques.

9.5 Lignes directrices pour la diffusion de coefficients de variation

Avant de diffuser et/ou de publier toute estimation établie à partir de l'ECE-SSP, les utilisateurs devraient premièrement déterminer le niveau de qualité de cette estimation. Les niveaux de qualité sont *acceptable*, *médiocre* et *inacceptable*. Les erreurs d'échantillonnage et non dues à l'échantillonnage, dont il a été question au chapitre 8.0, influencent la qualité des données. Aux fins du présent document, cependant, on ne déterminera le niveau de qualité d'une estimation qu'à partir d'une erreur d'échantillonnage dont rend compte le coefficient de variation indiqué à l'intérieur du tableau qui figure ci-dessous. Les utilisateurs devraient néanmoins s'assurer de lire le chapitre 8.0 pour être plus pleinement informés des caractéristiques relatives à la qualité de ces données.

On devrait premièrement déterminer le nombre de répondants retenus pour le calcul de l'estimation. Si ce nombre est inférieur à 30, il faudrait considérer l'estimation pondérée comme étant de qualité inacceptable.

Pour les estimations pondérées fondées sur des tailles d'échantillon de 30 ou plus, les utilisateurs devraient déterminer le coefficient de variation de l'estimation et suivre les lignes directrices relatives au niveau de qualité qui figurent ci-dessous. Celles-ci devraient être appliquées, pour la détermination du niveau de qualité d'une estimation, aux estimations pondérées arrondies.

On peut considérer qu'il est possible de divulguer toutes les estimations. Celles d'un niveau de qualité médiocre ou inacceptable doivent cependant être accompagnées d'une mise en garde pour avertir les utilisateurs subséquents.

Lignes directrices relatives au niveau de qualité de l'estimation

Niveau de qualité de l'estimation	Lignes directrices
1) Acceptable	<p>Les estimations proviennent d'une taille d'échantillon de 30 ou plus, et présentent de faibles coefficients de variation, de l'ordre de 0,0 à 16,5 %.</p> <p>Aucune mise en garde n'est requise.</p>
2) Médiocre	<p>Les estimations proviennent d'une taille d'échantillon de 30 ou plus, et présentent des coefficients de variation élevés, de l'ordre de 16,6 à 33,3 %.</p> <p>Ces estimations devraient être signalées par la lettre E (ou un quelconque identificateur similaire). Elles devraient être accompagnées d'une mise en garde avertissant les utilisateurs subséquents des niveaux élevés d'erreur associés aux estimations.</p>
3) Inacceptable	<p>Les estimations proviennent d'une taille d'échantillon inférieure à 30, ou présentent des coefficients de variation très élevés, supérieurs à 33,3 %.</p> <p>Statistique Canada recommande de ne pas diffuser d'estimations de qualité inacceptable. Si un utilisateur choisit cependant de le faire, ces estimations devraient alors être signalées à l'aide de la lettre F (ou d'un quelconque identificateur similaire) et devraient être accompagnées de la mise en garde suivante :</p> <p>« Nous informons l'utilisateur que ces estimations (désignées avec la lettre F) ne respectent pas les normes de qualité de Statistique Canada. Les conclusions qui reposeront sur ces données ne seront pas fiables et seront très probablement invalides. »</p>

9.6 Seuils pour la diffusion des estimations pour l'Enquête canadienne sur l'expérience des soins de santé primaires

Le tableau ci-dessous fournit une indication de la précision des estimations démographiques ainsi que les seuils de diffusion associés aux trois niveaux de qualité de l'estimation présentés à la section précédente. Ces seuils proviennent des tables de coefficients de variation (CV) dont il sera question au chapitre 10.0.

Par exemple, d'après le tableau, la qualité d'une estimation pondérée de 145 000 femmes possédant une caractéristique donnée chez les 65 ans et plus est médiocre.

Veillez noter que ces seuils de diffusion correspondent aux estimations de chiffres de population seulement. Dans le cas d'estimations de rapports, les utilisateurs ne devraient pas utiliser la valeur du numérateur (ni le dénominateur) afin de trouver le niveau de qualité de l'estimation correspondant. La règle 4 à la section 10.1 ainsi que l'exemple 4 à la section 10.1.1 expliquent la bonne procédure à suivre dans le cas d'un rapport.

