



**Enquête sur la santé dans les
collectivités canadiennes (ESCC)**

Cycle 2.2 (2004)

**Nutrition : Volets général sur la santé (y compris les
suppléments de vitamines et de minéraux) et du
rappel alimentaire de 24 heures**

Guide de l'utilisateur

Avril 2008



Statistique
Canada

Statistics
Canada

Canada

Quoi de neuf

1. Fichiers sur les suppléments de vitamines et de minéraux

La présente diffusion des données de l'ESCC 2.2 contient trois nouveaux fichiers ayant trait au module des suppléments de vitamines et de minéraux du volet général sur la santé. Il s'agit du fichier de description des vitamines et des minéraux (VDC), du fichier de données détaillées sur les suppléments de vitamines et de minéraux (VMD) et du fichier de données sommaires sur les vitamines et les minéraux (VSD). En outre, les fichiers HS et HS-SIDE contiennent des indicateurs qui montrent si un répondant a consommé un nutriment particulier sous forme de supplément.

2. Mises à jour de deux variables des suppléments de vitamines et de minéraux diffusées antérieurement

Deux variables des suppléments de vitamines et de minéraux qui ont été diffusées antérieurement dans le cadre de la deuxième vague ont été corrigées pour la présente diffusion. Il s'agit de VSDD_01 (Au cours du dernier mois, avez-vous pris des vitamines ou des minéraux?) et VSDD_13 (Au cours du dernier mois, combien de produits de vitamines et minéraux différents avez-vous pris?).

3. Nouvelles variables dérivées sur le revenu

La présente diffusion contient trois nouvelles variables dérivées sur le revenu. Il s'agit de INCDDADR (Ratio ajusté du revenu du ménage – échelle nationale), INCDDRCA (Distribution du revenu du ménage – échelle nationale) et INCDDRPR (Distribution du revenu du ménage – échelle provinciale). La variable INCDDADR fournit un ratio ajusté du revenu du ménage au seuil de faible revenu. Les variables INCDDRCA et INCDDRPR fournissent, pour chaque répondant, une mesure relative du revenu de leur ménage par rapport aux revenus des ménages pour l'ensemble des autres répondants, aux niveaux national et provincial. Est également incluse une variable géographique de taille de population, qui est utilisée dans les calculs pour créer les nouvelles variables dérivées sur le revenu. Des renseignements supplémentaires sur ces nouvelles variables figurent dans la documentation sur les variables dérivées.

4. Valeurs révisées de l'apport en vitamine D pour certains codes d'aliments

Les valeurs de l'apport en vitamine D pour les codes d'aliments 3208 - corégone de lac, espèces diverses, crues, 3169 - corégone de lac, espèces diverses, four/grillées, 3088 - corégone de lac, espèces diverses, fumées, et 10913 - boulette de poisson ont été révisées depuis la dernière diffusion. Cette révision aura une incidence sur les variables suivantes : FDCD_DMG, FIDD_DMG et FSDDDDMG.

5. Valeurs révisées de l'apport en acide linoléique pour certains codes d'aliment

La valeur de l'apport en acide linoléique pour le code d'aliment 554 - vinaigrette, type mayonnaise, commerciale, réduite en matière grasse, a été révisée depuis la dernière diffusion. Cette révision aura une incidence sur les variables suivantes : FDCD_FAN, FIDD_FAN, FSDDDFAN et FSDDDENI.

6. Valeurs mises à jour de l'apport en sucres pour certains aliments

Les valeurs pour les variables FDCD_SUG, FIDD_SUG et FSDDDSUG ont été mises à jour depuis la dernière diffusion. Cette mise à jour réduira la sous-couverture de l'apport en sucres dans la diffusion des données de la vague 2. Par conséquent, le nombre de valeurs manquantes pour le sucre sera plus faible que dans le cas de la diffusion précédente.

7. Révisions de l'apport alimentaire d'aspartame pour certains répondants

Les valeurs de l'apport d'aspartame étaient codées incorrectement dans les fichiers de la deuxième vague pour huit cas. Ces valeurs ont été révisées pour la présente diffusion. Ces révisions donneront lieu à des valeurs modifiées pour les profils nutritionnels de ces personnes.

8. Valeurs de nutriments corrigées pour certaines préparations pour nourrissons

Les valeurs de nutriments ont été corrigées pour certaines préparations pour nourrissons depuis la dernière diffusion. Ces corrections auront une incidence sur les codes d'aliments suivants: 501026, 501031, 501037, 501042, et 501049.

9. Codes du Bureau des sciences de la nutrition (BSN)

Environ 2 % des codes de groupes d'aliments et de recettes ont été corrigés depuis la dernière diffusion. Ces changements auront une incidence sur les variables suivantes : FDCD_FGE, FDCD_FGF, FIDD_FGR, FIDD_GR1, FIDD_GR2, FIDD_GR3 et FIDD_GR4.

