

Enquête sur la santé dans les collectivités canadiennes (ESCC) - Vieillesse en santé

Guide de l'utilisateur

Décembre 2010

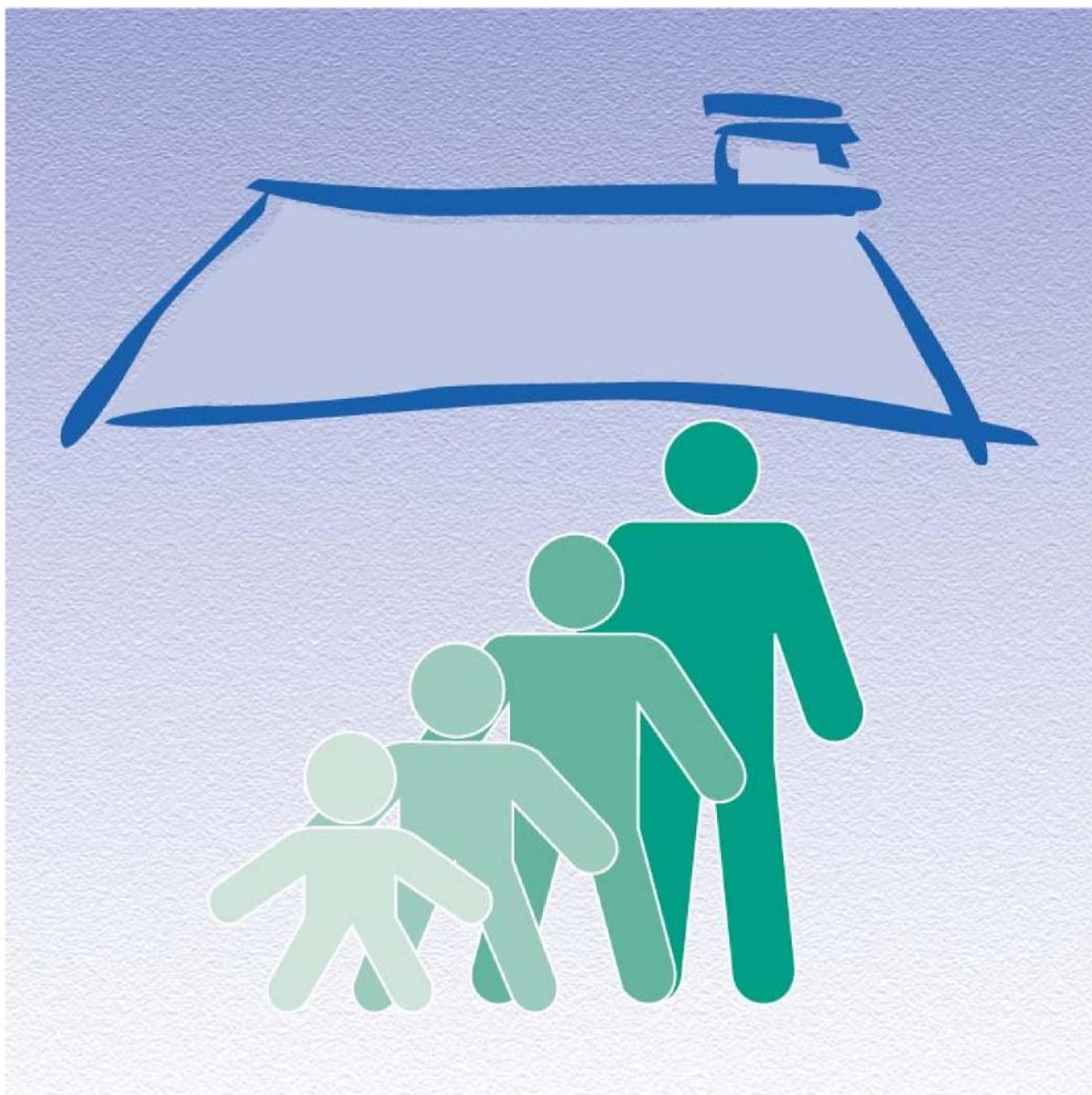


TABLE DES MATIÈRES

1.0	Introduction.....	1
2.0	Contexte et objectifs.....	2
2.1	Contexte de l'Enquête sur la santé dans les collectivités canadiennes (ESCC)	2
2.2	Contexte de l'Enquête sur la santé dans les collectivités canadiennes – Vieillesse en santé	3
2.3	Objectifs.....	3
3.0	Contenu de l'enquête	4
3.1	Essais qualitatifs.....	5
3.2	Enquête pilote	6
3.3	Contenu du questionnaire final	7
4.0	Plan d'échantillonnage	24
4.1	Population cible	24
4.2	Taille et répartition de l'échantillon.....	24
4.3	Plan d'échantillonnage.....	25
4.3.1	Sélection des régions géographiques	25
4.3.2	Sélection des ménages à l'intérieur de chaque grappe.....	26
4.3.3	Échantillonnage des répondants à l'intérieur de chaque ménage	27
5.0	Collecte des données	29
5.1	Interview assistée par ordinateur	29
5.2	Développement de l'application de l'ESCC – Vieillesse en santé.....	30
5.3	Formation des intervieweurs.....	31
5.4	Interview	31
5.5	Opérations sur le terrain.....	33
5.6	Contrôle de la qualité et gestion de la collecte	33
6.0	Traitement des données.....	35
6.1	Contrôle.....	35
6.1.1	Vérification du module des activités physiques.....	35
6.2	Codage	35
6.2.1	Codage du module sur la cognition	36
6.3	Création de variables dérivées	36
6.4	Interviews partielles	37
6.5	Pondération	37
7.0	Pondération	38
7.1	Pondération de l'échantillon	38

7.1.1	Poids initial	38
7.1.2	Retrait des unités hors champ	39
7.1.3	Non-réponse à l'échelle du ménage	39
7.1.4	Création du poids-personne	40
7.1.5	Non-réponse à l'échelle de la personne	40
7.1.6	Méthode d'estimation de Winsor	41
7.1.7	Calage aux marges	41
8.0	Qualité des données.....	42
8.1	Taux de réponse	42
8.1.1	Formules pour le calcul des taux de réponse et exemple	44
8.1.2	Restrictions dans le calcul du taux de réponse.....	45
8.2	Erreurs dans l'enquête.....	45
8.2.1	Erreurs d'échantillonnage	46
8.2.2	Erreurs non dues à l'échantillonnage	46
9.0	Lignes directrices pour la totalisation, l'analyse et la diffusion	48
9.1	Lignes directrices pour l'arrondissement.....	48
9.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation	49
9.2.1	Définitions : estimations de type nominal; estimations quantitatives.....	49
9.2.2	Totalisation d'estimations de type nominal	50
9.2.3	Totalisation d'estimations quantitatives	51
9.3	Lignes directrices pour l'analyse statistique	51
9.4	Lignes directrices pour la diffusion	52
10.0	Tableaux de la variabilité d'échantillonnage approximative.....	54
10.1	Comment utiliser les tableaux de CV pour les estimations catégorielles	54
10.2	Exemples d'utilisation des tableaux de CV pour des estimations catégorielles ...	57
10.3	Comment utiliser les tableaux de CV pour calculer les limites de confiance.....	60
10.4	Exemple d'utilisation de tableaux de CV pour obtenir des intervalles de confiance	61
10.5	Comment utiliser les tableaux de CV pour effectuer un test Z.....	61
10.6	Exemple d'utilisation des tableaux de CV pour effectuer un test Z	62
10.7	Variances ou coefficients de variation exacts.....	62
10.8	Seuils pour la diffusion des estimations relatives à l'ESCC.....	64
11.0	Fichiers de microdonnées : description, accès et utilisation	65
11.1	Fichiers maîtres.....	65
11.1.1	Centre de données de recherche.....	65
11.1.2	Totalisations personnalisées.....	65
11.1.3	Télé-accès	65
11.2	Fichiers de partage	66
11.3	Fichiers de microdonnées à grande diffusion	66

11.4	Utilisation de la variable de pondération	67
11.5	Convention appliquée pour nommer les variables.....	67
11.6	Convention appliquée pour nommer les variables des enquêtes de l'ESCC précédentes.....	70
11.7	Lignes directrices pour l'utilisation des variables d'un sous-échantillon.....	70
11.8	Dictionnaires de données	71
11.9	Différences dans le calcul des variables de contenu commun fondé sur . différents fichiers.....	72

1.0 Introduction

L'Enquête sur la santé dans les collectivités canadiennes – Vieillessement en santé (ESCC–VS) est une enquête transversale qui a pour but de recueillir des renseignements sur les facteurs, les influences et les processus qui contribuent au vieillissement en santé dans le cadre d'une approche multidisciplinaire axée sur des déterminants sanitaires, sociaux et économiques.

L'ESCC–VS a recueilli les réponses de personnes âgées de 45 ans et plus, qui habitent dans les logements privés dans les dix provinces. Les résidents des trois territoires, les personnes vivant dans les réserves indiennes ou sur les terres de la Couronne, les résidents d'établissements, les membres à temps plein des Forces canadiennes et les résidents de certaines régions éloignées sont exclus de la base de sondage. Les données ont été recueillies entre décembre 2008 et novembre 2009.

L'objectif du présent document est de faciliter la manipulation des fichiers de microdonnées de l'ESCC–VS et de décrire la méthodologie utilisée. L'Enquête sur la santé dans les collectivités canadiennes – Vieillessement en santé produit trois types de fichiers de microdonnées : les fichiers-maîtres, les fichiers partagés et les fichiers de microdonnées à grande diffusion (FMGD). Les caractéristiques de chaque type de fichier sont expliquées dans le présent guide. Les FMGD seront diffusés en 2011.

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

Service d'aide aux utilisateurs des produits électroniques : 1-800-949-9491

Totalisations personnalisées ou renseignements généraux sur les données
Services personnalisés à la clientèle, Division de la statistique de la santé : 613-951-1746
Courriel : hd-ds@statcan.gc.ca

Aide pour accès à distance : 613-951-1653
Courriel : cchs-escc@statcan.gc.ca
Télécopieur : 613-951-0792

2.0 Contexte et objectifs

2.1 Contexte de l'Enquête sur la santé dans les collectivités canadiennes (ESCC)

En 1991, le Groupe de travail national sur l'information en matière de santé a cité un certain nombre de préoccupations et de problèmes touchant le système d'information sur la santé. Les membres du groupe estimaient que les données étaient fragmentées et incomplètes, qu'elles ne pouvaient être facilement partagées, qu'elles n'étaient pas analysées dans leur pleine mesure et que les résultats de la recherche n'atteignaient pas les Canadiens de façon régulière¹.

En réponse à ces préoccupations, l'Institut canadien d'information sur la santé (ICIS), Statistique Canada et Santé Canada ont uni leurs forces afin de créer un Carnet de route de l'information sur la santé. C'est à partir de ce projet qu'est née l'Enquête sur la santé dans les collectivités canadiennes. Le format, le contenu et les objectifs de l'ESCC ont évolué dans le cadre de vastes consultations avec les principaux experts, de même qu'avec les partenaires fédéraux, provinciaux et communautaires des régions sociosanitaires afin de déterminer les besoins de ces derniers en matière de données².

Pour répondre à de nombreux besoins en données, l'ESCC a eu un cycle de collecte de deux ans. Jusqu'au remaniement en 2007, la première année du cycle d'enquête, désignée par « .1 », consistait en une enquête générale sur la santé de la population, visant à fournir des estimations fiables à l'échelle de la région sociosanitaire. La deuxième année du cycle d'enquête, désignée par « .2 », avait un plus petit échantillon et visait à fournir des résultats à l'échelle provinciale sur des domaines particuliers de la santé.

Nouvelles désignations pour les cycles .1 et .2

Depuis 2007, l'ESCC régionale recueille des données sur une base continue. Pour éviter toute confusion avec les enquêtes axées sur la santé, on a cessé d'utiliser les désignations .1 et .2 pour différencier les deux composantes. Désormais, le cycle x.1 de l'ESCC est appelé « composante annuelle » de l'ESCC. Le titre au long est « Enquête sur la santé dans les collectivités canadiennes - Composante annuelle, 2008 » et le titre abrégé est simplement « ESCC – 2008 ».

La composante du contenu ciblé demeure la même. Elle continuera d'étudier dans le détail des populations ou des sujets plus précis. Cette composante sera désignée par le nom de l'enquête suivi du sujet des thèmes couverts par chaque enquête (p. ex. « Enquête sur la santé dans les collectivités canadiennes – Vieillessement en santé », ou « ESCC-Vieillessement en santé », ou « ESCC-VS »).

¹ 1999. *Health Information Roadmap: Responding to Needs*, Santé Canada, Statistique Canada. p. 3.

² 1999. *Health Information Roadmap: Beginning the Journey*. Institut canadien d'information sur la santé /Statistique Canada. ISBN 1-895581-70-2. p. 19.

2.2 Contexte de l'Enquête sur la santé dans les collectivités canadiennes – Vieillessement en santé

L'Enquête sur la santé dans les collectivités canadiennes – Vieillessement en santé (ESCC–VS) est le troisième cycle à contenu ciblé de l'ESCC. La santé mentale et le bien-être en 2002 (cycle 1.2) et la nutrition en 2004 (cycle 2.2) constituaient les deux cycles précédents. L'Enquête canadienne sur les mesures de la santé (ECMS) a remplacé le cycle de 2006 de l'enquête.

En janvier 2006, la Division de la statistique de la santé a mené une enquête auprès des partenaires primaires afin de déterminer leurs besoins en matière de données pour 2008-2009 et le futur contenu ciblé des cycles de l'ESCC. Le processus de consultation portait sur la priorité donnée par les partenaires aux divers thèmes en ce qui a trait à la pertinence, aux enjeux, au choix du moment et aux sujets englobés par ces thèmes. Parmi les partenaires participants, citons Santé Canada, l'Agence de la santé publique du Canada, des représentants des provinces et des régions sociosanitaires, et des membres du milieu de la recherche. À la suite de ces consultations, le vieillissement en santé a été choisi comme thème pour le contenu ciblé de l'ESCC de 2008-2009.

2.3 Objectifs

Les objectifs de l'enquête sont les suivants :

- Mieux comprendre le processus de vieillissement des Canadiens âgés de 45 ans et plus, grâce à la collecte de données sur divers aspects de leur santé et de leur bien-être, l'utilisation des services de soins de santé, le soutien social et l'appartenance à un réseau social, de même que le travail et la transition vers la retraite.
- Examiner comment les facteurs du mode de vie affectent la santé des personnes à mesure qu'elles vieillissent.
- Examiner les liens entre le vieillissement en santé et les variables ou caractéristiques sociales, démographiques, géographiques et économiques au moyen d'une approche multidisciplinaire.
- Fournir des renseignements sur le vieillissement en santé selon le groupe d'âge et le sexe.

3.0 Contenu de l'enquête

La présente section décrit le processus de consultation utilisé pour élaborer le contenu de l'enquête et résume le contenu final qui allait être inclus dans l'ESCC – Vieillessement en santé.

Le contenu de l'ESCC – Vieillessement en santé a été élaboré par suite d'un processus de consultation en plusieurs étapes entre la Division de la statistique de la santé de Statistique Canada, Santé Canada et l'Agence de la santé publique du Canada, avec la contribution importante d'experts chargés de l'Étude longitudinale canadienne sur le vieillissement (ELCV). Les consultations ont aussi fait intervenir des représentants de Ressources humaines et Développement social Canada et des ministères provinciaux et territoriaux de la Santé.

La conception du questionnaire a commencé à l'automne 2006 et a exigé des consultations exhaustives avec les intervenants. Des révisions du contenu ont été effectuées jusqu'en juin 2008.