Groupe d'âge	Sexe	CV acceptable 0,0 à 16,5 %	CV médiocre 16,6 à 33,3 %	CV inacceptable > 33,3 %
Moins de 65 ans	Hommes	690 000 et plus	180 000 à < 690 000	moins de 180 000
	Femmes	605 000 et plus	155 000 à < 605 000	moins de 155 000
	Tous	665 000 et plus	167 000 à < 665 000	moins de 167 000
65 ans et plus	Hommes	381 000 et plus	111 000 à < 381 000	moins de 111 000
	Femmes	330 000 et plus	92 000 à < 330 000	moins de 92 000
	Tous	382 000 et plus	101 000 à < 382 000	moins de 101 000
Tous	Hommes	654 000 et plus	168 000 à < 654 000	moins de 168 000
	Femmes	560 000 et plus	142 000 à < 560 000	moins de 142 000
	Tous	616 000 et plus	155 000 à < 616 000	moins de 155 000

10.0 Tables de variabilité d'échantillonnage approximative

Afin de fournir des coefficients de variation (CV) qui pourraient s'appliquer à une gamme étendue d'estimations catégoriques produites à partir de ce fichier de microdonnées et auxquels il serait facilement possible pour l'utilisateur d'avoir accès, un ensemble de tables de variabilité d'échantillonnage approximative a été produit. Ces tables de CV permettent à l'utilisateur d'obtenir un coefficient de variation approximatif fondé sur la taille de l'estimation calculée à partir des données de l'enquête.

Les coefficients de variation sont calculés à l'aide de la formule de la variance pour un échantillonnage aléatoire simple et en y incorporant un facteur qui reflète la nature du plan d'échantillonnage, qui est à plusieurs degrés et qui prévoit la formation de grappes. Ce facteur, appelé l'effet du plan, a été déterminé en calculant premièrement les effets du plan pour une gamme étendue de caractéristiques, puis en choisissant parmi ceux-ci une valeur prudente, dans ce cas-ci le 75^e percentile, à utiliser à l'intérieur des tables de CV qui s'appliqueraient ensuite à l'ensemble entier des caractéristiques.

Le tableau ci-dessous indique la valeur prudente des effets du plan, ainsi que les tailles de l'échantillon et les chiffres de population selon le groupe d'âge et le sexe, qui ont été utilisés pour produire les tables de variabilité d'échantillonnage approximative de l'Enquête canadienne sur l'expérience des soins de santé primaires (ECE-SSP).

Groupe d'âge	Sexe	Effet du plan	Taille de l'échantillon	Population
Moins de 65 ans	Hommes	1,45	757	10 539 613
	Femmes	1,51	913	10 558 375
	Tous	1,48	1 670	21 097 988
65 ans et plus	Hommes	1,40	196	1 822 869
	Femmes	1,53	328	2 254 479
	Tous	1,47	524	4 077 348
Tous	Hommes	1,45	953	12 362 482
	Femmes	1,54	1 241	12 812 854
	Tous	1,51	2 194	25 175 336

Tous les coefficients de variation inclus dans les tables de variabilité d'échantillonnage approximative sont approximatifs et donc non officiels. Des estimations de la variance réelle pour des variables précises peuvent être obtenues auprès de Statistique Canada, contre remboursement des frais. Étant donné que le CV approximatif est une estimation prudente, l'utilisation de la variance réelle estimée pourrait faire passer l'estimation d'un niveau de qualité à un autre. Par exemple, une estimation *médiocre* pourrait devenir *acceptable* si elle était fondée sur le calcul du CV exact.

Rappelez-vous que : Si le nombre d'observations sur lesquelles une estimation est basée est inférieur à 30, l'estimation pondérée est très probablement inacceptable et Statistique Canada recommande de ne pas diffuser une telle estimation, quelle que soit la valeur du coefficient de variation.

10.1 Comment utiliser les tables de coefficients de variation pour des estimations catégoriques

Les règles qui suivent devraient permettre à l'utilisateur de déterminer les coefficients de variation approximatifs à partir des tables de variabilité d'échantillonnage approximative pour des estimations du nombre, de la proportion ou du pourcentage de membres de la population visée par l'enquête possédant une certaine caractéristique et pour des rapports et des différences entre de telles 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 la table de variabilité d'échantillonnage approximative pour le groupe âge/sexes approprié, repérez le nombre estimé dans la colonne la plus à gauche (intitulée « Numérateur du pourcentage ») et suivez les astérisques (le cas échéant) jusqu'au premier chiffre rencontré. Ce chiffre est le coefficient de variation approximatif.