10. Variables mises à jour dans le fichier HS SIDE

Dans la diffusion des données de la vague 2, environ 350 enregistrements contenaient des codes de non-réponse (99999.999999) pour les variables suivantes : FSDDDEAL, FSDDDECA, FSDDDEEI, FSDDDELI, FSDDDEMO, FSDDDENI, FSDDDEPO, FSDDDEPR et FSDDDESA. Ces valeurs ont maintenant été fixées à manquantes (« . ») afin d'assurer l'uniformité avec les autres valeurs de non-réponse.

Table des matières

1.	INTRODUCTION	1
2.	CONTEXTE ET OBJECTIFS DE L'ESCC	2
3.	ENQUÊTE SUR LA NUTRITION – CONTEXTE ET OBJECTIFS.....	5
4.	CONTENU DE L'ENQUÊTE	7
4.1	VOLET « GÉNÉRAL SUR LA SANTÉ»	7
4.2	VOLET « RAPPEL ALIMENTAIRE DE 24 HEURES »	14
5.	PLAN D'ÉCHANTILLONNAGE	16
5.1	POPULATION CIBLE.....	16
5.2	TAILLE ET RÉPARTITION DE L'ÉCHANTILLON.....	16
5.3	ACHATS PROVINCIAUX D'UNITÉS D'ÉCHANTILLON.....	17
5.4	ÉCHANTILLONNAGE DES MÉNAGES À PARTIR DES DIFFÉRENTES BASES DE SONDAGE	18
5.4.1	ÉCHANTILLONNAGE DES MÉNAGES À PARTIR DE LA BASE ARÉOLAIRE	18
5.4.2	ÉCHANTILLONNAGE DES MÉNAGES DE LA BASE LISTE CRÉÉE À PARTIR DE L'ESCC (CYCLE 2.1).....	20
5.4.3	ÉCHANTILLONNAGE DES MÉNAGES DE LA BASE LISTE D'ADRESSES DE LOGEMENT CRÉÉE À PARTIR DES REGISTRES DE SOINS DE SANTÉ PROVINCIAUX	20
5.5	ÉCHANTILLONNAGE DES RÉPONDANTS	21
5.6	RÉPARTITION DE L'ÉCHANTILLON SUR LA PÉRIODE DE COLLECTE DE DONNÉES.....	21
5.7	ACHAT SPÉCIAL D'UNITÉS D'ÉCHANTILLON DE LA PART DE SANTÉ CANADA	21
5.8	ÉVALUATION DE LA VARIABILITÉ INTRA-INDIVIDUELLE	22
6.	COLLECTE DES DONNÉES	23
6.1	CONCEPTION DU QUESTIONNAIRE	23
6.2	ESSAI SUR LE TERRAIN	23
6.3	PERSONNEL DE COLLECTE (FORMATION, SUPERVISION ET CONTRÔLE)	23
6.4	MÉTHODE DE COLLECTE DES DONNÉES	24
6.5	INTERVIEWS PAR PROCURATION.....	25
6.6	MINIMISER LA NON-RÉPONSE	26
6.7	CIRCONSTANCES SPÉCIALES DURANT LA COLLECTE DES DONNÉES	26
6.7.1	MESURE DE LA TAILLE ET DU POIDS.....	26
6.7.2	COLLECTE DE DONNÉES SUR LES SUPPLÉMENTS DE VITAMINES ET DE MINÉRAUX.....	29
6.7.3	APPLICATION INFORMATISÉE POUR LE RAPPEL ALIMENTAIRE DE 24 HEURES	29
6.7.3.1	CARACTÉRISTIQUES DE L'AMPM SERVANT À RECUEILLIR DES DÉTAILS SUR LES ALIMENTS ET SUR LES QUANTITÉS CONSOMMÉS	30
6.7.4	INTERVIEW DES ENFANTS ET DES JEUNES.....	31
6.8	RESPECT DE LA VIE PRIVÉE.....	31
7.	TRAITEMENT DES DONNÉES	32
7.1	VÉRIFICATION	32
7.1.1	VÉRIFICATION DES DONNÉES DU VOLET « GÉNÉRAL SUR LA SANTÉ »	32
7.1.2	VÉRIFICATION DES DONNÉES SUR LES SUPPLÉMENTS DE VITAMINES ET DE MINÉRAUX.....	32
7.1.3	CALCUL DE LA CONCENTRATION DU SUPPLÉMENT	33
7.1.4	VÉRIFICATION DES DONNÉES DU « RAPPEL ALIMENTAIRE DE 24 HEURES ».....	34
7.2	EXIGENCES MINIMALES DE RÉPONSE	34
7.3	CODAGE DES QUESTIONS OUVERTES.....	35
7.3.1	QUESTIONS NON LIÉES AUX ALIMENTS.....	35
7.3.2	CODAGE DES SUPPLÉMENTS DE VITAMINES ET DE MINÉRAUX.....	35