La sélection du contenu a été fondée sur les critères suivants :

- question(s) présentant des lacunes statistiques d'après les résultats des consultations avec les intervenants;
- preuves solides que le vieillissement exerce une influence importante sur la santé et la qualité de vie, la morbidité et la mortalité prématurée;
- nombre important de personnes visées par la question cible;
- répercussions significatives sur les familles, les collectivités et les coûts des soins de santé;
- données susceptibles d'entraîner une amélioration de la santé, par suite d'une intervention stratégique;
- question(s) jugée(s) prioritaire(s) pour le soutien/l'élaboration des programmes et de la politique, des exigences en matière de surveillance et de la recherche; et
- question(s) considérée(s) comme pertinente(s) pour le cadre du vieillissement en santé.

3.1 Essais qualitatifs

On a procédé à un examen du questionnaire et à des essais qualitatifs en anglais et en français, par l'entremise du Centre de ressources en conception de questionnaires (CRCQ) de Statistique Canada. Le questionnaire a fait l'objet d'essais qualitatifs en trois étapes, qui se sont tenues en mars, mai et juin 2007 et qui ont pris la forme d'interviews individuelles et de groupes de discussion.

Au total, 52 interviews individuelles qualitatives ont été menées à Ottawa, Gatineau et Montréal. Les questionnaires ont été administrés en personne aux répondants. Ces interviews individuelles ont porté sur les quatre étapes du processus cognitif de réponse : compréhension de la question et des catégories de réponses, remémoration/recherche des renseignements demandés, réflexion au sujet de la réponse et décision quant aux données à déclarer, et réponse. Les résultats de chaque étape des essais qualitatifs ont joué un rôle de premier plan pour la révision du contenu du questionnaire, en vue des rondes subséquentes d'essais qualitatifs.

Outre les interviews individuelles, le CRCQ a aussi tenu des groupes de discussion. Au total, huit (8) groupes de discussion ont été menés en anglais et en français à Ottawa, Calgary et Montréal. On a demandé aux participants des groupes de discussion leurs réactions et leur rétroaction concernant l'enquête proposée. Dans l'ensemble, les participants des groupes de discussion ont eu des réactions et des attitudes favorables à l'égard d'une enquête sur le vieillissement en santé. Par ailleurs, selon les participants, aucun des sujets possibles de l'enquête n'était particulièrement délicat.

Toutes les interviews qualitatives ont été menées par des intervieweurs formés au CRCQ et observés par des membres de l'équipe de projet de l'ESCC – Vieillesse en santé. Certains des principaux résultats des essais qualitatifs sont abordés ci-après.

Principaux résultats des essais qualitatifs

Dans l'ensemble, les participants ont indiqué que le questionnaire était simple et facile à comprendre, mais long. En général, les répondants ne croyaient pas que les questions étaient de nature trop délicate, sauf les questions sur les réseaux sociaux, les actifs et héritages et la solitude. Les essais qualitatifs ont aussi fait ressortir qu'il était possible d'évaluer la cognition dans une enquête auprès de la population. Les répondants se sont montrés prêts à accomplir les tâches. Selon les résultats, même si la tâche de mémorisation et celle de l'alternance étaient assez complexes pour certains répondants, elles n'ont pas causé trop de stress ni d'anxiété.

Par ailleurs, les essais qualitatifs ont fourni des lignes directrices utiles sur la façon d'adapter certains instruments, comme l'échelle d'activité physique des personnes âgées (PASE) et le Multidimensional Functional Assessment Questionnaire®, Older American Resources and Services (OARS), à une interview sur place assistée par ordinateur. Pendant les essais qualitatifs, les sections sur la prestation de soins et les soins reçus ont

été considérées comme très longues et exigeantes. Des modifications ont été apportées pour simplifier ces modules et préciser ce qui devait être déclaré.

3.2 Enquête pilote

Une enquête pilote a été menée en novembre et au début de décembre 2007 à Halifax, Montréal, Toronto et Winnipeg. La base de sondage de l'Enquête sur la santé dans les collectivités canadiennes de 2005 (ESCC 3.1) a servi à sélectionner environ 1 000 répondants. Ces derniers ont été interviewés à domicile au moyen de la méthode d'interview sur place assistée par ordinateur (IPAO). Les objectifs de l'essai pilote consistaient à observer la réaction des répondants à l'enquête, à obtenir des estimations du temps consacré aux divers modules de contenu, à déterminer l'efficacité de la formation des intervieweurs et de la stratégie de communication, à vérifier les procédures et les documents de formation, à fournir une indication préliminaire des taux de réponse, à mettre à l'essai de façon plus poussée le questionnaire informatisé et à évaluer la faisabilité d'utiliser l'ESCC – Vieillessement en santé comme cohorte de départ d'une enquête longitudinale.

Le nombre et la longueur des modules du questionnaire élaborés pour l'enquête pilote ont dépassé largement la capacité de Statistique Canada d'administrer le questionnaire dans la durée d'interview prévue. Pour l'enquête pilote, le contenu a été réparti en deux (2) questionnaires (enquête pilote A et enquête pilote B). Étant donné la possibilité d'une enquête longitudinale, tout le contenu proposé a été vérifié, avec comme objectif, par suite de l'enquête pilote, de sélectionner certains modules pour l'enquête principale, et de reporter les modules qui restent aux vagues futures d'une enquête longitudinale.

La sélection des modules du contenu final de la version principale de l'ESCC – Vieillessement en santé a été fondée sur plusieurs facteurs, y compris :

- l'analyse des résultats de l'enquête pilote, y compris la rétroaction reçue des intervieweurs et du personnel sur le terrain;
- les contraintes liées à l'administration du questionnaire;
- l'intervention des experts et la rétroaction reçue des intervenants, y compris Santé Canada, l'Agence de la santé publique du Canada et l'Étude longitudinale canadienne sur le vieillissement qui ont soumis des commentaires particuliers et des priorités; et
- la décision finale de concevoir l'enquête principale comme une enquête transversale.

Par suite de l'essai pilote, des décisions ont été prises en vue d'exclure les modules plus longs, complexes ou de nature plus délicate. Les questions sur les réseaux sociaux, les actifs et les héritages, les antécédents de travail, la retraite du conjoint et la cognition autodéclarée ont été supprimées, afin de réduire la fatigue pour le répondant et le fardeau de réponse.

3.3 Contenu du questionnaire final

La présente section fait état des modules qui constituent le contenu du questionnaire de l'ESCC – Vieillesse en santé. Le questionnaire comprenait 37 modules, en excluant les modules d'entrée et de sortie. Le tableau 3.1 fournit une description sommaire de chaque module.

Tableau 3.1 Description sommaire des modules

	Module	Description sommaire du module et source	Commentaires additionnels
1	Interview par procuration (GR)	Ce module a servi à recueillir des données concernant les interviews par procuration et comprenait des questions concernant l'identité du répondant par procuration et les raisons de la tenue d'une telle interview. Source : Enquête sur la santé dans les collectivités canadiennes de 2008.	On a autorisé l'interview par procuration auprès d'un autre membre du ménage si le répondant sélectionné était incapable de participer en raison d'un problème de santé mentale ou physique.
2	État de santé général (GEN)	Le module sur l'état de santé général a servi à recueillir des données sur l'état de santé perçu, l'état de santé mentale perçu, le stress perçu et le sentiment d'appartenance à la communauté locale. Source : Enquête sur la santé dans les collectivités canadiennes de 2008.	Le module a été modifié pour utiliser une question différente sur la satisfaction à l'égard de la vie (GEN_Q02AA).
3	Sommeil 2 (SL2)	Dans ce module, on a posé aux répondants une question sur la difficulté à s'endormir ou à rester endormi. Source : Enquête sur la santé dans les collectivités canadiennes de 2008.	
4	Taille et poids –	Dans ce module, on a	

	Module	Description sommaire du module et source	Commentaires additionnels
	autodéclarés (HWT)	<p>demandé aux répondants leur taille et leur poids, ainsi que s'ils considéraient faire de l'embonpoint, être trop maigres ou être à peu près normaux.</p> <p>Les données servent à calculer l'indice de masse corporelle (IMC), qui peut être utilisé pour déterminer si un répondant est trop maigre ou obèse.</p> <p>Source : Enquête sur la santé dans les collectivités canadiennes de 2008.</p>	
5	Problèmes de santé chroniques (CCC)	<p>Ce groupe de questions portait sur les problèmes de santé de longue durée diagnostiqués par un professionnel de la santé, ayant duré ou qui pourraient durer au moins six mois.</p> <p>Source : Enquête sur la santé dans les collectivités canadiennes de 2008.</p>	Ce module a été modifié, afin d'inclure aussi les problèmes de santé davantage présents dans les populations plus âgées : angine, ostéoporose, maladie de Parkinson, cataracte et maladie de la glande thyroïde.
6	Indice de l'état de santé (HUI)	<p>Ce module comprenait sept ensembles de questions. Chacun portait sur des aspects différents de la santé : vision, ouïe, élocution, mobilité, dextérité, sentiments et cognition (mémoire et pensée). Les réponses aux sept ensembles de questions, plus les questions du module HUP sur les douleurs et malaise, sont combinés pour obtenir un score qui indique l'état de santé fonctionnel global.</p>	

	Module	Description sommaire du module et source	Commentaires additionnels
		Source : Enquête sur la santé dans les collectivités canadiennes de 2008.	
7	Douleurs et malaise (HUP)	Ce module comprenait trois questions concernant les douleurs et malaise. Ces questions correspondent à la huitième dimension de l'Indice de l'état de santé. Source : Enquête sur la santé dans les collectivités canadiennes de 2008.	
8	Échelle de satisfaction à l'égard de la vie (SLS)	Dans ce module, on demandait aux répondants s'ils étaient satisfaits de leur vie globalement. Les réponses à ce module sont utilisées pour calculer un score de satisfaction à l'égard de la vie pour le répondant. Source : Échelle de satisfaction à l'égard de la vie élaborée par Ed Diener (Université de l'Illinois).	
9	Cognition (COG)	L'objectif du module sur la cognition était d'évaluer le fonctionnement cognitif des répondants, ainsi que de déterminer les changements graduels liés au vieillissement. Le module sur la cognition comportait quatre tâches chronométrées. <ul style="list-style-type: none"> • La première tâche était un test de mémoire : on demandait au répondant de se remémorer une liste de mots. • La deuxième tâche avait pour but de tester les 	Des procédures de codage et de pondération distinctes ont été requises pour ce module. Pour plus de renseignements, voir la section 6.2.1 « Codage du module sur la cognition ».

	Module	Description sommaire du module et source	Commentaires additionnels
		<p>fonctions exécutives. On demandait au répondant de nommer autant d'animaux qu'il pouvait en une minute.</p> <ul style="list-style-type: none"> • La troisième tâche, le test d'alternance mentale, comprenait trois sous-tâches : compter de 1 à 20, réciter l'alphabet et alterner des chiffres et des lettres (1A, 2B, 3C, etc.). • Dans la quatrième tâche, on demandait au répondant de se remémorer la liste de mots entendus à la première tâche. <p>Sources :</p> <ul style="list-style-type: none"> • Pour le test de mémoire, on a utilisé la « Liste de mots de REY », une version modifiée du Rey Auditory Verbal Learning Test (RAVLT). • Le test des animaux est un test couramment utilisé dans de nombreuses batteries uniformisées pour la mesure de la cognition. • Le test d'alternance mentale (MAT)© est modélisé sur le Trail making test, qui constitue une mesure courante utilisée par les neuropsychologues. Le MAT a été développé par Evelyn Teng (University of Southern California). Statistique Canada a reçu l'autorisation de 	

	Module	Description sommaire du module et source	Commentaires additionnels
		M ^{me} Teng pour utiliser cet instrument dans l'enquête.	
10	Activités physiques 2 (PA2)	<p>Les questions de ce module ont été légèrement modifiées par rapport à l'échelle d'activité physique des personnes âgées (PASE) et portaient sur les activités légères, modérées et ardues, et sur les exercices faits au cours des sept derniers jours, ainsi que sur l'activité physique pendant les tâches ménagères, le bénévolat et le travail, au cours des sept derniers jours. Les réponses à ces questions ont servi à calculer le score d'activité physique du répondant.</p> <p>D'autres questions se sont ajoutées à la suite du PASE (à partir de PA2_Q11A) et visaient à déterminer si l'activité déclarée au cours des 7 derniers jours représentait le niveau d'activité habituel des 12 derniers mois, si le répondant voulait participer davantage à des activités physiques et, le cas échéant, qu'est-ce qui l'avait empêché de le faire.</p> <p>Source : Ce module représente une modification de l'échelle de l'activité physique des personnes âgées (PASE)© des New England Research Institutes. Statistique Canada a obtenu les droits pour l'utilisation de cet instrument dans</p>	<p>Au cours du traitement des données, on a découvert que plusieurs réponses dans le champ « Autre – veuillez préciser » ont été mal codées. Un examen exhaustif a été effectué. Pour plus de renseignements, voir la section 6.1.1 « Vérification du module des activités physiques ».</p> <p>Voir l'annexe A pour les changements particuliers apportés à l'instrument original.</p>

	Module	Description sommaire du module et source	Commentaires additionnels
		l'enquête. La version française est fondée sur le libellé utilisé dans l'Étude longitudinale québécoise sur la nutrition comme déterminant du vieillissement réussi (NuAge).	
11	Risque nutritionnel (NUR)	<p>Dans ce module, on a utilisé l'outil de sélection Seniors in the Community: Risk Evaluation for Eating and Nutrition (SCREEN) pour interroger les répondants concernant leur changement de poids; leurs habitudes alimentaires; les problèmes alimentaires; la consommation de fruits, de légumes et de liquides, ainsi que la satisfaction éprouvée lors des repas. Les réponses ont servi à calculer un score de risque nutritionnel.</p> <p>Source : Ce module est une version modifiée de SCREEN©, élaboré par Heather Keller (Université de Guess) et est utilisé avec la permission de l'auteur.</p>	
12	Santé bucco-dentaire 3 (OH3)	Dans ce module, on a demandé aux répondants de décrire la santé de leur bouche, leur port de prothèses, le brossage des dents et l'utilisation de soie dentaire, ainsi que de déclarer la fréquence à laquelle ils évitaient certains aliments et éprouvaient un inconfort lorsqu'ils mangeaient tout aliment, en raison de problèmes buccaux. On a aussi inclus des	

	Module	Description sommaire du module et source	Commentaires additionnels
		<p>questions concernant les visites chez un professionnel des soins dentaires.</p> <p>Source : La majorité des questions ont été tirées de l'Enquête canadienne sur les mesures de la santé (ECMS) (module OHM) et de l'Enquête sur la santé dans les collectivités canadiennes de 2008 (modules OH2 et DEN).</p>	
13	Consommation de médicaments (MED)	<p>Ce module comportait des questions concernant les médicaments sur ordonnance et sans ordonnance pris par le répondant au cours du dernier mois. Une question de suivi a été ajoutée pour déterminer si les répondants utilisaient un médicament particulier sur une base quotidienne.</p> <p>Source : Enquête sur la santé dans les collectivités canadiennes de 2005.</p>	
14	Usage de suppléments alimentaires – Vitamines et minéraux (DS2)	<p>Les questions de ce module portaient sur la fréquence de consommation de suppléments de multivitamines, de calcium, ainsi que de vitamines D et de B12.</p> <p>Source : Nouveau module fondé sur l'Enquête sur la santé dans les collectivités canadiennes de 2008 et l'Enquête canadienne sur les mesures de la santé.</p>	
15	Usage du tabac (SMK)	Ce module comprenait une série de questions concernant	

	Module	Description sommaire du module et source	Commentaires additionnels
		<p>l'usage actuel et passé du tabac, y compris la fréquence du tabagisme, le nombre de cigarettes fumées par jour et l'âge auquel le répondant a commencé à fumer quotidiennement.</p> <p>Source : Enquête sur la santé dans les collectivités canadiennes de 2008 et Enquête canadienne sur les mesures de la santé.</p>	
16	Consommation d'alcool (ALC)	<p>Ce module portait sur la consommation d'alcool et la fréquence de l'abus d'alcool.</p> <p>Source : Enquête sur la santé dans les collectivités canadiennes de 2008.</p>	
17	Changements faits pour améliorer la santé (CIH)	<p>Ce module comprenait une série de questions concernant les changements de comportement adoptés par les répondants au cours de la dernière année pour améliorer leur santé, y compris la perte de poids et le renoncement au tabac. Dans le module, on cherchait également à connaître les facteurs qui les avaient empêchés de changer leur comportement, y compris le manque de temps ou le manque de volonté.</p> <p>Source : Enquête sur la santé dans les collectivités canadiennes de 2008.</p>	
18	Chutes (FAL)	<p>Les questions de ce module portaient sur les chutes subies au cours des 12 derniers mois où le</p>	<p>Ce module était destiné uniquement aux répondants de 65 ans et plus.</p>