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

Le coefficient de variation d'une proportion estimée ou d'un pourcentage estimé dépend à la fois de la taille de la proportion ou du pourcentage et de la taille du total sur lequel la proportion ou le pourcentage repose. Les proportions estimées ou les pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque la proportion ou le pourcentage repose sur un sous-groupe de la population. La proportion, par exemple, des personnes prenant des médicaments régulièrement qui avaient subi des effets secondaires au cours des 12 derniers mois est plus fiable que le nombre estimé de personnes prenant des médicaments régulièrement qui avaient subi des effets secondaires au cours des 12 derniers mois.

Lorsque la proportion ou que le pourcentage repose sur la population totale visée par la table, le CV de la proportion ou du pourcentage est le même que le CV du numérateur de la proportion ou du pourcentage. Dans ce cas, la règle 1 peut être appliquée.

Lorsque la proportion ou que le pourcentage repose sur un sous-ensemble de la population totale (p. ex. les personnes souffrant d'une maladie chronique), on devrait faire référence à la proportion ou au pourcentage (dans le haut de la table) et au numérateur de la proportion ou du pourcentage (dans la colonne de gauche de la table). L'intersection de la rangée et de la colonne appropriées donne le coefficient de variation.

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

L'erreur-type d'une différence entre deux estimations est approximativement égale à la racine carrée de la somme des carrés de chaque erreur-type considérée séparément. C'est-à-dire que l'erreur-type d'une différence ($\hat{d} = \hat{X}_1 - \hat{X}_2$) 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. Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}}/\hat{d}$. Cette

formule est exacte pour la différence entre des caractéristiques distinctes et non corrélées, mais n'est autrement qu'approximative.

Règle 4 : Estimations de rapports

Dans le cas où le numérateur est un sous-ensemble du dénominateur, le rapport devrait être converti en un pourcentage et la règle 2 appliquée. Cela s'appliquerait, par exemple, au cas où le dénominateur est le nombre de personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois et le numérateur, le nombre de personnes qui, au cours des 12 derniers mois, avaient éprouvé des difficultés à obtenir ces services dont ils avaient besoin.

Dans le cas où le numérateur n'est pas un sous-ensemble du dénominateur, comme par exemple le rapport du nombre de personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois comparativement au nombre de personnes ayant eu besoin de soins immédiats pour un problème de santé mineur au cours de la même période, l'erreur-type du rapport des estimations est approximativement égale à la racine carrée de la somme des carrés de chaque coefficient de variation considéré séparément, multipliée par \hat{R} . L'erreur-type d'un rapport ($\hat{R} = \hat{X}_1 / \hat{X}_2$) 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 de \hat{X}_2 respectivement. Le coefficient de variation de \hat{R} est donné par $\sigma_{\hat{R}} / \hat{R}$. La formule tendra à surestimer l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélés positivement et à la sous-estimer si \hat{X}_1 et \hat{X}_2 sont corrélés négativement.

Règle 5 : Estimations de différences entre des rapports

Dans ce cas, les règles 3 et 4 sont combinées. On détermine premièrement les CV pour les deux rapports à l'aide de la règle 4, puis on trouve le CV de leur différence au moyen de la règle 3.

10.1.1 Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques

Les exemples ci-dessous utilisent des données du fichier de l'ECE-SSP et sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter.

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

Supposons qu'un utilisateur estime à 8 914 814 le nombre de personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Reportez-vous à la table des coefficients de variation pour le CANADA – Tous les âges.
- 2) L'agrégat estimé 8 914 814 ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, c'est-à-dire 9 000 000.

- 3) On trouve le coefficient de variation pour un agrégat estimé en se reportant à la première entrée autre que des astérisques sur cette rangée, c'est-à-dire 3,4 %.
- 4) Le coefficient de variation approximatif de l'estimation est donc 3,4 %. Le résultat selon lequel il y aurait 8 914 814 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1) de personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois précédant l'enquête peut donc être publié sans réserve.