7.3.3	QUESTIONS LIÉES AUX ALIMENTS	36
7.4	NSS ET LE CODAGE DES ALIMENTS	36
7.4.1	CODAGE AUTOMATISÉ DES ALIMENTS	36
7.4.2	CODAGE MANUEL DES ALIMENTS.....	37
7.4.3	CALCUL DES TAILLES DES PORTIONS.....	38
7.4.4	IMPORTATION DES DONNÉES SUR LES ALIMENTS DANS LE NSS	39
7.4.5	VUE D'ENSEMBLE : NUTRITION SURVEY SYSTEM (NSS)	39
7.5	VÉRIFICATION/VALIDATION DES DONNÉES SUR LES ALIMENTS.....	40
7.6	VÉRIFICATION/VALIDATION DES DONNÉES SUR LES SUPPLÉMENTS DE VITAMINES ET DE MINÉRAUX.....	41
7.7	CRÉATION DE VARIABLES DÉRIVÉES ET DE VARIABLES REGROUPÉES	42
7.8	PONDÉRATION	42
7.9	SUPPRESSION DE RENSEIGNEMENTS CONFIDENTIELS.....	42
8.	STRUCTURE DES FICHIERS DE DONNÉES	43
8.1	DESCRIPTION DES FICHIERS DE DONNÉES.....	43
8.2	VARIABLES CLÉS IMPORTANTES ET DÉFINITIONS.....	49
8.3	VARIABLES CLÉS POUR LE COUPLAGE DES FICHIERS DE DONNÉES.....	49
8.4	PRINCIPAUX FICHIERS DE DONNÉES SUR LES ALIMENTS (FID ET FRL) ET VARIABLES CLÉS	50
9.	PONDÉRATION	53
9.1	PONDÉRATION DE L'ÉCHANTILLON	53
9.1.1	PONDÉRATION DE L'ÉCHANTILLON PROVENANT DE LA BASE ARÉOLAIRE.....	55
9.1.2	PONDÉRATION DE LA BASE LISTE DU CYCLE 2.1	57
9.1.3	PONDÉRATION DU REGISTRE DE SANTÉ DE L'ÎLE-DU-PRINCE-ÉDOUARD	60
9.1.4	PONDÉRATION DU REGISTRE DE SANTÉ DU MANITOBA	61
9.1.5	PONDÉRATION DES CAS PROVENANT DE LA BASE SUPPLÉMENTAIRE AUTOCHTONE	63
9.1.6	INTÉGRATION DE LA BASE ARÉOLAIRE AVEC LES BASES LISTES (I1).....	65
9.1.7	INTÉGRATION DE LA BASE SUPPLÉMENTAIRE AUTOCHTONE AVEC LES AUTRES BASES (I2)	67
9.1.8	EFFET SAISONNIER ET DU JOUR DE LA SEMAINE (I3).....	68
9.1.9	POSTSTRATIFICATION (I4)	69
9.2	PONDÉRATION DU SOUS-ÉCHANTILLON AYANT DES MESURES DE TAILLES ET DE POIDS	69
10.	QUALITÉ DES DONNÉES.....	71
10.1	LIMITES DES DONNÉES	71
10.1.1	LIMITES DU RAPPEL ALIMENTAIRE DE 24 HEURES.....	71
10.1.2	LIMITES DES DONNÉES SUR LES SUPPLÉMENTS DE VITAMINES ET DE MINÉRAUX.....	72
10.2	TAUX DE RÉPONSE	73
10.3	ERREURS DANS LES ENQUÊTES	76
10.3.1	ERREURS NON DUES À L'ÉCHANTILLONNAGE	76
10.3.2	ERREURS DUES À L'ÉCHANTILLONNAGE.....	76
11.	LIGNES DIRECTRICES POUR LA TOTALISATION, L'ANALYSE ET LA DIFFUSION	78
11.1	LIGNES DIRECTRICES POUR L'ARRONDISSEMENT	78
11.2	LIGNES DIRECTRICES POUR LA PONDÉRATION DE L'ÉCHANTILLON EN VUE DE LA TOTALISATION	79
11.2.1	DÉFINITIONS DES CATÉGORIES D'ESTIMATIONS : DE TYPE NOMINAL PAR OPPOSITION À QUANTITATIVES.....	79
11.2.2	TOTALISATION D'ESTIMATIONS DE TYPE NOMINAL.....	80
11.2.3	TOTALISATION D'ESTIMATIONS QUANTITATIVES	80
11.3	LIGNES DIRECTRICES POUR L'ANALYSE STATISTIQUE	81
11.4	SOFTWARE FOR INTAKE DISTRIBUTION (SIDE).....	81
11.5	LIGNES DIRECTRICES POUR LA DIFFUSION.....	82
12.	TABLEAUX DE LA VARIABILITÉ D'ÉCHANTILLONNAGE APPROXIMATIVE.....	83
12.1	COMMENT UTILISER LES TABLEAUX DE CV POUR LES ESTIMATIONS DE TYPE NOMINAL.....	83