	Module	Description sommaire du module et source	Commentaires additionnels
		<p>répondant s'était blessé assez gravement pour limiter une partie ou la totalité de ses activités normales. On a aussi posé des questions plus détaillées sur les soins à la suite de la chute et la peur de tomber à nouveau.</p> <p>Source : Nouveau module. Contenu similaire à celui de l'Enquête sur la santé dans les collectivités canadiennes de 2008 (module sur les blessures), de la US Health and Retirement Study et de l'Enquête sur le vieillissement en santé et la retraite en Europe. Les questions concernant la crainte ou la peur de tomber sont fondées sur les travaux originaux de Tennstedt S. et coll., 1998.</p>	
19	Activités instrumentales de la vie quotidienne (IAL)	<p>Ce module est une version modifiée du Multidimensional Functional Assessment Questionnaire, Older American Resources and Services (OARS). On a posé aux répondants une série de questions concernant leur capacité à exercer seuls diverses activités quotidiennes, comme utiliser le téléphone, voyager, magasiner, cuisiner, faire les tâches ménages, prendre des médicaments et s'occuper de leurs finances.</p> <p>Source : Version modifiée du Multidimensional Functional Assessment Questionnaire©,</p>	

	Module	Description sommaire du module et source	Commentaires additionnels
		Older American Resources and Services (OARS). Gerda G. Fillenbaum (Université Duke) a autorisé l'utilisation de cet instrument.	
20	Activités de base de la vie quotidienne (ADL)	Cette version modifiée du Multidimensional Functional Assessment Questionnaire, Older American Resources and Services (OARS) visait à déterminer la capacité du répondant à accomplir certaines activités de base de la vie quotidienne, comme se nourrir, s'habiller, soigner son apparence, marcher, se lever et se coucher, se laver, et s'il avait des problèmes d'incontinence. Source : Voir IAL ci-dessus.	
21	Utilisation des soins de santé 2 (HC2)	Ce module comportait une série de questions concernant l'utilisation des soins de santé, comme le fait d'avoir un médecin traitant, la consultation de spécialistes et le recours aux praticiens de médecines douces. Source : Modifié à partir de l'Enquête sur la santé dans les collectivités canadiennes de 2008 (module HCU).	
22	Soins reçus 1 (CR1)	Dans ce module, on demandait aux répondants s'ils avaient reçu des services de soins à domicile fournis par des professionnels en raison d'un problème de santé affectant leurs activités quotidiennes. On a demandé aux	Par « services de soins professionnels à domicile », on entend les soins de santé, les services d'entretien ménager ou d'autres services de soutien reçus à domicile en raison de problèmes de santé ou de limitations, peu

	Module	Description sommaire du module et source	Commentaires additionnels
		<p>répondants de déclarer l'aide fournie par des travailleurs et des organismes rémunérés pour diverses activités, comme les soins médicaux, les soins personnels, le ménage, le transport et la préparation ou la livraison des repas. On a aussi posé des questions aux répondants concernant les soins à domicile dont ils avaient besoin mais qu'ils ne recevaient pas.</p> <p>Source : Nouveau module, contenu similaire à l'Enquête sociale générale (ESG), cycle 21 et à l'Enquête sur la santé dans les collectivités canadiennes de 2008 (module HMC).</p>	<p>importe la source de financement de ces services.</p> <p>Dans l'ESCC de 2008, on posait des questions sur les soins de santé, les services d'entretien ménager ou d'autres services de soutien reçus à domicile dont le coût était entièrement ou partiellement financé par le gouvernement.</p>
23	Soins reçus 2 (CR2)	<p>Dans ce module, on demandait aux répondants de déclarer l'aide fournie par des membres de la famille, des amis ou des voisins (aidants naturels), en raison d'un problème de santé ayant des répercussions sur les activités quotidiennes.</p> <p>On a demandé aux répondants d'indiquer l'aide reçue pour diverses activités, comme les soins médicaux, les soins personnels, le ménage, le transport, ainsi que la préparation ou la livraison des repas.</p> <p>Le module comprenait une série de questions liées aux soins non officiels reçus à</p>	

	Module	Description sommaire du module et source	Commentaires additionnels
		<p>domicile, comme l'identité de la personne apportant l'aide, ainsi que la durée et l'intensité des soins reçus.</p> <p>Source : Nouveau module, contenu similaire à celui de l'Enquête sociale générale (ESG) (cycles 16 et 21).</p>	
24	Soutien social – Disponibilité (SSA)	<p>Ce module comprenait une série de questions détaillées sur la disponibilité du soutien social. Les questions de ce module ont été combinées pour déterminer la probabilité que le répondant ait accès à du soutien social.</p> <p>Source : Enquête sur la santé dans les collectivités canadiennes de 2008.</p>	
25	Participation sociale (SPA)	<p>Ce module portait sur le type et la fréquence de la participation à diverses activités sociales, comme des activités sportives et des services religieux, ainsi que sur les limitations d'une telle participation.</p> <p>Source : Nouveau module.</p>	
26	Prestations de soins (CAG)	<p>Dans ce module, on demandait au répondant s'il avait aidé d'autres personnes ayant un problème de santé ou une limitation au cours des 12 derniers mois.</p> <p>Les questions de ce module portaient sur des sujets liés à l'aide prodiguée aux autres, comme le type d'aide prodiguée, le nombre total de personnes aidées, des</p>	

	Module	Description sommaire du module et source	Commentaires additionnels
		<p>renseignements sur la personne à laquelle le plus d'aide a été apportée, les répercussions de l'aide prodiguée sur le travail et la santé, ainsi que les aspects positifs ou négatifs du fait de prodiguer de l'aide.</p> <p>Source : Nouveau module, contenu similaire à celui de l'Enquête sociale générale (ESG) (cycles 16 et 21).</p>	
27	Dépenses liées à la prestation de soins (CGE)	<p>Ce sous-module de la prestation de soins (CAG) traitait des dépenses personnelles encourues par les répondants à cause de la prestation de soins, y compris les achats, les services, le transport et les dépenses supplémentaires, comme pour le logement ou la nourriture.</p> <p>Source : Nouveau module, contenu similaire à celui de l'Enquête sur la participation et les limitations d'activités (EPLA).</p>	
28	Dépression (DEP)	<p>Ce module comprenait une série de questions détaillées sur la dépression.</p> <p>Source : Enquête sur la santé dans les collectivités canadiennes de 2008.</p>	<p>Il est recommandé que l'analyse se limite à l'examen de la dépression comme corrélat d'autres comportements et caractéristiques en matière de santé. On n'encourage pas l'utilisation des données pour calculer la prévalence.</p>
29	Solitude (LON)	<p>Ce module a été tiré du score de solitude à trois items. On</p>	

	Module	Description sommaire du module et source	Commentaires additionnels
		<p>a posé des questions aux répondants concernant le manque de compagnie, le sentiment de rejet et le sentiment d'isolement. On a par la suite calculé un score de solitude.</p> <p>Source : Score de solitude à trois items élaboré par Mary Elizabeth Hugues (Johns Hopkins Bloomberg School of Public Health). L'auteure a autorisé l'utilisation de l'instrument.</p>	
30	Transport (TRA)	<p>Les questions du module portent sur la possession d'un permis de conduire, les moyens de transport les plus utilisés et l'accès au transport en commun.</p> <p>Source : Nouveau module, module similaire dans l'ESCC de 2001 et dans l'Enquête sociale générale (ESG), cycle 19.</p>	
31	Activité sur le marché du travail (LBF)	<p>Ce module a servi à recueillir des données concernant la vie professionnelle des répondants, y compris s'ils étaient occupés, en chômage ou à la retraite. Des questions concernant la profession, les raisons de ne pas occuper un emploi et l'horaire habituel de travail ont aussi été posées.</p> <p>Source : Modifié à partir de l'Enquête sur la santé dans les collectivités canadiennes de 2005.</p>	<p>Les questions de ce module sont posées uniquement aux répondants de 45 à 74 ans.</p> <p>Une question a été ajoutée pour tenir compte des personnes qui n'ont jamais travaillé.</p>
32	Raisons de la retraite (RET)	<p>Les questions de ce module portaient sur l'âge de la retraite, les principales</p>	<p>Les questions de ce module sont posées uniquement aux</p>

	Module	Description sommaire du module et source	Commentaires additionnels
		<p>raisons de la retraite, la participation à la vie active et la capacité de travailler.</p> <p>Ce module portait aussi sur la retraite partielle dans le cas des répondants ayant officiellement pris leur retraite, mais continuant de travailler ou étant en retraite progressive.</p> <p>Source : Nouveau module, contenu similaire à celui de l'Enquête sociale générale (ESG) (cycles 16 et 21).</p>	répondants de 45 à 85 ans.
33	Planification de la retraite (RPL)	<p>Les questions de ce module portaient sur la planification de la retraite : âge souhaité de la retraite, planification de la retraite, pensions de retraite et plans élaborés avant la retraite.</p> <p>Source : Nouveau module, contenu similaire à celui de l'Enquête sociale générale (ESG) (cycles 16 et 21).</p>	Les questions de ce module sont posées uniquement aux répondants de 45 à 85 ans qui ne sont pas à la retraite ou qui sont partiellement à la retraite.
34	Propriétaires (OWN)	<p>Ce module comprenait des questions modifiées de l'Enquête sur les dépenses des ménages, de l'Enquête sur le vieillissement et l'autonomie et de l'Enquête sociale générale, cycle 21. On a demandé aux répondants s'ils étaient propriétaires de leur logement, la valeur de leur propriété et le montant de leur hypothèque.</p> <p>Source : Nouveau module. Modifié à partir de l'Enquête</p>	

	Module	Description sommaire du module et source	Commentaires additionnels
		sur les dépenses des ménages, l'Enquête sur le vieillissement et l'autonomie et l'Enquête sociale générale (ESG), cycle 21.	
35	Revenu (INC)	<p>Dans ce module, on a demandé aux répondants d'indiquer leur revenu personnel et le revenu du ménage, afin de déterminer toutes les sources de revenu du ménage et de revenu personnel, ainsi que la principale source.</p> <p>Des questions de suivi ont été posées concernant les prestations du Régime de pensions du Canada ou du Régime de rentes du Québec. On a aussi demandé aux répondants d'estimer la valeur de leurs épargnes et investissements et si leur revenu couvrait leurs dépenses de base.</p> <p>Source : Modifié à partir de l'Enquête sur la santé dans les collectivités canadiennes de 2008, avec l'ajout de nouvelles questions.</p>	
36	Renseignements sociodémographiques (SDC)	<p>Ce module a servi à recueillir des données sociales et démographiques, y compris le statut d'immigrant, le pays de naissance, l'origine ethnique et la langue.</p> <p>Source : Enquête sur la santé dans les collectivités canadiennes de 2008.</p>	
37	Renseignements administratifs	Ce module a permis de recueillir des données	Les questions du sous-module du CLS ont été

	Module	Description sommaire du module et source	Commentaires additionnels
	(ADM)	<p>administratives, y compris l'autorisation du couplage et du partage des données.</p> <p>Ce module comportait un sous-module appelé « Consentement à partager » (CLS). Les répondants ont reçu une trousse d'information et un formulaire de consentement qu'ils devaient passer en revue et remplir concernant la communication des données sur le contact et/ou de l'enquête à trois universités qui gèrent une initiative appelée « Étude longitudinale canadienne sur le vieillissement ».</p> <p>Source : Enquête sur la santé dans les collectivités canadiennes de 2008 et nouveau sous-module (CLS).</p>	posées uniquement aux répondants de 45 à 85 ans.

4.0 Plan d'échantillonnage

4.1 Population cible

« L'ESCC – Vieillessement en santé » vise la population de 45 ans et plus vivant dans des logements privés et résidant dans les dix provinces. Sont exclues du champ de l'enquête les résidents des trois territoires, les personnes vivant sur les réserves indiennes et les terres de la Couronne, les résidents des établissements, les membres à temps pleins des Forces canadiennes et les personnes vivant dans certaines régions éloignées.

4.2 Taille et répartition de l'échantillon

La population cible originale était 55 ans et plus. L'échantillon supplémentaire de 5 000 répondants âgés entre 45 et 54 ans a été financé par l'Étude longitudinale canadienne sur le vieillissement. On souhaitait un échantillon de 32 000 unités déclarantes sur une année pour réaliser les objectifs de l'enquête, soit estimer le vieillissement en santé parmi les Canadiens âgés de 45 ans et plus dans des domaines d'intérêt particuliers pour chaque province et compte tenu du budget accordé à l'enquête. Selon les 32 000 répondants, 5 000 d'entre eux auraient entre 45 et 54 ans et 27 000, 55 ans et plus. Le but était de produire des estimations fiables par province selon cinq groupes d'âge (45-54, 55-64, 65-74, 75-84 et 85+) et le sexe.

Une stratégie en deux étapes a servi à répartir l'échantillon entre les provinces. Cent vingt-cinq unités d'échantillonnage ont d'abord été rattachées à chaque domaine d'intérêt (dix groupes selon âge-sexe) dans chaque province. On a ainsi attribué à chaque province 1 250 unités dans un premier temps, soit au total 12 500 unités. Les autres 19 500 unités ont été réparties entre les provinces selon une méthode de répartition de puissance où $q=0,7^3$. La taille de l'échantillon total d'une province donnée correspondait à la somme des tailles obtenues aux deux étapes.

Afin d'assurer une représentation urbaine et rurale adéquate dans chaque province, on a subséquemment réparti l'échantillon en deux strates : les strates urbaines et rurales. Celles-ci sont définies à partir des îlots du Recensement de 2006. L'échantillon provincial a ainsi été réparti selon la strate urbaine et la strate rurale proportionnellement au nombre de logements comptant des personnes âgées de 45 ans et plus au sein de chaque strate. On a ensuite accru les effectifs de l'échantillon avant la collecte des données pour tenir compte du taux prévu de non-réponse et des logements hors champ. Les logements hors champ étaient tout particulièrement importants dans cette enquête puisque les ménages échantillonnés seraient rejetés s'il n'y avait personne de la population cible (de 45 ans et plus). La taille brute de l'échantillon requise de 32 000 répondants était estimée à 52 010. Le tableau 4.1 donne les tailles brutes de l'échantillon annuel pour l'ESCC – Vieillessement en santé par province et statut urbain/rural.

³ Bankier, M.D. (1988). Power Allocations: Determining Sample Sizes for Subnational Areas. *The American Statistician*, 42, 174-177.

Tableau 4.1 Tailles brutes de l'échantillon par province et statut urbain/rural

Province	Urbain	Rural	Total
Terre-Neuve-et-Labrador	1,885	1,320	3,205
Île-du-Prince-Édouard	1,397	1,160	2,557
Nouvelle-Écosse	2,201	1,560	3,761
Nouveau-Brunswick	2,016	1,440	3,456
Québec	7,490	1,316	8,806
Ontario	9,762	1,280	11,042
Manitoba	3,045	740	3,785
Saskatchewan	2,555	1,079	3,634
Alberta	4,585	675	5,260
Colombie-Britannique	5,904	600	6,504
Canada	40,840	11,170	52,010

4.3 Plan d'échantillonnage

L'échantillon des répondants de l'ESCC – Vieillesse en santé était sélectionné selon un plan d'échantillonnage à trois étapes. À la première étape, des régions géographiques ou grappes ont été sélectionnées. À la deuxième, un échantillon de ménages a été sélectionné à l'intérieur de chaque grappe et finalement, un répondant par ménage a été choisi au hasard. Chacune de ces étapes est expliquée dans les sections qui suivent.