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

Supposons qu'un utilisateur estime à $2\,362\,719 / 8\,914\,814 = 26,5\%$ la proportion de personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois qui déclareraient avoir eu de la difficulté à obtenir ces services. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Reportez-vous à la table des coefficients de variation pour le CANADA – Tous les âges.
- 2) Parce que l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c'est-à-dire les personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois), il faut utiliser à la fois le pourcentage (26,5 %) et le numérateur du pourcentage (2 362 719) pour déterminer le coefficient de variation.
- 3) Le numérateur, 2 362 719, ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, soit 2 000 000. De même, l'estimation du pourcentage ne figure dans l'en-tête d'aucune colonne; il faut donc utiliser la proportion qui s'en rapproche le plus, soit 25,0 %.
- 4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 8,1 %, est le coefficient de variation à employer.
- 5) Le coefficient de variation approximatif de l'estimation est donc 8,1 %. Le résultat selon lequel 26,5 % des personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois déclareraient avoir eu de la difficulté à obtenir ces services peut être publié sans réserve.

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

Supposons qu'un utilisateur estime que la proportion de personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois qui déclareraient avoir eu de la difficulté à obtenir ces services était de $1\,942\,741 / 7\,991\,443 = 24,3\%$ parmi les personnes ayant déclaré avoir un médecin habituel et de $419\,978 / 923\,371 = 45,5\%$ parmi les personnes n'ayant pas un médecin habituel. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

- 1) L'utilisation de la table des coefficients de variation du CANADA – Tous les âges de la façon décrite dans l'exemple 2 donne un CV de 8,1% pour l'estimation pour les personnes ayant un médecin habituel et un CV de 14,7% pour les personnes n'ayant pas un médecin habituel.

- 2) En utilisant la règle 3, l'erreur-type d'une différence ($\hat{d} = \hat{X}_1 - \hat{X}_2$) 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 (personnes ayant un médecin habituel), \hat{X}_2 est l'estimation 2 (personnes n'ayant pas un médecin habituel) et α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et de \hat{X}_2 respectivement.

C'est-à-dire que l'erreur-type de la différence $\hat{d} = 0,243 - 0,455 = -0,212$ est :

$$\begin{aligned} \sigma_{\hat{d}} &= \sqrt{[(0,243)(0,081)]^2 + [(0,455)(0,147)]^2} \\ &= \sqrt{(0,000387) + (0,004474)} \\ &= 0,070 \end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,070 / 0,212 = 0,330$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 33,0 %. La différence entre les estimations est considérée médiocre (tout juste au seuil de l'inacceptable) et Statistique Canada recommande de publier cette estimation accompagnée de la lettre E (ou un autre identificateur semblable) et d'une mise en garde avertissant les utilisateurs subséquents du niveau élevé d'erreur dû à l'échantillonnage y étant associé.

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime qu'au cours des 12 derniers mois, 8 914 814 de personnes avaient eu besoin de services de santé de routine ou de suivi, comparativement à 7 414 864 de personnes qui avaient eu besoin de soins immédiats nécessaires pour un problème de santé mineur. L'utilisateur est intéressé à comparer ces deux estimations sous la forme d'un rapport. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation d'un rapport, où le numérateur de l'estimation (\hat{X}_1) est le nombre de personnes qui avaient eu besoin de services de santé de routine ou de suivi. Le dénominateur de l'estimation (\hat{X}_2) est le nombre de personnes qui avaient eu besoin de soins immédiats nécessaires pour un problème de santé mineur.
- 2) Reportez-vous à la table des coefficients de variation pour le CANADA.
- 3) Le numérateur de cette estimation de rapport est 8 914 814. Le chiffre qui s'en rapproche le plus est 9 000 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 3,4 %.
- 4) Le dénominateur de cette estimation de rapport est 7 414 864. Le chiffre qui s'en rapproche le plus est 7 000 000. On trouve le coefficient de variation pour cette

estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 4,2 %.