12.2	EXEMPLES D'UTILISATION DES TABLEAUX DE CV POUR DES ESTIMATIONS DE TYPE NOMINAL	86
12.3	COMMENT UTILISER LES TABLEAUX DE CV POUR CALCULER LES LIMITES DE CONFIANCE	88
12.4	EXEMPLE D'UTILISATION DE TABLEAUX DE CV POUR OBTENIR DES LIMITES DE CONFIANCE	89
12.5	COMMENT UTILISER LES TABLEAUX DE CV POUR EFFECTUER UN TEST Z	90
12.6	EXEMPLE D'UTILISATION DES TABLEAUX DE CV POUR EFFECTUER UN TEST Z	90
12.7	VARIANCES OU COEFFICIENTS DE VARIATION EXACTS	90
12.8	SEUILS POUR LA DIFFUSION DES ESTIMATIONS RELATIVES À L'ESCC	92
13.	UTILISATION DU FICHIER.....	93
13.1	POIDS D'ÉCHANTILLONNAGE PRINCIPAL	93
13.2	POIDS D'ÉCHANTILLONNAGE POUR LES MESURES EXACTES DE L'INDICE DE MASSE CORPORELLE	93
13.3	CONVENTION APPLIQUÉE POUR NOMMER LES VARIABLES.....	95
13.3.1	STRUCTURE ÉLÉMENTAIRE DES NOMS DES VARIABLES DE L' ESCC	95
13.3.2	POSITIONS 1 À 3 : NOM DE LA VARIABLE/SECTION DU QUESTIONNAIRE	95
13.3.3	POSITION 4 : CYCLE	96
13.3.4	POSITION 5 : TYPE DE VARIABLE	96
13.3.5	POSITIONS 6 À 8 : NOM DE LA VARIABLE.....	96
13.4	ACCÈS AU FICHIER MAÎTRE.....	97
	ANNEXE 1 : QUESTIONS ANALYTIQUES.....	99
1.	DONNÉES SUR LES SUPPLÉMENTS VITAMINIQUES ET MINÉRAUX	99
2.	ÉTAT DE SÉCURITÉ ALIMENTAIRE DES MÉNAGES	104

L'information fournie dans le présent guide comprend les données de l'Enquête sur la santé dans les collectivités canadiennes, cycle 2.2 (2004) – Nutrition, incluant toutes les données diffusées en juillet 2005, juillet 2006 et en avril 2008. En raison de leur nature confidentielle, quelques-unes de ces données sont uniquement disponibles dans les établissements de Statistique Canada, y compris les Centres de données de recherche.

Pour obtenir plus de renseignements, communiquez avec le *Service d'information et d'accès aux données*, Division de la statistique de la santé au cchs-escr@statcan.ca.

1. Introduction

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

Le présent fichier de microdonnées contient les données de la quatrième année de collecte de l'ESCC (cycle 2.2). Les renseignements ont été recueillis de janvier 2004 à janvier 2005 dans les dix provinces. Les données du cycle 2.2 de l'ESCC sont recueillies auprès des personnes de tous les âges vivant dans des logements privés. Sont exclus de la base de sondage 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 plein des Forces canadiennes et les personnes vivant dans certaines régions éloignées.

Le présent document a pour but de faciliter la manipulation du fichier de microdonnées du cycle 2.2 de l'ESCC qui est décrit en détails dans le texte et les annexes qui suivent.

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

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

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

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

2. Contexte et objectifs de l'ESCC

En 1991, le Groupe de travail national sur l'information en matière de santé a relevé plusieurs problèmes posés par le système d'information sur la santé. Selon ses membres, les données étaient :

- fragmentées et incomplètes,
- ne pouvaient être partagées facilement,
- n'étaient pas analysées aussi pleinement que possible, et
- les résultats des études réalisées n'atteignaient pas de façon régulière la population canadienne.¹

Pour résoudre ces problèmes, l'Institut canadien d'information sur la santé (ICIS), Statistique Canada et Santé Canada ont conjugué leurs efforts en vue de créer un Carnet de route de l'information sur la santé.

L'initiative du Carnet de route a été lancée en réponse directe aux préoccupations et aux souhaits exprimés par plus de 500 personnes représentant un large éventail d'organismes et de groupes d'intérêt. Au début de 1999, les trois organismes nationaux susmentionnés ont mené une consultation nationale à grande échelle sur les besoins d'information en matière de santé. Les participants ont insisté sur le fait que les organismes nationaux doivent collaborer en vue de renforcer le système canadien d'information sur la santé et mettre à profit les investissements et les compétences considérables aux niveaux local, régional et provincial/territorial.²

Le Carnet de route représente une contribution importante à l'édification d'un système national complet d'information sur la santé et de l'infrastructure requise pour donner aux Canadiens l'information dont ils ont besoin pour entretenir et améliorer le système de santé et la santé de la population du Canada.³ Un plan d'action coordonné est requis. Le gouvernement seul ou une seule organisation ne peut pas lutter contre les problèmes mentionnés plus haut. La collaboration à tous les niveaux — organismes de santé nationaux, provinciaux, territoriaux, régionaux et locaux — est une condition préalable au succès.⁴

Le plan d'action commence par poser deux questions cruciales :⁵

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

La première question englobe l'efficacité, l'efficience et la réceptivité du système de santé. En règle générale, un système de santé efficace, efficient et réceptif est un système qui offre aux Canadiens les soins de qualité auxquels ils s'attendent.⁶