4.3.1 Sélection des régions géographiques

Afin de contrôler les coûts et d'assurer une distribution raisonnable des interviews sur le terrain puisqu'on a utilisé des interviews sur place, on a créé des groupes de ménages à partir de la proximité géographique. Ces groupes s'appellent des grappes. Les grappes ont été créées afin que chaque logement au Canada fasse partie d'une (et d'une seule) grappe. Les îlots du Recensement de 2006 étaient utilisés comme des unités élémentaires pour les grappes. Pour créer des régions plus grandes, on a regroupé des îlots contigus en utilisant les critères suivants (d'après les chiffres du Recensement de 2006) :

- Au moins 100 logements par grappe;
- Un nombre minimum de logements ayant au moins un membre âgé de 75 à 84 ans et de 85 ans et plus (ce nombre variait par province et statut urbain/rural);
- Pas plus d'une certaine superficie pour chaque grappe (encore une fois, ce nombre variait par province et statut urbain/rural);
- Pas de groupements d'îlots de recensement urbains et ruraux dans la même grappe, et pas de grappes qui franchissent des frontières provinciales;
- Accent mis sur l'accessibilité (pouvoir se déplacer d'un point de la grappe à un autre sans en sortir) et la compacité.

Un niveau différent d'importance pourrait être assigné à chacun de ces critères. Les grappes ne répondaient pas toutes à l'ensemble des critères, mais l'ensemble de grappes final s'approchait le plus de ces exigences. Au total, près de 17 000 grappes ont été créées.

Afin de déterminer le nombre de grappes à échantillonner, le nombre de logements à choisir dans chaque grappe devait être établi. Ceci a été fait de façon à équilibrer l'efficacité et la précision des coûts. Il y aurait un coût de collecte réduit si on sélectionnait moins de grappes et plus d'interviews dans chaque grappe, alors qu'il y aurait une meilleure précision si on sélectionnait plus de grappes et moins d'interviews dans chaque grappe parce que « l'effet des grappes », qui est attribuable au fait que les logements dans la même grappe ont tendance à être plus similaires que les logements en général, est réduit. Il a ainsi été déterminé que la taille d'un échantillon brut de 35 cas par grappe urbaine et de 20 cas par grappe rurale était raisonnable.

Pour ce faire, les grappes ont été divisées entre deux groupes (un processus appelé stratification) par province : les grappes urbaines et rurales. Dans chaque province, l'échantillon total a été divisé entre « urbain » et « rurale » d'après la proportion des personnes de 85 ans et plus dans chacun des deux groupes. Ensuite, en utilisant la cible des 35 et 20 cas par grappe urbaine et rurale, le nombre de grappes requis à échantillonner dans les régions urbaines et rurales de chaque province était dérivé.

La sélection des grappes s'est faite à l'aide de la méthode d'échantillonnage avec probabilité proportionnelle à la taille. Avec cette méthode, plus il y avait de personnes de 45 ans et plus dans une grappe, plus cette grappe avait de chances d'être sélectionnée. Cette méthode avait pour but de mieux contrôler la variabilité introduite par la sélection de grappes de différentes tailles.

Finalement, chaque grappe a été assignée au hasard à une période de collecte. L'enquête qui s'est déroulée de décembre 2008 à novembre 2009 était divisée en six périodes de collecte de deux mois chacune. L'échantillon pour les deux premières périodes de collecte (de décembre 2008 à janvier 2009 et de février à mars 2009) a été réduit en raison de la diminution dans les ressources des interviewers. Il a été augmenté dans les autres périodes afin de compenser cet ajustement.

4.3.2 Sélection des ménages à l'intérieur de chaque grappe

Un des principaux défis de ce plan de sondage consistait à obtenir assez d'aînés (des Canadiens de 75 à 84 ans et de 85 ans et plus) dans l'échantillon cible, car la proportion d'aînés y était beaucoup plus élevée que dans la population. Plusieurs sources ont été considérées, mais on a décidé d'utiliser le Recensement de la population de 2006 comme base de sondage à partir de laquelle les logements seraient sélectionnés. L'information du Recensement pourrait ensuite servir à cibler des logements qui sont plus susceptibles d'avoir des personnes de 45 ans et plus.

Seuls les logements composés d'au moins une personne de 45 ans et plus selon la composition du ménage du Recensement (avec un ajustement d'âge qui reflète l'âge au moment de la collecte) étaient considérés.

Dans chaque grappe sélectionnée, les logements étaient divisés en trois strates :

- Ceux composés d'au moins une personne de 85 ans et plus;
- Ceux composés seulement de personnes de 55 et moins (mais au moins d'une personne de 45 à 54 ans);
- Tous les autres logements.

Les deux premières strates ont été créées afin que le nombre de personnes sélectionnées dans chaque groupe d'âge soit approprié. Sans cette stratification, l'échantillon aurait inclus trop de personnes du groupe 45 à 54 ans et pas assez de personnes du groupe 85 ans et plus.

L'échantillon de 35 logements dans les régions urbaines et de 20 logements dans les régions rurales a ensuite été réparti entre chacune des trois strates, avec une répartition variant selon la province. L'un des défis de cette méthode était que l'échantillon était sélectionné selon la composition des ménages au moment du Recensement, lequel avait eu lieu deux ans et demi à trois ans et demi plus tôt, soit avant la collecte de cette enquête. La composition de plusieurs de ménages a changé durant cette période. Afin de déterminer la répartition de l'échantillon parmi les trois strates, on a fait des simulations en utilisant la composition des ménages au moment du Recensement et ensuite, on a fait des changements dans la composition des ménages à partir des observations de l'enquête pilote. L'échantillon de l'enquête pilote provenait des répondants du cycle 3.1 de l'ESCC (2005), mais qui a eu lieu de deux ans à deux ans et demi plus tard. On a révisé cette répartition plus tard à partir des observations des deux premières périodes de collecte.

Le nombre de logements requis au sein de chaque strate de l'échantillon était obtenu par échantillonnage aléatoire simple.

4.3.3 Échantillonnage des répondants à l'intérieur de chaque ménage

Dans le cadre de l'ESCC – Vieillesse en santé, une personne par ménage a été choisie comme répondant. Au moment de la visite d'un logement choisi, on a obtenu la composition du ménage au moment de l'enquête. Des logements sans répondant admissible (des logements composés seulement de personnes de 44 ans et moins ou de personnes qui ne sont pas dans la population cible) ont été classés comme hors champ. Pour les autres ménages, un répondant a été sélectionné au hasard parmi tous les répondants admissibles selon des probabilités de sélection variées.

On a assigné à chaque membre d'un ménage de 45 ans et plus un facteur de probabilité de sélection d'après les cinq groupes d'âge pour lesquels des estimations étaient requises (45-54, 55-64, 65-74, 75-84 et 85+). Les probabilités de sélection variaient selon la

province pour atteindre le nombre de répondants visé au sein de chaque groupe d'âge. Les probabilités de sélection ont été créées au moyen d'une étude en simulation utilisant la composition du ménage de l'enquête pilote.

5.0 Collecte des données

La collecte de l'Enquête sur la santé dans les collectivités canadiennes – Vieillesse en santé s'est déroulée de décembre 2008 à novembre 2009. Au cours de la période de collecte, on a mené au total 30 865 interviews valides, au moyen d'un type d'interview assistée par ordinateur (IAO) appelée interview sur place assistée par ordinateur (IPAO).

5.1 Interview assistée par ordinateur

L'interview assistée par ordinateur (IAO) offre deux grands avantages par rapport aux autres méthodes de collecte des données. D'abord, la technique est étayée d'un système de gestion des cas et d'une fonctionnalité de transmission des données. Le système de gestion des cas enregistre automatiquement l'information de gestion importante sur chaque tentative visant un cas particulier et produit des rapports aux fins de la gestion du processus de collecte.

Le système de gestion des cas achemine les applications de questionnaires et les fichiers de l'échantillon du bureau central de Statistique Canada aux bureaux régionaux. Les données destinées au bureau central sont acheminées en sens inverse. Par souci de confidentialité, les données sont encodées avant la transmission. Elles sont ensuite désencodées une fois sauvegardées sur un ordinateur sécurisé distinct, sans accès à distance.

Deuxièmement, l'IAO permet de concevoir une interview personnalisée à l'intention de chaque répondant, en fonction de ses caractéristiques particulières et de ses réponses à l'enquête. Par exemple :

- l'application saute automatiquement les questions qui ne s'appliquent pas au répondant;
- des règles de vérification sont appliquées automatiquement pour repérer les réponses incohérentes ou non incluses dans la fourchette de valeurs permises, et des messages-guides s'affichent à l'écran quand une entrée est invalide;
- le répondant reçoit une rétroaction immédiate et l'intervieweur peut corriger toute incohérence;
- le libellé des questions est personnalisé automatiquement, y compris les périodes de référence et les pronoms, en fonction de facteurs comme l'âge et le sexe du répondant, la date de l'interview et les réponses aux questions précédentes.

5.2 Développement de l'application de l'ESCC – Vieillessement en santé

Pour l'ESCC – Vieillessement en santé, on a utilisé une application d'interview sur place assistée par ordinateur (IPAO). L'application comprenait des composantes Entrée, de contenu sur la santé (appelé C2) et Sortie.

Les composantes Entrée et Sortie comprennent des ensembles uniformes de questions auxquels l'intervieweur a pu se reporter pour prendre contact avec un répondant, recueillir de l'information importante sur l'échantillon, choisir les répondants et évaluer l'état des cas. La partie C2 est constituée des modules sur la santé et représente la plus grande partie de l'application. L'application a subi trois étapes d'essai : essai modulaire, essai intégré et essai de bout en bout.

L'essai modulaire consiste à mettre à l'essai indépendamment chaque module de contenu, afin de vérifier si les instructions « passez à », la logique d'enchaînement et le texte, dans les deux langues officielles, sont spécifiés correctement. Les instructions « passez à » et la logique d'enchaînement entre modules ne sont pas testées à cette étape, car chaque module est considéré comme un questionnaire autonome. Une fois que tous les modules sont vérifiés par plusieurs responsables des essais, ils sont ajoutés aux composantes Entrée et Sortie dans une application intégrée. Cette nouvelle application intégrée passe alors à l'étape suivante des essais.

L'essai intégré porte sur l'ensemble des modules testés à l'étape précédente, regroupés en applications intégrées avec les composantes Entrée et Sortie. Cette deuxième étape des essais fait en sorte que les données clés, comme l'âge et le sexe, passent de la composante Entrée aux composantes C2 et Sortie de l'application. Elle confirme également que les variables qui influent sur les instructions « passez à » et la logique d'enchaînement sont transmises correctement de module en module à l'intérieur de la composante C2. Étant donné qu'à cette étape, le fonctionnement de l'application est essentiellement identique à ce qu'il sera sur le terrain, tous les scénarios possibles auxquels feront face les intervieweurs sont simulés, afin de s'assurer qu'ils fonctionnent bien. Les scénarios servent à tester divers aspects des composantes Entrée et Sortie, y compris la prise de contact, la collecte d'information sur le contact, la question de savoir si un cas est compris dans le champ de l'enquête, le listage de ménages, la prise de rendez-vous et la sélection des répondants.

L'essai de bout en bout situe l'application entièrement intégrée dans un environnement de collecte simulé. Les applications sont chargées dans des ordinateurs connectés à un serveur d'essai. Les données sont par la suite recueillies, transmises et extraites en temps réel, exactement comme elles le seraient sur le terrain. Cette dernière étape des essais permet la vérification de tous les aspects techniques de l'entrée, ainsi que de la transmission et de l'extraction des données pour l'application de l'ESCC – Vieillessement en santé. Elle fournit aussi une dernière occasion de trouver des erreurs dans les composantes Entrée, C2 et Sortie.

5.3 Formation des intervieweurs

En octobre et novembre 2008, des représentants de la Division de la planification et de la gestion de la collecte de Statistique Canada ont visité quatre bureaux régionaux (Halifax, Montréal, Toronto et Edmonton) pour coordonner la collecte des données pour l'ESCC – Vieillessement en santé. Ces visites visaient à former les gestionnaires de projet des bureaux régionaux et les intervieweurs principaux pour l'ESCC – Vieillessement en santé. Les membres de l'équipe de projet de l'ESCC – Vieillessement en santé de la Division de la statistique de la santé ont aussi participé aux séances de formation, en vue de présenter de l'information concernant le contexte et l'élaboration de l'enquête, ainsi que d'offrir du soutien additionnel et de préciser les questions ou les préoccupations qui ont été soulevées.

Ces séances visaient à mettre les intervieweurs à l'aise pour l'utilisation de l'application de l'ESCC – Vieillessement en santé, ainsi qu'à les familiariser avec le contenu de l'enquête. La formation était centrée sur :

- les buts et objectifs de l'enquête;
- les techniques d'enquête;
- les fonctionnalités des applications;
- l'examen du contenu du questionnaire, y compris des exercices;
- l'administration des exercices du module de la cognition pour l'enregistrement des réponses audio;
- la simulation d'interviews difficiles et de situations de non-réponse;
- la gestion de l'enquête; et
- les procédures de transmission.

5.4 Interview

Les unités sélectionnées à partir de la base aréolaire ont été interviewées par des intervieweurs sur place au moyen de l'IPAO. Les intervieweurs sur place ont effectué leur travail en autonomie à la maison, au moyen d'ordinateurs portables, et ils étaient surveillés à distance par des intervieweurs principaux. Les intervieweurs qui se servaient de la méthode de l'IPAO ont reçu la formation nécessaire pour procéder à une première prise de contact sur place avec chaque ménage échantillonné. Tous les efforts ont été déployés pour mener les interviews sur place, et 94 % des interviews ont été menées exclusivement de cette façon. On a autorisé la collecte par téléphone uniquement lorsqu'un répondant demandait à être interviewé dans l'autre langue officielle, mais qu'aucun intervieweur bilingue n'était disponible dans la région, ou lorsque le répondant ne parlait ni l'une ni l'autre des langues officielles, mais qu'un autre intervieweur était disponible pour traduire pour le répondant.

Dans tous les logements choisis, l'intervieweur demandait à un membre du ménage bien informé de fournir l'information démographique de base sur tous les occupants. Un

membre du ménage âgé de 45 ans et plus a alors été sélectionné pour une interview plus approfondie, appelée interview C2.

Dans un souci d'assurer la qualité des données recueillies, les intervieweurs avaient reçu instruction de prendre tous les moyens à leur disposition pour mener en privé l'interview avec le répondant choisi. Là où la situation était inévitable, le répondant a été interviewé en présence d'une autre personne. Des indicateurs dans les fichiers de microdonnées signalent si une personne autre que le répondant était présente pendant l'interview (ADM_N10) et si, de l'avis de l'intervieweur, les réponses du répondant étaient influencées par la présence d'une autre personne (ADM_N11).