- 5) Le coefficient de variation approximatif de l'estimation du rapport est donc donné par la règle 4, qui est :

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

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

C'est-à-dire que :

$$\begin{aligned} \alpha_{\hat{R}} &= \sqrt{(0,034)^2 + (0,042)^2} \\ &= \sqrt{0,001156 + 0,001764} \\ &= 0,054 \end{aligned}$$

- 6) Le rapport obtenu entre les personnes ayant eu besoin de services de santé de routine ou de suivi et celles ayant eu besoin de soins immédiats nécessaires pour un problème de santé mineur est 8 914 814 / 7 414 864, c'est-à-dire 1,20 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1). Le coefficient de variation de cette estimation est 5,4 %, ce qui fait qu'on peut la diffuser sans réserve.

Exemple 5 : Estimations de différences de rapports

Supposons qu'un utilisateur estime que le rapport entre les personnes ayant eu besoin de services de santé de routine ou de suivi et celles ayant eu besoin de soins immédiats nécessaires pour un problème de santé mineur au cours des 12 derniers mois s'établit à 1,06 pour les hommes tandis qu'il est de 1,34 pour les femmes. L'utilisateur est intéressé à comparer les deux rapports pour voir s'il y a une différence statistique entre ceux-ci. Comment l'utilisateur détermine-t-il le coefficient de variation de cette différence?

- 1) Tout d'abord calculez le coefficient de variation approximatif pour le rapport chez les hommes (\hat{R}_1) et le rapport chez les femmes (\hat{R}_2) tel qu'il est décrit dans l'exemple 4. Il faut pour cela utiliser respectivement les tables « Hommes, tous les âges » et « Femmes, tous les âges ». Le CV approximatif pour le rapport chez les hommes est 7,8 % et celui chez les femmes est 6,7 %.
- 2) En utilisant la règle 3, l'erreur-type d'une différence ($\hat{d} = \hat{R}_1 - \hat{R}_2$) est :

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

où α_1 et α_2 sont les coefficients de variation de \hat{R}_1 et \hat{R}_2 respectivement. C'est-à-dire que l'erreur-type de la différence $\hat{d} = 1,06 - 1,34 = -0,28$ est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(1,06)(0,078)]^2 + [(1,34)(0,067)]^2} \\ &= \sqrt{(0,006836) + (0,008060)} \\ &= 0,122\end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,122 / (-0,28) = -0,436$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 43,6 %. La différence entre les estimations est considérée inacceptable et Statistique Canada recommande de ne pas la publier. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre F (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les utilisateurs subséquents en garde contre les hauts taux d'erreur associés à l'estimation.

10.2 Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance

Bien que les coefficients de variation soient beaucoup utilisés, l'intervalle de confiance d'une estimation est une mesure plus intuitivement significative de l'erreur d'échantillonnage. Un intervalle de confiance constitue une déclaration du niveau de confiance selon laquelle la valeur vraie pour la population se situe à l'intérieur d'une gamme précisée de valeurs. Par exemple, un intervalle de confiance de 95 % peut être décrit comme suit :

Si l'échantillonnage de la population est répété indéfiniment, chaque échantillon menant à un nouvel intervalle de confiance pour une estimation, l'intervalle englobera alors dans 95 % des échantillons la valeur vraie de la population.

En utilisant l'erreur-type d'une estimation, des intervalles de confiance pour des estimations peuvent être obtenues en partant de l'hypothèse qu'au terme d'un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique donnée de la population se répartiront normalement 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 pour 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. Ces différents degrés de confiance sont désignés sous le nom de niveaux de confiance.

Des intervalles de confiance pour une estimation \hat{X} sont généralement exprimés sous forme de deux chiffres, un inférieur et un supérieur à l'estimation, comme étant $(\hat{X} - k, \hat{X} + k)$, où k est déterminé suivant le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

Des intervalles de confiance pour une estimation peuvent être calculés directement à partir des tables de variabilité d'échantillonnage approximative, en déterminant d'abord à partir de la table appropriée le coefficient de variation de l'estimation \hat{X} , puis en utilisant la formule suivante pour le convertir à un intervalle de confiance ($IC_{\hat{x}}$) :

$$IC_{\hat{x}} = (\hat{X} - t\hat{X}\alpha_{\hat{x}}, \hat{X} + t\hat{X}\alpha_{\hat{x}})$$

où $\alpha_{\hat{x}}$ est le coefficient de variation déterminé de \hat{X} , et

- $t = 1$ si l'on désire un intervalle de confiance de 68 %;
- $t = 1,6$ si l'on désire un intervalle de confiance de 90 %;
- $t = 2$ si l'on désire un intervalle de confiance de 95 %;
- $t = 2,6$ si l'on désire un intervalle de confiance de 99 %.