La deuxième question est plus générale et traite des objectifs de base du système : la santé des Canadiens s'améliore-t-elle? Afin de répondre à cette question nous avons besoin d'un système solide d'information sur la santé.⁷ Ce système doit posséder six grandes caractéristiques.⁸

1. Institut canadien d'information sur la santé, Santé Canada et Statistique Canada. 1999. *Carnet de route de l'information sur la santé : répondre aux besoins*. p. 3. ISBN 1-895581-30-3. (<http://www.cihi.ca>)

2. Ibid. p. 1.

3. Ibid. p. 1.

4. Ibid. p. 3.

5. Ibid. p. 3.

6. Ibid. p. 3.

7. Ibid. p. 5.

Le système d'information doit être:

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

Ce nouveau système d'information sur la santé doit être à jour, fournir des renseignements orientés vers la personne et s'appuyer sur des normes de données communes à d'autres enquêtes sur la santé de la population canadienne, telles que l'Enquête nationale sur la santé de la population (ENSP). Il doit également fournir de nouveaux ensembles de données ou des ensembles de données étoffées, des données sur les services de santé, des données sur les résultats relatifs à la santé, l'état de santé et les déterminants non médicaux de la santé, des données sur les résultats d'interventions particulières, des études spéciales portant sur des questions prioritaires, des données sur les coûts selon le service, des protocoles d'échange de données, une plus grande capacité d'analyse des données, ainsi que des rapports publics sur le système de santé.⁹

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

Les objectifs principaux de l'ESCC sont les suivants :

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

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

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

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

9. Ibid. p. 11-14.

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

Le présent Guide est publié en réponse à l'exigence de recueillir des données fiables et pertinentes sur les services de santé, l'état de santé et les questions relatives à la santé revêtant une importance pour la population canadienne et de diffuser cette information au public.

3. Enquête sur la nutrition – contexte et objectifs

L'alimentation est considérée comme un facteur de risque pour la santé qui touche tous les Canadiens et qu'il est possible de modifier. La dernière enquête nationale auprès de la population ayant trait à la consommation d'aliments et la dernière évaluation connexe de la nutrition ont été effectuées en 1972. Tout au long des années 90, Santé Canada a collaboré avec chacune des provinces à la réalisation d'enquêtes visant l'évaluation de l'alimentation. En raison de nombreux facteurs, les données de ces enquêtes provinciales sur la nutrition n'ont pu fournir de l'information nationale complète sur les habitudes alimentaires des Canadiens.

Compte tenu du besoin impératif de disposer de renseignements plus détaillés et plus actuels sur l'alimentation des Canadiens, il a été décidé que la nutrition serait le thème du cycle 2.2 de l'ESCC. Cette enquête, qui ciblait un échantillon de 30 000 répondants appartenant à tous les groupes d'âge (0 et plus) vivant dans un logement privé occupé, dans l'une des dix provinces, a été effectuée du 14 janvier 2004 au 21 janvier 2005. Les résidents des trois territoires, les personnes qui vivent dans des réserves indiennes ou sur des terres de la Couronne, les pensionnaires d'un établissement, les membres à temps plein des Forces canadiennes et les résidents de certaines régions éloignées ont été exclus de la base de sondage.

L'Enquête sur la nutrition vise surtout à fournir une information fiable et actuelle sur l'apport alimentaire et le bien-être nutritionnel, de même que sur leurs déterminants clés afin de documenter et d'orienter les programmes, les politiques et les activités des gouvernements fédéral et provinciaux ainsi que ceux des services locaux de santé.

Voici certains des principaux objectifs de cette enquête :

- estimer la distribution de l'apport alimentaire habituel selon les aliments, les groupes d'aliments, les compléments alimentaires, les éléments nutritifs et les habitudes alimentaires dans un échantillon représentatif de Canadiens aux échelons national et provincial;
- recueillir des mesures anthropométriques pour évaluer précisément la masse corporelle;
- mesurer la prévalence de l'insécurité alimentaire¹¹ des ménages;
- recueillir des données sur certains problèmes de santé et sur les caractéristiques socioéconomiques et démographiques des répondants.

L'enquête qui en résulte se compose de deux volets distincts mais complémentaires, soit (1) le volet **général sur la santé** (y compris les suppléments de vitamines et de minéraux), et (2) le volet du **rappel alimentaire de 24 heures**.

Comme son nom l'indique, le volet général sur la santé a recueilli de l'information sur l'état de santé général du répondant, y compris les problèmes de santé chroniques, ainsi que sur des comportements liés à la santé comme la consommation des suppléments de vitamines et de minéraux, la consommation de fruits et de légumes, les activités physiques et sédentaires, le tabagisme et la consommation d'alcool. Ce volet a également permis la collecte des mesures physiques de la taille et du poids des répondants ainsi que leurs caractéristiques sociodémographiques.

11. Le terme « insécurité alimentaire » renvoie à l'insuffisance d'aliments adéquats sur le plan nutritif, où à l'incertitude quant à la possibilité de s'en procurer, ou à la capacité limitée d'acquérir des aliments acceptables de façon acceptable sur le plan social, ou à l'incertitude quant à cette capacité.