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

a) Lettres de présentation

Avant le début de chaque période de collecte, des lettres et des brochures de présentation expliquant l'objectif de l'enquête ont été envoyées aux ménages sélectionnés. Elles visaient à expliquer l'importance de l'enquête et à fournir des exemples de la façon dont les données de l'ESCC – Vieillesse en santé seraient utilisées.

b) Prise de contact

Les intervieweurs ont reçu instruction de mettre en œuvre tous les moyens raisonnables pour obtenir des interviews. Lorsque le moment de la visite de l'intervieweur ne convenait pas, on a pris rendez-vous pour mener l'interview à un moment plus propice. Si personne n'était à la maison au moment de la première visite, un avis de visite et de l'intention de prendre contact a été laissé à la porte. De nombreuses visites répétées au logement ont été faites à différents moments, différents jours.

c) Conversion des cas de refus

Si une personne refusait d'abord de participer à l'enquête, le bureau régional de Statistique Canada lui faisait parvenir une lettre qui soulignait l'importance de l'enquête et de la collaboration du ménage. Cela a été suivi par une deuxième visite (ou appel) d'un intervieweur principal, d'un superviseur de projet ou d'un autre intervieweur, afin de tenter de convaincre le répondant de l'importance de participer à l'enquête.

d) Obstacles linguistiques

Afin de supprimer les obstacles linguistiques susceptibles de nuire aux interviews, chacun des bureaux régionaux de Statistique Canada recrute des intervieweurs qui ont une vaste gamme de compétences linguistiques. Au besoin, les cas ont été transférés à un intervieweur capable de remplir le questionnaire dans la langue voulue.

e) Interviews par procuration

Dans les cas où le répondant sélectionné était, pour des raisons de santé physique ou mentale, incapable de répondre à l'interview, les renseignements à son sujet ont été fournis par un autre membre bien informé du ménage. Cette façon de faire est qualifiée d'interview par procuration. Même si les personnes interviewées par procuration sont souvent capables de fournir des réponses exactes à la plupart des questions de l'enquête, les questions plus délicates ou personnelles dépassaient la portée de leurs connaissances. Par conséquent, certaines questions posées dans le cadre des interviews par procuration sont demeurées sans réponse. Il fallait donc tout tenter pour réduire au minimum le nombre d'interviews par procuration. La variable ADM_PRX indique si l'interview d'un cas a été menée par procuration.

5.5 Opérations sur le terrain

L'échantillon de l'ESCC – Vieillesse en santé a été divisé en six périodes de collecte de deux mois, sans chevauchement. Les bureaux de collecte régionaux ont été informés d'utiliser les quatre premières semaines de chaque période de collecte pour interviewer la majorité de l'échantillon, les quatre semaines suivantes étant consacrées aux interviews restantes et au suivi des cas de non-réponse. À la deuxième semaine de chaque période, des tentatives devaient avoir été effectuées relativement à tous les cas. Les cas initialement codés « Pas de contact » ou « Absent pour la durée de l'enquête » ont été renvoyés à deux autres périodes de collecte : 1) du 15 août au 30 novembre 2009 et 2) du 15 octobre au 30 novembre 2009.

Les fichiers d'échantillon ont été envoyés environ six semaines avant le début de chaque période de collecte aux quatre bureaux régionaux chargés de la collecte pour la planification de la charge de travail. Chaque échantillon d'IPAO comprenait une série de cas fictifs. Ces cas ont été confiés à des intervieweurs principaux, afin de s'assurer que toutes les procédures de transmission des données fonctionnaient pendant l'ensemble du cycle de collecte. Après réception des échantillons, il incombait aux surveillants de projet de planifier les tâches des intervieweurs chargés des interviews selon la méthode de l'IPAO.

Les interviews par IPAO terminées ont été transmises quotidiennement du domicile de l'intervieweur, directement au bureau central de Statistique Canada, au moyen d'un système de transmission téléphonique sécuritaire.

Au terme de la collecte de données, un taux de réponse de 74 % a été atteint au niveau national. Des détails complets concernant les taux de réponse se trouvent au chapitre 8, section 8.1.

5.6 Contrôle de la qualité et gestion de la collecte

Au cours de la collecte des données de l'ESCC – Vieillesse en santé, plusieurs méthodes ont été utilisées pour assurer la qualité des données et optimiser la collecte. Il

s'agissait notamment de mesures internes en vue de vérifier le rendement des intervieweurs, de l'observation des interviews sur le terrain, ainsi que de bulletins d'information à l'intention des intervieweurs, ainsi que d'une série de rapports permanents pour contrôler les divers objectifs de collecte et la qualité des données.

Les représentants du bureau central ont observé les interviews en accompagnant les intervieweurs pendant les interviews à domicile. L'objectif principal de ces observations était de faire en sorte que des techniques et des procédures d'interview appropriées (lecture de la question telle que libellée dans l'application, absence de questions d'approfondissement, etc.) étaient suivies.

Trois bulletins à l'intention des intervieweurs ont été produits et distribués à tous les intervieweurs. Cet outil de communication a permis de fournir de la rétroaction directe aux intervieweurs concernant les questions de qualité des données. Ils ont aussi permis de faire le point sur la collecte, de répondre aux questions fréquentes concernant la collecte, de donner des renseignements additionnels et de renforcer les procédures et les techniques d'interview appropriées.

Les responsables ont produit une série de rapports visant à contrôler et à gérer efficacement les objectifs de collecte, ainsi qu'à mettre au jour les problèmes posés par la collecte.

Des rapports cumulatifs ont été produits à la fin de chaque période de collecte, afin de montrer les taux de réponse, de lien, de partage et de procuration pour chaque région de collecte. Des rapports ont aussi été produits par les méthodologistes d'enquête pour contrôler les grappes d'échantillon et les groupes d'âge qui étaient inférieurs aux objectifs de collecte, afin de permettre aux bureaux régionaux de mettre l'accent sur les régions concernées.

Par ailleurs, des rapports personnalisés ont été établis et utilisés pour examiner des questions particulières de qualité des données qui se sont posées au cours de la collecte. Par exemple, le protocole de l'ESCC – Vieillesse en santé permet des interviews par procuration lorsque les répondants sélectionnés sont incapables mentalement ou physiquement de répondre à l'enquête. Les interviews par procuration ont une valeur limitée parce que certains modules sont sautés pendant ces interviews pour des raisons de qualité des données. Il est donc important de déterminer les taux d'interview par procuration et les motifs derrière ces interviews. Un rapport d'interviews par procuration a été créé pour assurer le suivi de ces interviews et les justifier. Grâce à ces rapports, il a été possible d'identifier les intervieweurs qui ne semblaient pas savoir clairement quand une interview par procuration était acceptable. Des rapports similaires ont été produits pour contrôler les résultats en ce qui a trait au taux de refus, de lien et de partage, pour un suivi ultérieur et la gestion de la collecte permanente. Ces intervieweurs ont par la suite reçu de la formation additionnelle/un rappel, afin de passer en revue les procédures pertinentes.

6.0 Traitement des données

6.1 Contrôle

La plupart du contrôle des données a été effectué au moment de l'interview, grâce à l'application d'interview sur place assistée par ordinateur (IPAO). Les intervieweurs ne pouvaient pas entrer de valeurs non valides et les erreurs d'enchaînement ont fait l'objet d'un contrôle grâce à des instructions programmées. Par exemple, l'application d'IPAO a fait en sorte que les questions qui ne s'appliquaient pas au répondant n'ont pas été posées.

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 ».

6.1.1 Vérification du module des activités physiques

Dans le module 2 des activités physiques, des questions ont été posées concernant le niveau d'activité physique du répondant. Les activités ont été consignées à quatre niveaux (légères, modérées, ardues, et force musculaire et endurance) et, pour chaque niveau, on a inscrit la fréquence hebdomadaire et la durée quotidienne de toutes les activités. Les variables de fréquence hebdomadaire et de durée étaient des variables catégoriques. Des problèmes se sont produits lorsque des activités qui auraient dû être enregistrées à un niveau l'ont été à un autre (p. ex., déclarer la bicyclette comme une activité légère, alors qu'il s'agissait d'une activité ardue). Pour corriger ce problème de classification, ces activités ont été transférées au niveau approprié.

Pour chaque activité transférée, la fréquence hebdomadaire et la durée quotidienne des deux niveaux touchés ont été corrigées. Une partie proportionnelle de la fréquence hebdomadaire a été supprimée du niveau d'activité précédent et transférée au nouveau niveau d'activité. En ce qui a trait à la durée quotidienne, la même durée a été reportée du niveau précédent au nouveau niveau. Lorsque les activités existaient déjà dans le nouveau niveau, des valeurs ont été ajoutées aux fréquences hebdomadaires et aux durées quotidiennes existantes. De cette façon, la quantité totale d'activités (correspondant à la fréquence multipliée par la durée) est maintenue, mais le niveau d'effort relatif à une activité particulière est modifié afin d'assurer l'uniformité entre tous les enregistrements.

6.2 Codage

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

Dans les cas où la réponse donnée par le répondant ne pouvait être attribuée facilement à une catégorie existante, l'intervieweur pouvait poser plusieurs questions lui permettant d'entrer une réponse en toutes lettres dans la catégorie « Autre – veuillez préciser ». Par suite de la collecte, toutes ces questions ont été passées en revue au moment du traitement au bureau central. Dans certains cas, on a donné aux réponses en toutes lettres le code d'une catégorie figurant sur la liste, si la réponse faisait double emploi. On tiendra compte des réponses « Autre – veuillez préciser » fournies pour toutes les questions lors du perfectionnement des catégories de réponses, en vue des cycles futurs de l'enquête.

6.2.1 Codage du module sur la cognition

Le module sur la cognition de l'enquête exigeait des répondants qu'ils effectuent quatre tâches, dont les réponses ont été enregistrées (avec leur permission). Ces tâches étaient les suivantes : nommer le plus grand nombre d'animaux possible, compter de 1 à 20, réciter les lettres de l'alphabet et alterner des chiffres et des lettres. Les répondants ont eu au maximum 1 minute pour s'acquitter de la tâche consistant à nommer des animaux et 30 secondes pour chacune des autres tâches.

Le traitement et le codage des réponses enregistrées se sont faits en plusieurs étapes. Tout d'abord, les enregistrements sonores (fichiers audio) ont été renvoyés au bureau central de Statistique Canada en blocs de fichiers audio distincts des données d'enquête, chacun comportant un nom unique, afin de pouvoir être identifié ultérieurement et associé aux données d'enquête du répondant. Les blocs de fichiers audio ont été déballés et triés en bases de données distinctes pour chacune des quatre tâches. Les fichiers audio de chaque tâche de cognition ont par la suite été versés dans la base de données appropriée. Une interface élaborée au niveau interne a permis aux membres de l'équipe de codage des données d'accéder aux fichiers audio de ces bases de données, d'écouter les enregistrements et de les coder. On leur a fourni des règles de codage. Une fois les quatre tâches codées pour un répondant, les données codées étaient intégrées dans les données d'enquête du répondant.

6.3 Création de variables dérivées

Afin de faciliter l'analyse des données et de réduire le risque d'erreur, un certain nombre de variables du fichier ont été calculées au moyen de questions figurant dans les questionnaires de l'ESCC – Vieillesse en santé. Les variables dérivées comportent généralement un « D », « G » ou « F » au quatrième caractère du nom de la variable. Dans certains cas, les variables dérivées sont simples, correspondant au regroupement de catégories de réponses. Dans d'autres cas, on a combiné plusieurs variables pour en créer une nouvelle. La *Documentation sur les variables dérivées (VD)* fournit des détails sur la façon de dériver les variables plus complexes. Pour plus de renseignements sur la convention de désignation, voir la section 11.5.

6.4 Interviews partielles

Dans certains cas, l'intervieweur a pu procéder à une partie de l'interview, mais n'a pu l'achever pour diverses raisons. Certains répondants n'étaient disposés à consacrer qu'un temps limité à la réponse au questionnaire. Dans d'autres cas, l'intervieweur a procédé à une partie de l'interview avec le répondant et pris rendez-vous pour la poursuivre à un autre moment, mais n'a pas pu reprendre contact avec le répondant. Dans ces cas, il a fallu déterminer des critères types pour décider du sort de ces interviews « partielles ».

Pour l'ESCC – Vieillesse en santé, il a été décidé que pour qu'une interview partielle soit acceptée, le répondant devait avoir répondu aux modules d'entrée et de sortie, ainsi qu'à une petite partie du questionnaire sur la santé générale, jusqu'à la première question du module sur la participation sociale (SPA) inclusivement. Toute réponse n'atteignant pas ce niveau a été considérée comme une non-réponse, ce qui signifie que le ménage a été éliminé de l'échantillon de répondants. La variable ADMD_STA indique si un cas a été partiellement ou complètement terminé, au moyen des chiffres 71 et 70 respectivement.

6.5 Pondération

Le principe qui sous-tend l'estimation dans un échantillon probabiliste comme celui de l'ESCC – Vieillesse en santé est que chaque personne de l'échantillon « représente », en plus d'elle-même, plusieurs autres personnes ne figurant pas dans l'échantillon. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne en représente 50. Dans la terminologie en usage ici, nous dirons que nous avons attribué à chaque personne un facteur de pondération de 50.

L'étape de détermination des facteurs de pondération donne lieu au calcul du poids d'échantillonnage de chaque personne échantillonnée. Ce poids figure dans le fichier et doit être utilisé pour calculer des estimations utiles à partir 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 des détails sur la façon dont on calcule les poids d'échantillonnage à la section 7.

7.0 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, les utilisateurs doivent incorporer les poids de sondage dans leurs calculs. Un poids de sondage est attribué à chaque personne incluse dans l'échantillon final, c'est-à-dire dans l'échantillon de personnes ayant répondu à l'enquête. Ce poids correspond au nombre de personnes représentées par le répondant dans l'ensemble de la population de l'enquête.

7.1 Pondération de l'échantillon

Il y a sept ajustements distincts visés par la stratégie de pondération. Les sections suivantes décrivent le processus de pondération de l'enquête. La sous-section 7.1.1 explique le poids initial, puis la sous-section 7.1.2 fournit les détails du retrait des unités hors champ. Le traitement des ménages non-répondants est présenté dans la sous-section 7.1.3. On retrouve ensuite les ajustements du poids-personne dans la sous-section 7.1.4 et ceux de la non-réponse à l'échelle de la personne à la sous-section 7.1.5. Enfin, on couvre les étapes finales de la pondération, la méthode d'estimation de Winsor et le calage aux marges, dans les sous-sections 7.1.6 et 7.1.7.

7.1.1 Poids initial

La pondération pour l'échantillon commence avec un poids initial du ménage (L1). Le poids initial couvre les deux premières étapes du plan d'échantillon, telles que décrites à la sous-section 4.3 : la sélection des grappes et la sélection des ménages dans chaque grappe.

Le poids de la grappe est l'inverse de la probabilité de sélection de la grappe. Telles que mentionnées à la sous-section 4.3.1, les grappes dans chaque province ont été stratifiées entre des grappes urbaines et rurales. Ensuite, le nombre de grappes désiré est échantillonné proportionnellement au nombre de personnes de 45 ans et plus dans chaque grappe. Selon l'algorithme Hanurav-Vijayan⁴, la probabilité de sélection pour chaque grappe est :

$$\mathbf{PoidsGrappe}_i = n \frac{M_i}{M_{total}} ,$$

où i = la grappe sélectionnée, n = le nombre de grappes échantillonné par province et statut urbain/rural, M_i = le nombre de personnes de 45 ans et plus dans la grappe sélectionnée et M_{total} = le nombre total des personnes de 45 ans et plus dans tous les grappes de la même province et du même statut urbain/rural.

⁴ Vijayan, K. (1968). An Exact πps Sampling Scheme – Generalization of a Method of Hanurav. *Journal of the Royal Statistical Society*, Series B, 30, p. 556-566.

Le poids du logement est l'inverse de la probabilité de sélection du logement dans chaque grappe. La sous-section 4.3.2 décrit cette procédure où les grappes sont divisées en trois strates et un nombre de logements précis est sélectionné dans chacune. La probabilité de sélection pour chaque logement est calculée par échantillonnage aléatoire simple de la façon suivante :

$$\mathbf{PoidsLogement}_j = \frac{N_j}{n_j} ,$$

où j = le logement sélectionné, N_j = le nombre total des logements dans la grappe par strate et n_j = le nombre de logements échantillonné dans la grappe par strate.