Nota : Les lignes directrices pour la diffusion des estimations s'appliquent également aux intervalles de confiance. S'il est impossible, par exemple, de diffuser une estimation, on ne peut alors pas non plus communiquer un intervalle de confiance.

10.2.1 Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance

Un intervalle de confiance de 95 % pour la proportion estimée de personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois qui déclareraient avoir eu de la difficulté à obtenir ces services (d'après l'exemple 2 à la section 10.1.1) serait calculé comme suit :

$$\hat{X} = 26,5 \text{ \% (ou exprimé sous forme de proportion 0,265)}$$

$$t = 2$$

$\alpha_{\hat{x}} = 8,1 \text{ \% (0,081 exprimé sous forme de proportion)}$ est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.

$$IC_{\hat{x}} = \{0,265 - (2) (0,265) (0,081), 0,265 + (2) (0,265) (0,081)\}$$

$$IC_{\hat{x}} = \{0,265 - 0,043, 0,265 + 0,043\}$$

$$IC_{\hat{x}} = \{0,222, 0,308\}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 22,2 % et 30,8 % des personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois avaient eu de la difficulté à obtenir ces services.

10.3 Comment utiliser les tables de coefficients de variation pour effectuer un test t

Des erreurs-types peuvent aussi être utilisées pour effectuer des tests d'hypothèses, une procédure destinée à distinguer des paramètres d'une population à l'aide d'estimations d'un échantillon. Ces estimations peuvent être des chiffres, des moyennes, des pourcentages, des rapports, etc. Les tests peuvent être effectués à divers niveaux de signification, où un niveau de signification est la probabilité de conclure que les caractéristiques sont différentes quand, en fait, elle sont identiques.

Supposons que \hat{X}_1 et \hat{X}_2 sont des estimations d'un échantillon pour deux caractéristiques qui nous intéressent. Supposons également que l'erreur-type de la différence $\hat{X}_1 - \hat{X}_2$ est $\sigma_{\hat{d}}$.

Si $t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ se situe entre -2 et 2, aucune conclusion à propos de la différence entre les caractéristiques n'est alors justifiée au niveau de signification de 5 %. Si, cependant, ce rapport est inférieur à -2 ou supérieur à +2, la différence observée est significative au niveau de 0,05. C'est-à-dire que la différence entre les estimations est significative.

10.3.1 Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t

Supposons que l'utilisateur désire tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y a pas de différence entre la proportion de personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois qui déclaraient avoir eu de la difficulté à obtenir ces services chez les personnes ayant déclaré avoir un médecin habituel et chez les autres personnes. D'après l'exemple 3 à la section 10.1.1, il s'est avéré que l'erreur-type de la différence entre ces deux estimations était 0,070. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0,243 - 0,455}{0,070} = \frac{-0,212}{0,070} = -3,03$$

Puisque $t = -3,03$ est inférieur à -2, il faut en conclure qu'il existe une différence significative entre les deux estimations au niveau de signification de 0,05.

10.4 Coefficients de variation pour des estimations quantitatives

Il faudrait produire des tables spéciales afin de déterminer l'erreur d'échantillonnage d'estimations quantitatives, ce qui n'a pas été fait, parce que la plupart des variables pour l'ECE-SSP sont principalement de nature catégorique.

En général cependant, le coefficient de variation d'un total quantitatif sera supérieur au coefficient de variation de l'estimation de la catégorie correspondante (c'est-à-dire l'estimation du nombre de personnes retenues dans l'estimation quantitative). S'il est impossible de diffuser l'estimation de la catégorie correspondante, on ne pourra pas non plus communiquer l'estimation quantitative. Par exemple, le coefficient de variation du nombre total de fois où les gens ont personnellement utilisé le service d'urgence d'un hôpital au cours des 12 derniers mois serait supérieur au coefficient de variation de la proportion correspondante de personnes les ayant utilisées. Si, par conséquent, le coefficient de variation de la proportion est inacceptable (rendant la proportion non diffusable), il en sera de même du coefficient de variation de l'estimation quantitative correspondante (rendant cette estimation quantitative non diffusable).