Le volet du rappel alimentaire de 24 heures a été développé de sorte à recueillir de l'information sur tous les aliments et boissons consommés au cours de la période de 24 heures, de minuit à minuit, le jour précédant l'interview. On a demandé aux répondants de fournir des renseignements sur le moment où les aliments ont été consommés, l'occasion (p. ex., petit déjeuner, dîner), sur ce qui était ajouté aux aliments (p.ex., du beurre sur les rôties), la description détaillée des aliments, les quantités consommées et si le repas a été préparé à la maison ou ailleurs.

4. Contenu de l'enquête

La présente section résume de façon générale le contenu choisi pour l'Enquête sur la nutrition. Le contenu de cette enquête a été élaboré à partir des consultations qui ont constamment eu lieu depuis le début de 2000 jusqu'à la fin de 2003 entre Statistique Canada et les membres d'un groupe consultatif d'experts. Ces membres venaient de partout au Canada, notamment du Groupe de travail du Système de surveillance alimentaire et nutritionnelle de Santé Canada, de membres du Groupe fédéral-provincial-territorial en nutrition, de spécialistes des enquêtes provinciales sur la nutrition, de nutritionnistes régionaux travaillant dans les services de santé locaux et d'universitaires. Le questionnaire d'enquête, résultat final de ces consultations, comporte deux volets : (1) le **Volet général sur la santé** et (2) le **Rappel alimentaire de 24 heures**.

Le volet général sur la santé recueille de l'information telle que la taille et le poids des répondants, leurs activités physiques, leurs problèmes de santé chroniques, leurs consommations des suppléments de vitamines et de minéraux ainsi que leurs caractéristiques sociodémographiques. Le volet Rappel alimentaire de 24 heures recueille de l'information sur tous les aliments et toutes les boissons que le répondant a consommés au cours des 24 heures précédant l'enquête, de minuit à minuit. Les répondants sont questionnés sur le moment où les aliments ont été consommés, le nom de l'occasion de manger (p. ex., déjeuner, dîner), ce qui a été ajouté aux aliments (p. ex., du beurre sur les rôties), la description détaillée des aliments et les quantités d'aliments qu'ils ont consommés, et si le repas était préparé à la maison ou ailleurs.

Trente pour cent environ des répondants de tous les groupes d'âge ont participé à un deuxième Rappel alimentaire de 24 heures. Cette deuxième interview – en général réalisée de 3 à 10 jours après l'entrevue initiale – sert à mesurer à quel point l'alimentation d'une personne se modifie d'une journée à l'autre. Veuillez remarquer que pendant cette deuxième interview, seul le module Rappel alimentaire de 24 heures est administré.

Vous trouverez ci-dessous un résumé de chacune des composantes du questionnaire.

4.1 Volet « général sur la santé »

Le volet général sur la santé est formé des 18 modules énumérés ci-dessous dans l'ordre dans lequel ils ont été présentés aux répondants :

- 1 – État de santé général
- 2 – Activités physiques
- 3 – Activités sédentaires
- 4 – Activités physiques des enfants
- 5 – Taille et poids, autodéclaration
- 6 – Suppléments de vitamines et minéraux
- 7 – Détails des suppléments de vitamines et minéraux
- 8 – Taille et poids mesurés
- 9 – Santé des femmes
- 10 – Consommation de fruits et de légumes
- 11 – Problèmes de santé chroniques
- 12 – Usage du tabac
- 13 – Consommation d'alcool
- 14 – Sécurité alimentaire
- 15 – Caractéristiques sociodémographiques

- 16 – Population active
- 17 – Revenu
- 18 – Administration

Le tableau 4.1 ci-dessous décrit brièvement chacun des modèles ainsi que l'univers auquel le module s'applique.

Tableau 4.1 – Description sommaire des modules du volet général sur la santé

	Nom du module/ acronyme	Source et description sommaire du module
1	<p>État de santé général (GEN)</p> <p><u>Univers</u> : Répondants âgés de 12 ans et plus sauf la question GEN_Q07, qui s'applique aux 15 ans et plus. Cette question porte sur le niveau de stress dans la vie des répondants.</p>	<p>Source : Module sur l'état de santé général du cycle 2.1 de l'ESCC (l'Enquête sur la nutrition n'a pas posé toutes les questions).</p> <p>Aperçu : Les questions de ce module portent sur l'état de santé général, le niveau de stress et le sentiment d'appartenance à la collectivité locale.</p>
2	<p>Activités physiques (PAC)</p> <p><u>Univers</u> : 12 ans et plus</p>	<p>Source : Module sur les activités physiques du cycle 2.1 de l'ESCC (l'Enquête sur la nutrition n'a pas posé toutes les questions).</p> <p>Aperçu : Ce module recueille de l'information sur les activités physiques effectuées par les répondants pendant leurs loisirs. Les réponses à ce module servent au calcul de l'indice des activités physiques du répondant.</p>
3	<p>Activités sédentaires (SAC)</p> <p><u>Univers</u> : Jeunes âgés de 12 à 17 ans</p>	<p>Source : Module sur les activités sédentaires du cycle 2.1 de l'ESCC. Nota : Dans le cycle 2.1 de l'ESCC, ce module s'applique aux répondants âgés de 12 ans et plus, tandis que dans l'Enquête sur la nutrition, il s'applique aux jeunes âgés de 12 à 17 ans.</p> <p>Aperçu : Le terme « activité sédentaire » renvoie au temps passé assis à ne rien faire ou à faire peu d'activité physique. Dans ce module, on demande aux répondants combien d'heures par semaine, au cours des trois mois précédant l'enquête, ils sont en général restés assis à un ordinateur pendant leurs loisirs. On leur demande aussi combien d'heures par semaine ils ont joué à des jeux vidéos, ont regardé la télévision ou ont lu pour le plaisir.</p>