Le poids initial est simplement le produit de ces deux composants :

$$\mathbf{L1} = \mathbf{PoidsGrappe}_i \times \mathbf{PoidsLogement}_j .$$

7.1.2 Retrait des unités hors champ

Parmi tous les logements échantillonnés, une certaine proportion de ceux-ci est identifiée, lors de la collecte, comme étant hors du champ de l'enquête. Des logements qui sont à l'extérieur de la population cible comme spécifiée à la sous-section 4.1 (y compris ceux sans personnes de 45 ans et plus) ainsi que des logements détruits ou en construction, les logements vacants, saisonniers ou secondaires sont des exemples de cas hors champ de l'ESCC – Vieillesse en santé. Ces logements et leur poids connexe sont tout simplement retirés de l'échantillon, ne laissant plus que les logements faisant partie du champ de l'enquête. Les logements qui restent dans l'échantillon conservent le même poids qu'à l'étape précédente que l'on appelle maintenant poids L2 :

$$\mathbf{L2} = \begin{cases} \mathbf{L1} & \text{si l'unité fait partie du champ} \\ 0 & \text{si l'unité est hors champ} \end{cases} .$$

7.1.3 Non-réponse à l'échelle du ménage

Lors de la collecte, une certaine proportion des ménages échantillonnés a inévitablement résulté en non-réponse. Ceci survient habituellement lorsqu'un ménage refuse de participer à l'enquête, fournit des données inutilisables, ou encore, ne peut être rejoint pour réaliser l'interview. Les poids des ménages non-répondants sont redistribués aux répondants à l'intérieur de groupes homogènes de réponse (GHR). Pour créer ces GHR, on utilise la méthode des scores fondée sur les modèles de régression logistique afin de déterminer la propension de réponse. Ces probabilités de réponse sont par la suite utilisées afin de diviser l'échantillon par groupes ayant des propriétés de réponse similaires. L'information disponible pour les non-répondants étant limitée, on utilise des caractéristiques du modèle de régression comme la période de collecte et les

renseignements géographiques ainsi que des parodontées, ce qui inclut le nombre d'essais pour contacter le ménage, l'heure/le jour de l'essai et si les essais ont été faits au cours de la semaine ou de la fin de semaine. Un facteur d'ajustement a donc été calculé à l'intérieur de chaque GHR de la façon suivante :

$$\frac{\text{Somme des poids L2 pour tous les ménages}}{\text{Somme des poids L2 pour tous les ménages répondants}}$$

Le poids L2 des ménages répondants est multiplié par ce facteur d'ajustement pour produire le poids L3. Les ménages non-répondants sont éliminés du processus de pondération à partir de ce point.

7.1.4 Création du poids-personne

Puisque l'unité d'échantillonnage désiré pour l'ESCC – Vieillesse en santé est la personne, le poids-ménage calculé jusqu'ici doit être converti en un poids-personne. Celui-ci est obtenu en multipliant le poids L3 par l'inverse de la probabilité de sélection de la personne choisie dans le ménage. Nous obtenons ainsi le poids I1. Rappelons que la probabilité de sélection d'une personne change en fonction du nombre de personnes de 45 ans et plus qui se trouvent dans le ménage et de l'âge de ces personnes (voir la sous-section 4.3.3 pour plus de détails).

7.1.5 Non-réponse à l'échelle de la personne

Dans le cadre de l'ESCC – Vieillesse en santé, une interview peut être vue comme un processus en deux étapes. Dans un premier temps, l'interviewer obtient la liste complète des personnes vivant dans le ménage, puis il (elle) interviewe la personne sélectionnée dans le ménage. Dans certains cas, les interviewers 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 répondants pour compenser cette non-réponse. Tout comme pour la non-réponse à l'échelle du ménage, l'ajustement est appliqué à l'intérieur des GHR définis à partir des caractéristiques disponibles pour les répondants et non-répondants. Toutes les caractéristiques recueillies lors du listage des membres du ménage, en plus de l'information géographique et certaines parodontées, étaient en fait disponibles pour créer ces groupes. La méthode des scores est utilisée afin de définir les groupes. Le facteur d'ajustement est calculé de la façon suivante :

$$\frac{\text{Somme des poids I1 pour toutes les personnes sélectionnées}}{\text{Somme des poids I1 pour toutes les personnes sélectionnées répondantes}}$$

Le poids I1 des personnes répondantes a donc été multiplié par ce facteur d'ajustement pour produire le poids I2. Les personnes non-répondantes sont éliminées du processus de pondération à partir de ce point.

7.1.6 Méthode d'estimation de Winsor

Notez qu'à la suite à la série d'ajustements appliqués aux répondants, il est possible que certaines unités se retrouvent avec des poids se démarquant des autres poids de leur domaine d'intérêt (selon la province, le groupe d'âge et le sexe). Certains répondants peuvent effectivement représenter une proportion anormalement élevée de leur domaine d'intérêt et influencer fortement sur la variance, et la variance, de ces domaines d'intérêt. Afin d'éviter cette situation, le poids des répondants qui contribuent de façon aberrante à leur groupe province-âge-sexe est ajusté à la baisse selon la « méthode d'estimation de Winsor ». Pour la plupart des unités qui ne sont pas des valeurs aberrantes, le poids I3 produit à ce point sera le même que le poids I2.

7.1.7 Calage aux marges

La dernière étape nécessaire afin d'obtenir le poids final de l'ESCC – Vieillesse en santé est le calage aux marges (I4). Le calage est appliqué en utilisant Calmar⁵ afin de s'assurer que la somme des poids finaux corresponde aux estimations de populations définies à l'échelle des provinces, pour chacun des 10 groupes d'intérêt âge-sexe, c'est-à-dire les cinq groupes d'âge 45-54, 55-64, 65-74, 75-84 et 85+, pour chacun des deux sexes. Un minimum de 20 répondants est nécessaire dans chaque domaine pour un calage à l'échelon provincial par groupe âge-sexe. Pour les domaines ayant moins de 20 répondants, des regroupements sont effectués au niveau de la province ou du sexe. Il est à noter que le calage aux marges a été fait en utilisant une géographie la plus à jour possible qui diffère peut-être de la géographie utilisée lors de l'échantillonnage.

Les estimations de la population sont fondées sur les chiffres du Recensement les plus récents, de même que sur les chiffres ayant trait aux naissances, aux décès, à l'immigration et l'émigration depuis ce temps. La moyenne des estimations mensuelles pour chacune des strates a posteriori province-âge-sexe par période de collecte a été retenue pour réaliser le calage. Le poids I3 a donc été ajusté à l'aide de Calmar pour obtenir le poids final I4. Le poids I4 correspond au **poids final de l'ESCC – Vieillesse en santé** que l'on retrouve dans le fichier de données portant le nom de variable WTS_M.

⁵ Sautory, O. (2003). Calmar 2: Une nouvelle version du programme Calmar de redressement d'échantillon par calage. Recueil du Symposium 2003 de Statistique Canada : Défis reliés à la réalisation d'enquêtes pour la prochaine décennie, produit n° 11-522-XIF au catalogue de Statistique Canada.

8.0 Qualité des données

8.1 Taux de réponse

Au total, 41 496 des unités sélectionnées à l'ESCC – Vieillesse en santé faisaient partie du champ de l'enquête⁶. Parmi ces unités, 33 517 ont accepté de participer à l'enquête, ce qui entraîne un taux de réponse global à l'échelle du ménage de 80,8 %. Parmi ces ménages répondant, 33 517 personnes (une par ménage) ont été sélectionnées pour participer à l'enquête, parmi lesquelles 30 865 ont répondu, ce qui entraîne un taux de réponse global à l'échelle de la personne de 92,1 %. À l'échelle canadienne, un taux de réponse combiné (du ménage et de la personne) de 74,4 % a donc été observé pour l'ESCC – Vieillesse en santé. Le tableau 8.1 donne les taux de réponse par province et par groupe d'âge⁷.

Tableau 8.1: Taux de réponse par province et par groupe d'âge dans l'ESCC – Vieillesse en santé

Province	Taux de réponse provincial du ménage	Âge	À l'échelle de la personne			Taux de réponse combiné
			Nombre de personnes sélectionnées	Nombre de répondants	Taux de réponse	
10 - Terre-Neuve-et-Labrador	86,8	Total	2 185	2 010	92,0	79,9
		[45-54]	403	358	88,8	
		[55-64]	730	667	91,4	
		[65-75]	466	443	95,1	
		[75-84]	318	299	94,0	
		[85+]	268	243	90,7	
11 - Île-du-Prince-Édouard	83,3	Total	1 765	1 650	93,5	77,8
		[45-54]	322	293	91,0	
		[55-64]	526	485	92,2	
		[65-75]	350	334	95,4	
		[75-84]	313	300	95,8	

⁶ Parmi les unités sélectionnées, certaines ne font pas partie du champ de l'enquête. Ce sont, par exemple, des logements vacants ou démolis, ou des ménages dont aucun des résidents n'appartient au groupe d'âge ciblé par l'enquête. Ces unités sont identifiées pendant la collecte, autrement elles auraient été exclues avant la sélection de l'échantillon. Ces unités ne sont pas considérées dans le calcul des taux de réponse.

⁷ La stratification pour l'ESCC – Vieillesse en santé prenait en considération l'âge prévu des membres du ménage au moment de la collecte. Les taux de réponse pour les groupes d'âge [45-54], [55-64], [65-74], [75-84] et [85+] étaient suivis pour s'assurer que chaque groupe était bien représenté dans l'échantillon. Les taux de réponse sont présentés selon les mêmes regroupements.

			À l'échelle de la personne			
Province	Taux de réponse provincial du ménage	Âge	Nombre de personnes sélectionnées	Nombre de répondants	Taux de réponse	Taux de réponse combiné
		[85+]	254	238	93,7	
12 - Nouvelle-Écosse	82,4	Total	2 536	2 282	90,0	74,1
		[45-54]	399	353	88,5	
		[55-64]	794	706	88,9	
		[65-74]	595	551	92,6	
		[75-84]	393	358	91,1	
		[85+]	355	314	88,5	
13 - Nouveau-Brunswick	85,2	Total	2 396	2 225	92,9	79,1
		[45-54]	381	351	92,1	
		[55-64]	730	676	92,6	
		[65-74]	526	495	94,1	
		[75-84]	400	370	92,5	
		[85+]	359	333	92,8	
24 - Québec	80,0	Total	5 649	5 217	92,4	73,8
		[45-54]	964	857	88,9	
		[55-64]	1 749	1 623	92,8	
		[65-74]	1 278	1 187	92,9	
		[75-84]	1 046	988	94,5	
		[85+]	612	562	91,8	
35 - Ontario	79,1	Total	7 159	6 525	91,1	72,1
		[45-54]	1 152	1 023	88,8	
		[55-64]	2 187	1 984	90,7	
		[65-74]	1 624	1 496	92,1	
		[75-84]	1 245	1 150	92,4	
		[85+]	951	872	91,7	
46 - Manitoba	79,9	Total	2 386	2 177	91,2	72,9
		[45-54]	429	378	88,1	
		[55-64]	678	622	91,7	
		[65-74]	528	486	92,0	
		[75-84]	399	369	92,5	

			À l'échelle de la personne			
Province	Taux de réponse provincial du ménage	Âge	Nombre de personnes sélectionnées	Nombre de répondants	Taux de réponse	Taux de réponse combiné
		[85+]	352	322	91,5	
47 - Saskatchewan	80,7	Total	2 304	2 184	94,8	76,5
		[45-54]	371	336	90,6	
		[55-64]	672	639	95,1	
		[65-74]	499	476	95,4	
		[75-84]	429	417	97,2	
		[85+]	333	316	94,9	
48 - Alberta	78,8	Total	3 012	2 735	90,8	71,6
		[45-54]	633	558	88,2	
		[55-64]	853	775	90,9	
		[65-74]	648	597	92,1	
		[75-84]	473	436	92,2	
		[85+]	405	369	91,1	
59 - Colombie-Britannique	79,4	Total	4 125	3 860	93,6	74,3
		[45-54]	713	659	92,4	
		[55-64]	1 242	1 153	92,8	
		[65-74]	964	910	94,4	
		[75-84]	675	633	93,8	
		[85+]	531	505	95,1	
Canada	80,8	Total	33 517	30 865	92,1	74,4
		[45-54]	5 767	5 166	89,6	
		[55-64]	10 161	9 330	91,8	
		[65-74]	7 478	6 975	93,3	
		[75-84]	5 691	5 320	93,5	
		[85+]	4 420	4 074	92,2	

8.1.1 Formules pour le calcul des taux de réponse et exemple

On décrit dans ce qui suit les formules pour le calcul des taux de réponse et on donne un exemple illustrant de quelle façon les taux de réponse à l'échelle de la personne et du ménage doivent être manipulés pour calculer les taux de réponse combinés.

Taux de réponse à l'échelle du ménage

$$HHRR = \frac{\text{Nombre de ménages répondants}}{\text{Nombre de ménages faisant partie du champ de l'enquête}}$$

Taux de réponse à l'échelle de la personne

$$PPRR = \frac{\text{Nombre de personnes répondantes}}{\text{Nombre de personnes sélectionnées (une par ménage répondant)}}$$

Taux de réponse combiné

$$CombRR = HHRR \times PPRR$$

Exemple 1 – Calcul de CombRR pour Terre-Neuve-et-Labrador d'après le tableau 8.1

$$HHRR = \frac{\text{Nombre de ménages répondants}}{\text{Nombre de ménages faisant partie du champ de l'enquête}} = \frac{2185}{2516} = 86,8 \%$$

$$PPRR = \frac{\text{Nombre de personnes répondantes}}{\text{Nombre de personnes sélectionnées}} = \frac{2010}{2185} = 92,0 \%$$

$$CombRR = 86,8 \% \times 92,0 \% = \mathbf{79,9 \%}$$

8.1.2 Restrictions dans le calcul du taux de réponse

Deux variables d'intérêt dans l'ESCC – Vieillesse en santé, le groupe d'âge et le sexe, sont liées aux personnes et ainsi, seul le taux de réponse à l'échelle de la personne peut être présenté selon cette ventilation.

8.2 Erreurs dans l'enquête

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

d'un échantillon et celles qui découlent d'un recensement complet de la population cible dans des conditions semblables s'appelle l'erreur d'échantillonnage des estimations.

Des erreurs qui ne sont pas liées à l'échantillonnage peuvent être commises à presque chaque étape des opérations de l'enquête. Il est possible que les intervieweurs comprennent mal les instructions, que les répondants fassent des erreurs en remplissant le questionnaire, que les réponses soient mal saisies dans l'ordinateur 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.

8.2.1 Erreurs d'échantillonnage

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

Par exemple, supposons qu'on estime que 25 % des Canadiens de 45 ans et plus font de l'exercice au moins une fois par semaine et que cette estimation comporte un écart type de 0,003. On calcule alors le CV de cette estimation de la façon suivante :

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

Statistique Canada utilise fréquemment les résultats du CV pour l'analyse des données et conseille vivement aux utilisateurs produisant des estimations à partir des fichiers de données de l'ESCC – Vieillesse en santé de faire de même. Pour plus d'information sur le calcul des CV, voir la section 10. Pour consulter les lignes directrices sur la façon d'interpréter les résultats du CV, vous référer au tableau à la fin de la sous-section 9.4.

8.2.2 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 mises en place 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 ayant reçu une formation poussée sur les procédures à suivre et sur le questionnaire, et à l'observation des

entrevues afin de déceler les problèmes. La mise à l'essai de l'application IPAO et les essais sur le terrain ont également été au nombre des procédures essentielles pour réduire au maximum les erreurs dans la collecte de données.

L'effet de la non-réponse sur les résultats de l'enquête constitue une source importante d'erreurs non dues à l'échantillonnage dans les enquêtes. L'ampleur de la non-réponse varie d'une non-réponse partielle (le fait de ne pas répondre à une ou à quelques questions) à une non-réponse totale. Dans l'ESCC – Vieillesse en santé, il y a peu de non-réponse partielle, car une fois que les répondants commencent à répondre au questionnaire, ils avaient tendance à le terminer. Il y avait une non-réponse totale lorsque la personne sélectionnée pour participer à l'enquête refusait de le faire ou que l'intervieweur était incapable d'entrer en contact avec elle. On a géré la non-réponse totale en corrigeant les poids des personnes ayant répondu à l'enquête afin de compenser pour celles n'ayant pas répondu. Pour obtenir plus de renseignements sur la correction de la pondération pour la non-réponse, voir la section 7.