Des coefficients de variation de telles estimations peuvent être calculés, au besoin, pour une estimation précise à l'aide d'une technique appelée pseudo-répétition, ce qui veut dire diviser les enregistrements inclus dans les fichiers de microdonnées en sous-groupes (ou répétitions) et déterminer la variation à l'intérieur de l'estimation de répétition en répétition. Les utilisateurs qui désirent calculer des coefficients de variation pour des estimations quantitatives peuvent communiquer avec Statistique Canada afin d'obtenir des conseils sur l'allocation d'enregistrements à des répétitions appropriées et sur les formules à employer à l'intérieur de ces calculs.

10.5 Tables des coefficients de variation

Consulter le fichier ECE-SSP2006-2007_CVTabF.pdf pour les tables de coefficients de variation.

11.0 Pondération

Puisque l'Enquête canadienne sur l'expérience des soins de santé primaires (ECE-SSP) faisait appel à un sous-échantillon de l'échantillon de l'Enquête sur la santé dans les collectivités canadiennes (ESCC), cycle 3.1, le calcul des poids pour les enregistrements des données de l'enquête est clairement lié à la procédure de pondération utilisée aux fins de l'ESCC. La procédure de pondération employée pour l'ESCC est décrite brièvement ci-dessous.

11.1 Procédures de pondération pour l'Enquête sur la santé dans les collectivités canadiennes

Dans le cas de l'ESCC 3.1, une base aréolaire et une base téléphonique ont été utilisées. Tel que mentionné à la section 5.4, seuls les répondants provenant de la base aréolaire étaient admissibles à l'ECE-SSP. Dans l'ESCC, les répondants de chacune des deux bases sont pondérés séparément avant que les deux bases ne soient intégrées. Le poids de départ pour l'ECE-SSP est donc le poids des répondants de l'ESCC sélectionnés, tel que calculé avant l'intégration des bases. Ce poids est donc censé permettre à l'échantillon de la base aréolaire de représenter adéquatement l'ensemble de la population canadienne ciblée par l'enquête. La stratégie de pondération des unités de la base aréolaire de l'ESCC est détaillée dans le Guide de l'utilisateur du fichier de microdonnées à grande diffusion de l'ESCC 3.1. Le poids final de la base aréolaire de l'ESCC 3.1 tient entre autre compte de la probabilité de sélection de chaque ménage, de la non-réponse au niveau des ménages, de la sélection de la personne choisie au sein du ménage, de même que de la non-réponse au niveau des personnes.

11.2 Procédures de pondération pour l'Enquête canadienne sur l'expérience des soins de santé primaires

Le poids de départ de l'ECE-SSP est le poids final de la base aréolaire de l'ESCC 3.1. Il est ajusté pour tenir compte de la sélection d'un échantillon restreint de répondants de l'ESCC et de la non-réponse à l'ECE-SSP. Les poids sont aussi ajustés pour contrer la présence de poids aberrants et pour s'assurer que les estimations de l'ECE-SSP concordent avec les projections démographiques pour certains sous-groupes de la population. Tous ces ajustements sont expliqués dans la présente section.

Sélection de l'échantillon de l'ECE-SSP

Le poids de départ, celui de l'ESCC 3.1, fournit une représentation adéquate de la population ciblée à condition d'inclure tous les répondants. Pour l'ECE-SSP, un échantillon de 3 800 répondants a été choisi aléatoirement parmi tous les répondants admissibles de la base aréolaire. L'échantillon a été choisi de façon indépendante dans chaque région sociosanitaire (RSS) à l'aide de l'échantillonnage aléatoire systématique. Ainsi, chaque répondant de la base aréolaire de l'ESCC âgé de 18 ans et plus à l'intérieur d'une même RSS avait la même probabilité d'être choisi. Le poids de sélection de l'ECE-SSP a été combiné au poids de départ fourni par l'ESCC de façon à s'assurer que la somme des poids de tous les répondants à l'intérieur de chaque RSS soit préservée.