	Nom du module/ acronyme	Source et description sommaire du module
4	<p>Activité physique des enfants (CPA)</p> <p><u>Univers</u> : Enfants âgés de 6 à 11 ans</p>	<p>Source : Adapté du questionnaire de l'Enquête sur les comportements liés à la santé des enfants d'âge scolaire de 2001-2002.¹²</p> <p>Aperçu : L'activité physique des enfants diffère de celle des adultes. Ce module porte sur l'activité physique des enfants <u>de 6 à 11 ans</u>, à l'école et en dehors de l'école. Des questions ont aussi été posées sur la fréquence à laquelle les enfants écoutent la télévision et le temps qu'ils passent à l'ordinateur.</p>
5	<p>Taille et poids, autodéclaration (HWT)</p> <p><u>Univers</u> : Répondants âgés de 18 ans et plus, sous-échantillon</p>	<p>Source : Module sur la taille et le poids du cycle 2.1 de l'ESCC. Nota : Dans le cycle 2.1 de l'ESCC, ce module s'applique aux répondants de 12 ans et plus, tandis que dans l'Enquête sur la nutrition, il s'applique aux personnes de 18 ans et plus.</p> <p>Aperçu : Ce module comporte une série de questions demandant aux répondants de 18 ans et plus de faire état de leur taille et de leur poids.</p> <p>Nota : L'Enquête sur la nutrition comporte deux modules sur la taille et le poids : dans celui qui s'appelle « Taille et poids, autodéclaration » (HWT), on demande au répondant de <u>déclarer lui-même</u> sa taille et son poids; dans l'autre, appelé « Taille et poids mesurés » (MHW), l'intervieweur mesure la taille et le poids du répondant.</p> <p>Les questions du module MHW sont posées à tous les répondants âgés de 2 ans et plus, mais <u>les deux</u> modules, HWT et MHW, sont demandés à 10 % seulement de l'échantillon des personnes âgées de 18 ans et plus. On demande les deux modules, HWT et MHW, à certains répondants adultes pour essayer d'évaluer la différence entre la taille et le poids autodéclarés et mesurés.</p>

12. L'Enquête sur les comportements liés à la santé des enfants d'âge scolaire est parrainée par l'Organisation mondiale de la Santé (OMS). Cette enquête, effectuée tous les quatre ans dans plus d'une vingtaine de pays, est une étude multinationale sur les enfants d'âge scolaire. Elle est menée au Canada depuis 1990 par l'Université Queen's, en partenariat avec Santé Canada. Il est possible d'en apprendre davantage en ligne, à l'adresse www.phac-aspc.gc.ca/dca-dea/7-18yrs-ans/sante_ecole_f.html.

	Nom du module/ acronyme	Source et description sommaire du module
6	<p>Vitamines et suppléments minéraux (NSP)</p> <p><u>Univers</u> : Tous les âges</p>	<p>Source : Il s'agit d'un nouveau module conçu expressément pour le cycle 2.2 de l'ESCC.</p> <p>Aperçu : Ce module comporte des questions sur les suppléments de vitamines et de minéraux que le répondant a pris au cours du mois écoulé. Veuillez noter que l'enquête n'a pas recueilli d'information sur les préparations à base de plantes médicinales (p. ex., ginkgo, ginseng, millepertuis). Pour plus de renseignements sur la combinaison des données sur les aliments et les suppléments, veuillez consulter l'annexe.</p>
7	<p>Détails des vitamines et suppléments minéraux (NSD)</p> <p><u>Univers</u> : Tous les âges</p>	<p>Source : Il s'agit d'un nouveau module conçu expressément pour le cycle 2.2 de l'ESCC.</p> <p>Aperçu : Ce module constitue la suite du module sur les suppléments de vitamines et de minéraux (NSP); il recueille des précisions pour chacune des vitamines et chacun des suppléments minéraux dont il est fait état, notamment l'identification numérique de la drogue (DIN). Il comporte de plus des questions sur la fréquence et le montant auxquels chacun des suppléments a été pris. (Veuillez noter que l'application permettait aux répondants de donner des détails pour 10 suppléments au maximum.) Pour plus de renseignements sur la combinaison des données sur les aliments et les suppléments, veuillez consulter l'annexe.</p>
8	<p>Taille et poids mesurés (MHW)</p> <p><u>Univers</u> : 2 ans et plus</p>	<p>Source : Il s'agit d'un nouveau module conçu expressément pour le cycle 2.2 de l'ESCC.</p> <p>Aperçu : Après avoir obtenu la permission du répondant, l'intervieweur mesure la taille et le poids du répondant à l'aide d'un pèse-personne et d'un ruban à mesurer. Dans les cas où les répondants ne veulent pas être mesurés, l'intervieweur leur demande de <u>déclarer eux-mêmes</u> leur taille et leur poids. Si les répondants ne peuvent pas être mesurés (p. ex., ne peuvent pas se tenir debout sans aide ou sont en fauteuil roulant), l'intervieweur leur demande de déclarer eux-mêmes leur taille et leur poids.</p>