La non réponse à une question particulière (non réponse partielle) est généralement faible, mais pour diverses raisons, elle peut être plus élevée dans certains modules que dans d'autres. La non réponse partielle augmente à mesure que l'on progresse dans le questionnaire à cause de la fatigue du répondant. Par exemple, la non réponse partielle est de 0,1 % au début du questionnaire (GEN_Q1) et de 0,7 % à la fin (SDC_Q1). La non réponse partielle est 2,2 % plus élevée pour les modules ne pouvant pas être remplis par personne interposée (p. ex., Satisfaction à l'égard de la vie [SLS]). Les questions du module sur la Disponibilité du soutien social (SSA) qui ont trait aux amis, aux sentiments et à des situations hypothétiques donnent lieu à un taux de non réponse plus élevé (p. ex., 5,4 % pour SSA_Q2), principalement en raison du plus grand nombre de répondants qui choisissent la réponse « Ne sait pas ». Les questions sur le revenu sont particulièrement délicates et ont la réputation de produire un taux de non réponse partielle élevé : les taux de non réponse sont de 16,8 % pour le revenu du ménage (IN2DHH) et de 21,8 % pour le revenu personnel (IN2DPER), c'est à dire des taux comparables à ceux observés pour la composante annuelle de l'Enquête sur la santé dans les collectivités canadiennes. Cette tendance touche aussi les variables dérivées d'après d'autres modules, mais dans lesquelles interviennent les variables de revenu (p. ex. Situation de retraite – Objectif (RETDRS) pour laquelle la non réponse est de 9,6 %). Durant l'analyse, il est vivement conseillé aux utilisateurs des données de tenir compte des variations plus importantes, par exemple, en introduisant dans l'analyse par régression une variable indicatrice pour la non réponse concernant le revenu.

9.0 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 des données provenant des fichiers de données de l'enquête. Ces lignes directrices devraient permettre aux utilisateurs de microdonnées de produire des chiffres semblables à ceux produits par Statistique Canada et de produire aussi des chiffres, qui ne sont pas encore publiés, conformes aux lignes directrices établies.

9.1 Lignes directrices pour l'arrondissement

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

- a) Les estimations qui figurent dans le corps d'un tableau statistique doivent être arrondies à la centaine près par la méthode d'arrondissement classique. Selon cette méthode, si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre retenu ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente d'une unité (1) la valeur du dernier chiffre retenu. Par exemple, si l'on veut arrondir à la centaine près de la façon classique une estimation dont les deux derniers chiffres sont compris entre 00 et 49, il faut les remplacer par 00 et ne pas modifier le chiffre précédent (le chiffre des centaines). Si les deux derniers chiffres sont compris entre 50 et 99, il faut les remplacer par 00 et augmenter d'une unité (1) le chiffre précédent.
- b) Les totaux et sous-totaux marginaux 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 par les utilisateurs en aucune circonstance. Des estimations non arrondies donnent l'impression d'être beaucoup 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 cette enquête n'est pas autopondéré. Autrement dit, le poids d'échantillonnage n'est pas le même pour toutes les personnes qui font partie de l'échantillon. Même pour produire des estimations simples, y compris des tableaux statistiques ordinaires, l'utilisateur doit employer le poids d'échantillonnage approprié. Sinon, les estimations calculées à partir du fichier de microdonnées ne pourront être considérées comme représentatives de la population observée et ne correspondront pas à celles de Statistique Canada.

L'utilisateur ne doit pas non plus perdre de vue que certains logiciels ne permettent peut-être pas d'obtenir des estimations qui coïncident exactement avec celles de Statistique Canada, en raison de leur façon de traiter les poids d'échantillonnage.

9.2.1 Définitions : estimations de type nominal; estimations quantitatives

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

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 font partie d'une catégorie particulière. Le nombre de personnes âgées de 65 ans ou plus ayant fait au moins une chute au cours des douze derniers mois 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 :

Au cours des 12 derniers mois, ^AVEZ-^VOUS_T fait des chutes? (FAL_01)

- Oui
 Non

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 chutes au cours des douze derniers mois par des personnes âgées de 65 ans ou plus ayant chuté au moins une fois. Le numérateur correspond à l'estimation du nombre total de chutes subies au cours des 12 derniers mois, alors que le dénominateur correspond à l'estimation du nombre de personnes ayant subi au moins une chute au cours des 12 derniers mois. Seules les personnes de 65 ans et plus répondent au module sur les chutes, et celui-ci se limite aux chutes où les gens se sont fait suffisamment mal pour limiter certaines de leurs activités normales.

Exemple de question quantitative :

Combien de fois ^ÊTES-^VOUS_I tombé^E au cours des 12 derniers mois? (FAL_02)

|_|_| Nombre de fois

9.2.2 Totalisation d'estimations de type nominal

On peut obtenir, à partir du fichier de microdonnées, des estimations du nombre de personnes qui possèdent une caractéristique particulière en faisant la somme des poids finaux de tous les enregistrements possédant cette caractéristique.

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

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

9.2.3 Totalisation d'estimations quantitatives

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

- a) multiplier la valeur de la variable étudiée par le poids final, puis faire la somme de cette quantité pour tous les enregistrements visés pour obtenir le numérateur (\hat{X});
- b) faire la somme des poids finaux des enregistrements ayant la caractéristique 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 chutes au cours des douze derniers mois par des personnes de 65 ans ou plus ayant chuté au moins une fois, on calcule d'abord le numérateur (\hat{X}) en faisant la somme du produit entre la valeur de la variable **FAL_02** et le poids **WTS_M**, pour les enregistrements pour lesquels la valeur de la variable **FAL_01** est « oui ». On obtient ensuite le dénominateur (\hat{Y}) en faisant la somme du poids final de tous les enregistrements pour lesquels la valeur de la variable **FAL_01** est « oui ». Le nombre moyen de chutes au cours des douze derniers mois par des personnes de 65 ans ou plus ayant chuté au moins une fois est finalement obtenu en divisant (\hat{X}) par (\hat{Y}).

9.3 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'enquêtes aussi complexes 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 ou la définition de ces poids peut différer de celle appropriée dans le contexte d'une enquête par sondage. Par conséquent, si les estimations calculées au moyen de ces progiciels sont souvent exactes, les variances calculées n'ont, quant à elles, pratiquement aucune signification.

Dans le cas de nombreuses méthodes d'analyse (par exemple la régression linéaire, la régression logistique, l'analyse de la variance), une méthode existe pour que les progiciels courants produisent des résultats plus adéquats. Cette méthode consiste à réechelonner les poids qui figurent dans les enregistrements de façon à ce que le poids moyen soit égal à un (1). Les résultats produits par les progiciels classiques sont alors 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 la variabilité d'échantillonnage approximative (couramment appelées « Tableaux de 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 obtenir plus de détails, consulter la section 10.

9.4 Lignes directrices pour la diffusion

Avant de diffuser et/ou de publier des estimations tirées des fichiers de microdonnées, l'utilisateur doit d'abord déterminer le nombre de répondants dans l'échantillon ayant la caractéristique à l'étude (par exemple, le nombre de répondants qui fument lorsqu'on s'intéresse à la proportion de fumeurs pour une population donnée) pour s'assurer qu'il y a assez d'unités pour calculer une estimation de qualité. Pour les utilisateurs de FMGD, si ce nombre est inférieur à 30, l'estimation pondérée ne doit pas être diffusée, quelle que soit la valeur de son coefficient de variation. Pour les utilisateurs des fichiers-maîtres ou partagés, il est recommandé d'avoir au moins 10 observations. Pour les estimations pondérées basées sur des échantillons de 10 ou plus (30 ou plus pour le FMGD), l'utilisateur doit calculer le coefficient de variation de l'estimation et suivre les lignes directrices qui suivent.

Tableau 9.1 Lignes directrices relatives à la variabilité d'échantillonnage

Type d'estimation	CV (en %)	Lignes directrices
Acceptable	$0,0 \leq CV \leq 16,5$	On peut envisager une diffusion générale non restreinte des estimations. Aucune annotation particulière n'est nécessaire.
Marginale	$16,5 < CV \leq 33,3$	On peut envisager une diffusion générale non restreinte des estimations, en y joignant une mise en garde aux utilisateurs quant à la variabilité d'échantillonnage élevée liée aux estimations. Les estimations de ce genre doivent être identifiées par la lettre E (ou d'une autre manière similaire).
Inacceptable	$CV > 33,3$	Statistique Canada recommande de ne pas publier des estimations dont la qualité est inacceptable. Toutefois, si l'utilisateur choisit de le faire, il doit alors y adjoindre la lettre F (ou un autre identificateur semblable) et les diffuser avec l'avertissement suivant : « Nous avisons l'utilisateur que ...(précisez les données)... ne répondent pas aux normes de qualité de Statistique Canada pour ce programme statistique. Les conclusions tirées de ces données ne sauraient être fiables et seront fort probablement erronées. Ces données et toute conclusion qu'on pourrait en tirer ne doivent pas être publiées. Si l'utilisateur choisit de les publier, il est alors tenu de publier également le présent avertissement. »

10.0 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 catégorielles obtenues à partir de ce fichier de microdonnées à grande diffusion (FMGD), on a produit un ensemble de tableaux sur la variabilité d'échantillonnage approximative. Ces tableaux de référence permettent aux utilisateurs d'obtenir un coefficient de variation approximatif selon la taille de l'estimation calculée à partir des données de l'enquête.

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

Les effets de plan, les tailles d'échantillon et les chiffres de la population qui ont servi à produire les tableaux sur la variabilité d'échantillonnage approximative ainsi que les tableaux eux-mêmes sont disponibles dans un document qui est inclus sur le CD du FMGD. Tous les coefficients de variation des tableaux sur la variabilité d'échantillonnage approximative sont approximatifs, donc non officiels. Les possibilités concernant le calcul de coefficients de variation exacts sont traitées à la sous-section 10.7.

Rappel : Tel qu'indiqué dans « Les lignes directrices relatives à la variabilité d'échantillonnage » à la section 9.4, si le nombre d'observations sur lesquelles une estimation s'appuie 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 fondés sur des échantillons de petite taille sont trop imprévisibles pour être adéquatement représentés dans les tableaux.

10.1 Comment utiliser les tableaux de CV pour les estimations catégorielles

Les règles suivantes devraient permettre à l'utilisateur de calculer, à partir des tableaux sur la variabilité d'échantillonnage, les coefficients de variation approximatifs pour les estimations relatives au nombre, à la proportion ou au pourcentage de la population observée qui possède une caractéristique donnée ainsi que des ratios 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 approprié sur la variabilité d'échantillonnage approximative, il faut repérer l'estimation calculée dans la colonne d'extrême gauche (intitulée « Numérateur du pourcentage ») et suivre les astérisques (s'il y en a) de gauche à droite jusqu'au premier nombre. Puisque les valeurs possibles de l'estimation ne sont pas toutes disponibles, il faut prendre la plus petite et plus proche valeur (par exemple, si l'estimation vaut 1 700 et que les deux plus proches valeurs disponibles sont 1 000 et 2 000, il faut choisir la première). Ce nombre constitue le coefficient de variation approximatif.

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

Le coefficient de variation d'une proportion (ou d'un pourcentage) estimative dépend à la fois de la taille de la proportion et de celle du numérateur utilisées dans le calcul de la proportion. Les proportions estimatives 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 s'appuient sur le chiffre le plus élevé dans une rangée d'un tableau particulier, tandis que les coefficients de variation des estimations du premier type reposent 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 de 65 ans et plus ayant fait au moins deux chutes au cours des douze derniers mois parmi celles en ayant fait au moins une est plus fiable que le nombre estimatif de personnes ayant fait au moins deux chutes.

Lorsque la proportion (ou le pourcentage) est fondée sur la population totale représentée dans un tableau donné, 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 référer à la proportion (haut du tableau) et au numérateur de la proportion (côté gauche du tableau). Puisque les valeurs possibles de la proportion ne sont pas toutes disponibles, il faut prendre la plus petite et plus proche valeur (par exemple, si la proportion est de 23 % et que les deux plus proches valeurs disponibles dans la colonne sont de 20 % et de 25 %, il faut choisir 20 %). Le coefficient de variation se trouve à l'intersection de la rangée et de la colonne appropriées.

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 des sous-groupes indépendants mais n'est autrement qu'approximative. Cette formule mènera à une surestimation de l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélés positivement et à une sous-estimation de l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélés négativement.

Règle 4 : Estimations de ratios

Si le numérateur est un sous-ensemble du dénominateur, il faut convertir le ratio en pourcentage et appliquer la règle 2. Ce serait le cas, par exemple, si le numérateur était le nombre de personnes de 65 ans et plus ayant fait au moins une chute au cours des douze derniers mois et le dénominateur était le nombre de personnes âgées de 65 ans et plus.

Si le numérateur n'est pas un sous-ensemble du dénominateur (par exemple, le ratio du nombre de personnes de 65 ans et plus ayant fait une chute au cours des douze derniers mois au nombre de personnes en ayant fait plus d'une), l'écart-type du ratio 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 ratio des estimations ($\hat{R} = \hat{X}_1 / \hat{X}_2$). L'erreur type d'un ratio 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 ratios

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

10.2 Exemples d'utilisation des tableaux de CV pour des estimations catégorielles

Les exemples réels suivants ont pour but d'aider les utilisateurs à appliquer les règles décrites ci-dessus⁸.

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

Supposons qu'un utilisateur estime à 865 484 le nombre de personnes de 65 ans et plus ayant fait au moins une chute au cours des douze derniers mois. 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 estimatif (865 484) ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le nombre le plus petit qui s'en rapproche le plus, soit 750 000.
- 3) Le coefficient de variation d'un agrégat estimatif (exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 4,9 %.
- 4) Le coefficient de variation approximatif de l'estimation est donc 4,9 %. Par conséquent, selon les « Lignes directrices relatives à la variabilité d'échantillonnage » présentées à la section 9.4, l'estimation selon laquelle 865 484 personnes de 65 ans et plus ont fait au moins une chute au cours des douze derniers mois 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 $865\,484 / 4\,366\,101 = 19,8\%$ le pourcentage de personnes, parmi celles âgées de 65 ans et plus, qui ont fait au moins une chute au cours des douze derniers mois. Comment l'utilisateur fait-il pour déterminer le coefficient de variation de cette estimation?

- 1) Se reporter au tableau de CV pour le CANADA.

⁸ Les valeurs utilisées dans cette section proviennent du fichier maître de l'ESCC - Vieillesse en santé. Une estimation provenant de différents fichiers de micro-données diffèrera légèrement, par contre, la même application concernant les tableaux de coefficient de variation s'applique.

- 2) Parce que l'estimation est un pourcentage qui s'appuie sur un sous-ensemble de la population totale (c.-à-d. les personnes âgées de 65 ans et plus), il faut utiliser à la fois le pourcentage (19,8 %) et le numérateur du pourcentage (865 484) pour déterminer le coefficient de variation.
- 3) Le numérateur (865 484) ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le nombre le plus petit qui s'en rapproche le plus, soit 750 000. De même, l'estimation du pourcentage ne figure pas parmi les en-têtes de colonne; il faut donc utiliser le nombre le plus petit qui s'en rapproche le plus, soit 15,0 %.
- 4) Le nombre qui se trouve à l'intersection de la ligne et de la colonne utilisées, soit 4,8 %, est le coefficient de variation (exprimé en pourcentage) à employer.
- 5) Le coefficient de variation approximatif de l'estimation est donc 4,8 %. Par conséquent, selon les « Lignes directrices relatives à la variabilité d'échantillonnage » présentées à la section 9.4, l'estimation selon laquelle 19,8 % des Canadiens âgés de 65 ans et plus ont fait au moins une chute au cours des douze derniers mois peut être diffusée sans réserve.