Ainsi, le poids de sélection ajusté de l'ECE-SSP est donné par :

$$\text{Poids ECE - SSP} = \left(\frac{\text{Poids ESCC 3.1} * \sum \text{poids de tous les répondants de 18 ans et plus de la base aréolaire de l'ESCC 3.1 dans la RSS}}{\sum \text{poids des répondants choisis à l'ECE - SSP dans la RSS}} \right)$$

Ajustement pour la non-réponse

Pour différentes raisons, certaines personnes n'ont pas pu être interviewées pour l'ECE-SSP. Dans certains cas, une information de contact à jour n'a pu être obtenue. Dans d'autres cas, la période de collecte s'est terminée sans que le répondant ait pu être joint. D'autres personnes ont refusé de participer à l'enquête. Une partie de l'échantillon de départ de l'ECE-SSP est donc « perdu » et un facteur d'ajustement doit être appliqué aux poids des personnes répondantes pour compenser cette non-réponse.

Le poids des non-répondants est redistribué aux répondants à l'intérieur de groupes de réponse. Pour ce faire, la régression logistique est employée. Un modèle de prédiction de la probabilité de répondre ou non à l'enquête a été bâti à l'aide des variables disponibles pour toutes les personnes choisies à l'ECE-SSP. Étant donné la grande quantité d'information disponible de l'ESCC, un vaste éventail de possibilités s'offrait pour bâtir le modèle de réponse. À l'aide de ce modèle, les répondants ont été divisés en neuf groupes selon leur probabilité de réponse à l'enquête. Des groupes de taille égale ont été créés, sauf pour le groupe avec les plus faibles probabilités de réponse, qui avait deux fois la taille des huit autres. Chaque non-répondant a ensuite été ajouté au groupe qui correspondait à sa propre probabilité de réponse. Dans chaque groupe, le poids des répondants est ensuite augmenté d'un facteur correspondant à la somme des poids de toutes les unités dans le groupe de réponse divisé par le poids de tous les répondants dans le groupe.

La possibilité d'ajuster séparément pour la non-réponse de différents types (pas d'information de contact, etc.) a été étudiée mais comme cette façon de procéder ne semblait pas permettre un meilleur ajustement, un seul ajustement a été effectué.

Contrôle des poids aberrants

Étant donné la série d'ajustements successifs appliqués sur les poids des répondants, d'abord par l'ESCC puis par l'ECE-SSP, il est possible que certaines unités se retrouvent avec des poids se démarquant des poids des autres répondants du même groupe démographique au point même de devenir aberrants. Certains répondants peuvent effectivement représenter une proportion anormalement élevée de leur groupe et ainsi influencer fortement les estimations pour ces groupes. Afin d'éviter cette situation, le poids des répondants qui contribuent de façon aberrante à leur groupe démographique est ajusté à la baisse selon une méthode dite de « winsorization ». Étant donné la taille d'échantillon réduite, cet ajustement a dû être limité à des groupes assez larges. Les groupes utilisés étaient composés de :

- six régions : les provinces de l'Atlantique, le Québec, l'Ontario, l'Alberta, la Colombie-Britannique et le Yukon, et finalement les autres provinces et territoires;
- quatre groupes d'âge : 18 à 34 ans, 35 à 49 ans, 50 à 64 ans et 65 ans et plus;
- le sexe du répondant.

Très peu d'unités ont vu leur poids être ainsi modifié.

Post-stratification

La dernière étape nécessaire afin d'obtenir le poids final de l'ECE-SSP est la post-stratification. La post-stratification est appliquée afin d'assurer que la somme des poids finaux corresponde aux estimations de population pour chacun des 48 groupes mentionnés ci haut (six régions, quatre groupes d'âge et le sexe du répondant). Les estimations de population en date du 20 janvier 2007 ont été utilisées pour la post-stratification. Le poids final de l'ECE-SSP est donné par :

$$\text{Poids final} = \left(\frac{\text{Poids winsorié} * \text{projection démographique pour le groupe auquel appartient le répondant}}{\sum \text{poids de tous les répondants dans le groupe}} \right)$$

Le poids qui en résulte WTPS (pour « final weight » en anglais) est le poids final qui figure dans le fichier partagé de microdonnées de l'ECE-SSP.

12.0 Questionnaire

Consulter le fichier ECE-SSP2006-2007_QuestF.pdf pour la version française du questionnaire utilisé pour la collecte des données.

13.0 Cliché d'enregistrement à valeurs univariées

Consulter le fichier ECE-SSP2006-2007_LvCds.pdf pour le cliché d'enregistrement à chiffres univariés.