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

Supposons qu'un utilisateur estime que, parmi les personnes âgées de 65 à 74 ans, $409\,499 / 2\,407\,219 = 17,0\%$ ont fait au moins une chute au cours des douze derniers mois (estimation 1), alors que chez les personnes de 85 ans et plus, ce pourcentage est estimé à $132\,527 / 490\,775 = 27,0\%$ (estimation 2). Comment l'utilisateur fait-il pour déterminer le coefficient de variation de la différence entre ces deux estimations?

- 1) Le CV de l'estimation 1 est obtenu à l'aide du tableau de CV pour le Canada par groupe d'âge chez les 65 à 74 ans. Puisque la proportion est basée sur la population complète visée par le tableau, le CV de la proportion est le même que celui du numérateur de la proportion, soit 409 499. Puisque ce nombre ne figure pas dans la colonne de gauche, il faut utiliser le plus petit nombre s'en approchant le plus, soit 400 000. Ceci donne un CV de 4,0 % pour l'estimation 1. De façon similaire pour l'estimation 2, un CV de 5,0 % est obtenu en consultant le tableau de CV pour les 85 ans et plus.
- 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 l'estimation 2, et α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement. L'erreur type de la différence $\hat{d} = (0,270 - 0,170) = 0,100$ est donc :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0,270)(0,050)]^2 + [(0,170)(0,040)]^2} \\ &= 0,015\end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,015 / 0,100 = 0,150$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 15,0 % (exprimé en pourcentage). Par conséquent, toujours selon les « Lignes directrices relatives à la variabilité d'échantillonnage » présentées à la section 9.4, cette estimation peut être publiée sans réserve.

Exemple 4 : Estimations de ratios

Supposons qu'un utilisateur estime à 547 815 le nombre de personnes âgées de 65 ans et plus ayant fait une chute au cours des douze derniers mois et à 314 031 le nombre de celles du même groupe d'âge en ayant fait plus d'une. L'utilisateur veut comparer ces deux estimations sous la forme d'un ratio. Comment fait-il pour déterminer le coefficient de variation de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation de ratio, où le numérateur de l'estimation (= \hat{X}_1) est le nombre de personnes de 65 ans et plus ayant fait une chute au cours des douze derniers mois. Le dénominateur de l'estimation (= \hat{X}_2) est le nombre de personnes de 65 ans et plus ayant fait plus d'une chute au cours de la même période.
- 2) Se reporter au tableau de CV pour le CANADA.
- 3) Le numérateur de cette estimation de ratio est 547 815. Le nombre le plus petit qui se rapproche le plus de ce nombre est 500 000. Le coefficient de variation de cette estimation (exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 6,2 %.
- 4) Le dénominateur de cette estimation de ratio est 314 031. Le plus petit nombre qui se rapproche le plus de ce nombre est 300 000. Le coefficient de variation de cette estimation (exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 8,0 %.
- 5) Le coefficient de variation approximatif de l'estimation de ratio est donc donné par la règle 4,

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

c'est-à-dire,

$$\alpha_{\hat{R}} = \sqrt{(0,062)^2 + (0,08)^2}$$

$$= 0,101$$

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement. Le ratio des personnes de 65 ans et plus qui ont chuté une fois au nombre de ceux ayant chuté plus d'une fois est 547 815 / 314 031, soit 1,74:1. Le coefficient de variation de cette estimation est 10,1 % (exprimé en pourcentage); selon les « Lignes directrices relatives à la variabilité d'échantillonnage » présentées à la section 9.4, l'estimation peut donc être diffusée sans réserve.

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

Bien que les coefficients de variation soient largement utilisés, l'intervalle de confiance d'une estimation représente une mesure plus intuitive de l'erreur d'échantillonnage. Un intervalle de confiance est une façon d'énoncer la probabilité que la valeur réelle 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 réelle de la population dans 95 % des cas.

À l'aide de l'erreur type d'une estimation, on peut calculer des intervalles de confiance pour des estimations selon 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 réelle 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 réelle 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 la forme de deux nombres, l'un étant inférieur à l'estimation et l'autre supérieur à celle-ci, sous la forme $(\hat{X} - k, \hat{X} + k)$, où k varie selon le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

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

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

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

$z = 1$ si l'on désire un intervalle de confiance de 68 %

$z = 1,6$ si l'on désire un intervalle de confiance de 90 %

$z = 2$ si l'on désire un intervalle de confiance de 95 %

$z = 3$ si l'on désire un intervalle de confiance de 99 %

Nota : Les lignes directrices sur la diffusion des estimations de la section 9.4 s'appliquent aussi aux intervalles de confiance. Par conséquent, si l'estimation ne peut être diffusée, alors l'intervalle de confiance ne peut l'être non plus.

10.4 Exemple d'utilisation de tableaux de CV pour obtenir des intervalles de confiance

Voici la marche à suivre pour calculer un intervalle de confiance de 95 % pour la proportion estimative de personnes de 65 ans et plus ayant fait au moins une chute au cours des douze derniers mois (d'après l'exemple 2 de la sous-section 10.2).

$$\hat{X} = 0,198$$

$$z = 2$$

$$\alpha_{\hat{X}} = 0,046 \text{ (0,046 est le coefficient de variation de cette estimation selon les tableaux.)}$$

$$CI_{\hat{X}} = \{0,198 - (2) (0,198) (0,046), 0,198 + (2) (0,198) (0,046)\}$$

$$CI_{\hat{X}} = \{0,180, 0,216\}$$

10.5 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 fondées sur un échantillon. Ces estimations peuvent être des nombres, des moyennes, des pourcentages, des ratios, etc. Les tests peuvent être effectués à divers niveaux de signification; un niveau de signification étant 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 fondées sur un échantillon pour deux caractéristiques voulues. Supposons aussi que l'erreur type de la différence $\hat{X}_1 - \hat{X}_2$ est $\sigma_{\hat{d}}$. Si le ratio $z = (\hat{X}_1 - \hat{X}_2) / \sigma_{\hat{d}}$ 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 ratio est inférieur à -2 ou supérieur à +2, la différence observée est significative au niveau 5 %.

10.6 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 de personnes ayant fait au moins une chute au cours des douze derniers chez celles âgées de 65 à 74 ans ET cette même proportion chez les personnes de 85 ans et plus. Dans l'exemple 3 de la sous-section 10.2, nous avons déterminé que l'erreur type de la différence entre ces deux estimations est égale à 0,015. Par conséquent,

$$z = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0,17 - 0,27}{0,015} = \frac{-0,1}{0,015} = -6,7$$

Puisque $z = -6,7$ est inférieur à -2, on doit conclure qu'il existe une différence significative entre les deux estimations au niveau de signification de 5 %. À noter que les deux sous-groupes comparés sont considérés comme étant indépendants, ce qui fait en sorte que le résultat du test est valide.

10.7 Variances ou coefficients de variation exacts

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

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

Le calcul des coefficients de variation (ou tout autre mesure de précision) fait à l'aide de la méthode du bootstrap nécessite toutefois l'accès à de l'information considérée confidentielle qui n'est évidemment pas disponible dans le fichier de microdonnées à grande diffusion. Le calcul doit donc se faire à l'aide du fichier maître. L'accès au fichier maître est traité à la section 11.1.

Un programme macro, appelé « Bootvar », a été développé pour faciliter l'accès à la méthode du bootstrap aux utilisateurs. Le programme Bootvar est offert en formats SAS et SPSS, et est constitué de macros qui calculent les variances de totaux, de ratios, de différences entre les ratios ainsi que les régressions linéaires et logistiques.

Il y a diverses raisons pour lesquelles un utilisateur pourrait souhaiter obtenir une variance exacte. En voici quelques-unes.

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

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

Troisièmement, dans le cas de l'estimation de variables quantitatives, il est nécessaire d'utiliser des tableaux distincts pour déterminer l'erreur d'échantillonnage. Or, la plupart des variables de l'ESCC étant de type catégoriel, 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 exacte. À noter, toutefois, que le coefficient de variation d'un total quantitatif est généralement plus grand que celui de l'estimation catégorielle correspondante (c'est-à-dire, l'estimation du nombre de personnes qui contribuent à l'estimation quantitative). Si l'estimation catégorielle 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 chutes faites au cours des 12 derniers mois par les personnes de 65 ans et plus ayant chuté au moins une fois serait supérieur à celui de l'estimation correspondante du nombre de personnes de 65 ans et plus ayant chuté au moins une fois. 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 variation 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 sur 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 exacte produit le coefficient de variation précis associé à la variable en question.

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

Le document Tableaux de la variabilité d'échantillonnage approximative fourni avec le CD du FMGD présente également les tableaux indiquant les seuils minimums pour des estimations des totaux à l'échelle du Canada et des provinces, ainsi que pour les différents groupes d'âges à l'échelle nationale. Les estimations inférieures à la valeur indiquée dans la colonne « Marginal » ne peuvent en aucun cas être diffusées.

11.0 Fichiers de microdonnées : description, accès et utilisation

L'ESCC – Vieillesse en santé, produit trois types de fichiers de microdonnées : les fichiers maîtres, les fichiers de partage et les fichiers de microdonnées à grande diffusion (FMGD). La diffusion des FMGD est prévue pour 2011.

11.1 Fichiers maîtres

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

11.1.1 Centre de données de recherche

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

11.1.2 Totalisations personnalisées

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

11.1.3 Télé-access

En dernier lieu, le service de télé-access aux fichiers maîtres de l'enquête est un moyen d'accéder à ces données s'il est impossible de passer par un Centre de recherche en données. On peut fournir à l'acheteur d'un produit de microdonnées un fichier maître de données synthétique ou fichier « fictif » et le cliché d'enregistrement correspondant. Le fichier maître de données synthétique ou fichier « fictif » sera disponible en fin de 2010.

Grâce à ces outils, le chercheur peut mettre au point son propre ensemble de programmes analytiques. Il ne lui reste plus qu'à envoyer le code pour les totalisations personnalisées par courrier électronique à cchs-escs@statcan.gc.ca. Le code est transmis au réseau interne protégé de Statistique Canada et traité en regard du fichier maître approprié de données de l'ESCC – Vieillessement en santé. Les estimations générées seront communiquées à l'utilisateur, sujet aux directives sur l'analyse et la communication des données tel qu'exposé dans les grandes lignes à la section 9 de ce document. Les résultats sont filtrés pour vérifier qu'ils sont conformes aux normes de confidentialité et de fiabilité, puis, les données de sortie sont renvoyées au client. Ce service est gratuit.

11.2 Fichiers de partage

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

11.3 Fichiers de microdonnées à grande diffusion

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

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

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

11.4 Utilisation de la variable de pondération

La variable de pondération WTS_M représente le poids d'échantillonnage pour les fichiers principaux de l'enquête. Pour un répondant donné, ce poids d'échantillonnage peut être interprété comme étant le nombre de personnes que le répondant représente dans la population canadienne. Ce poids doit être utilisé en tout temps dans les calculs d'estimations statistiques, afin de permettre l'inférence à l'échelle de la population. La production de résultats non pondérés n'est pas recommandée. La répartition de l'échantillon, de même que les détails du plan de sondage, peuvent entraîner des résultats biaisés qui ne représentent pas correctement la population. Pour une description plus détaillée du calcul de ce poids, consulter la section 7 sur la pondération.

Le module Cognition a été administré en anglais et en français aux répondants non interviewés par personne interposée qui ont consenti à participer à l'enquête. Étant donné que le taux de réponse à ce module était sensiblement inférieur au taux de réponse global à l'enquête, il a été décidé de calculer des poids distincts pour le module Cognition de l'ESCC – Vieillesse en santé. La variable de pondération pour le module Cognition WTS_CM figure dans un fichier distinct. Cette variable doit être utilisée en remplacement de la variable WTS_M pour toutes analyses portant sur le module Cognition. La variable de poids WTS_M doit être utilisée pour toutes analyses ne portant pas sur ce module.

11.5 Convention appliquée pour nommer les variables

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

La convention appliquée pour nommer les variables de l'ESCC a été modifiée à partir de 2007. Ainsi, la lettre correspondant à l'édition de l'enquête (par exemple, A = cycle 1.1, C = cycle 2.1 et E = 3.1) n'est plus utilisée dans les noms de variables. Une nouvelle variable (REFPER, format = AAAAMM-AAAAMM) a été ajoutée aux fichiers de microdonnées afin d'identifier le début et la fin de la période de référence au cours de laquelle les données ont été recueillies.

Les noms de variables sont structurés de la manière suivante :

Positions 1 à 3 :	Nom du module/de la section du questionnaire
Position 4 :	Type de variable (sous-tiret, C, D, F ou G)

Positions 5 à 8 : Numéro de la question et option de réponse (s'il s'agit d'une question à réponse multiple)

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

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

L'exemple 2 présente la structure du nom de la variable correspondant à la question 6 du module Utilisation des soins de santé 2 (HC2_06) qui constitue une question à réponse multiple :

Positions 1 à 3 : HC2 Module sur l'utilisation des soins de santé
Position 4 : _ (sous-tiret = données recueillies)
Position 5 à 8 : 06A numéro de la question correspondant et option de réponse

Dans les positions 1 à 3, on retrouve l'acronyme de chacun des modules. Ces acronymes apparaissent à côté des noms de modules qui sont tous présentés dans la figure de l'annexe 3.3.

La position 4 désigne le type de variable selon qu'il s'agit d'une variable collectée directement à partir d'une question du questionnaire (« _ »), d'une variable codée (« C »), dérivée (« D »), groupée (« G ») ou d'une variable indicatrice (« F »)⁹.

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

⁹ Du mot anglais *flag*.

Tableau 11.1 Désignation des codes utilisés à la 4e position du nom des variables de l'ESCC

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

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

11.6 Convention appliquée pour nommer les variables des enquêtes de l'ESCC précédentes

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

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

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

11.7 Lignes directrices pour l'utilisation des variables d'un sous-échantillon

Pour encourager l'utilisation appropriée des données de l'ESCC – Vieillesse en santé, les poids de pondération pour le module Cognition décrits à la section 11.5 sont fournis dans un fichier distinct. Des poids d'échantillonnage qui leur sont propres sont donnés ainsi qu'un ensemble de poids bootstrap correspondant qu'il faut utiliser pour produire des estimations valables pour toutes les variables dans le fichier. Les poids d'échantillonnage du module Cognition sont calés de telle sorte que leur somme représente la population canadienne totale de 45 ans et plus. Le tableau 11.2 décrit les deux fichiers de données.

Tableau 11.3 Noms et contenu des fichiers de données de l'ESCC 2007

Fichiers	Nom des fichiers	Poids d'échantillonnage	Fichiers des poids bootstrap	Variables incluses	Enregistrements inclus
Fichier maître principal	HS.txt	WTS_M	b5.txt	Tous les modules sauf COG.	Enregistrements de tous les répondants

Fichiers	Nom des fichiers	Poids d'échantillonnage	Fichiers des poids bootstrap	Variables incluses	Enregistrements inclus
Fichier de partage	HS.txt	WTS_S	b5.txt	Tous les modules sauf COG et CLS.	Enregistrements de tous les répondants qui ont accepté de partager leurs données
Fichier maître du module Cognition	HSC.txt	WTS_CM	b5_c.txt	Tous les modules.	Enregistrements de tous les répondants ayant participé au module Cognition
Fichier de partage du module Cognition	HSC.txt	WTS_CS	b5_c.txt	Tous les modules sauf CLS.	Enregistrements de tous les répondants ayant participé au module Cognition qui ont accepté de partager leurs données

11.8 Dictionnaires de données

Des dictionnaires de données distincts comprenant des descriptions d'univers et des fréquences sont fournis pour le fichier maître principal et le fichier de sous-échantillon.

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

Les variables peuvent être estimées à partir du fichier maître, du fichier de partage ou du FMGD. Selon le fichier utilisé, des différences très faibles seront observées.

Toutes les estimations officielles de Statistique Canada des variables, sauf celles du module Cognition, sont fondées sur le poids d'échantillonnage du fichier maître principal. Toutes les estimations officielles de Statistique Canada portant sur les variables du module Cognition sont fondées sur les poids d'échantillonnage du fichier maître de données du module Cognition.