	Nom du module/ acronyme	Source et description sommaire du module
9	<p>Santé des femmes (WHC)</p> <p><u>Univers</u> : Filles et femmes âgées de 9 ans et plus</p>	<p>Source : Les questions WHC_Q03, WHC_Q04, WHC_Q05 et WHC_Q08 sont adaptées de cycles antérieurs de l'ESCC. Les autres questions sont nouvelles et conçues expressément pour le cycle 2.2 de l'ESCC.</p> <p>Aperçu : Les besoins nutritifs des femmes et des jeunes filles varient aux différentes étapes de leur vie. Par exemple, elles ont besoin de plus d'acide folique lorsqu'elles sont enceintes et de plus de fer durant les années pendant lesquelles elles sont menstruées. Ce module recueille de l'information auprès des répondantes à propos de l'âge de leurs premières menstruations et leur demande si elles sont enceintes, si elles allaitent ou si elles prennent des pilules anticonceptionnelles.</p>
10	<p>Consommation de fruits et de légumes (FVC)</p> <p><u>Univers</u> : 6 mois et plus</p>	<p>Source : Module sur la consommation de fruits et de légumes du cycle 2.1 de l'ESCC. Nota : Dans le cycle 2.1 de l'ESCC, ce module s'applique aux personnes âgées de 12 ans et plus, tandis que dans l'Enquête sur la nutrition, il s'applique aux personnes âgées de six mois et plus.</p> <p>Aperçu : Ce module comporte des questions sur la consommation habituelle de jus, de fruits et de légumes du répondant.</p>
11	<p>Problèmes de santé chroniques (CCC)</p> <p><u>Univers</u> : Tous les âges</p>	<p>Source : Module sur les problèmes de santé chroniques du cycle 2.1 de l'ESCC, auquel les modifications suivantes ont été apportées :</p> <ul style="list-style-type: none"> (1) les questions relatives à de nombreux problèmes de santé chroniques n'ayant pas de lien direct avec la nutrition n'ont pas été posées; (2) une nouvelle question (CCC_Q401) demandant aux répondants âgés de 49 ans et plus s'ils ont de « l'ostéoporose » a été ajoutée. <p>Aperçu : Ce module recueille de l'information sur les problèmes de santé et les maladies à long terme qui ont duré ou devraient durer au moins 6 mois et qui ont été diagnostiqués par un professionnel de la santé.</p>

	Nom du module/ acronyme	Source et description sommaire du module
12	Usage du tabac (SMK) <u>Univers</u> : 12 ans et plus	Source : Module sur l'usage du tabac du cycle 2.1 de l'ESCC (certaines questions qui ne sont pas nécessaires pour la production de variables calculées relatives à l'usage du tabac n'ont pas été posées). Aperçu : Ce module comporte une série de questions sur les habitudes passées et actuelles en ce qui concerne l'usage du tabac. On sait que fumer influence le métabolisme d'une personne et qu'il existe donc une forte relation entre le tabac et la nutrition.
13	Consommation d'alcool (ALC) <u>Univers</u> : 12 ans et plus	Source : Module sur la consommation d'alcool du cycle 2.1 de l'ESCC (certaines questions qui ne sont pas nécessaires pour la production de variables calculées relatives à la consommation d'alcool n'ont pas été posées). Aperçu : Ce module comporte des questions sur la fréquence de la consommation d'alcool, et la quantité d'alcool consommé au cours des 12 mois précédents. La consommation d'alcool joue un rôle important dans la consommation de calories et l'alimentation globale d'une personne.
14	Sécurité alimentaire (FSC) <u>Univers</u> : Tous les ménages	Source : Adapté de « Food-Security Core Module Questionnaire », publié par le Département de l'agriculture des États-Unis (USDA) en 2000. ¹³ Aperçu : Ce module questionne les répondants sur la situation alimentaire de leur ménage et leur demande s'ils ont pu se payer les aliments dont ils avaient besoin au cours des 12 mois précédents. Toutes les questions de ce module ont deux caractéristiques en commun : (1) Chacune vise à vérifier que la situation ou le comportement dont il est fait état se sont produits en conséquence des contraintes financières du ménage et comporte pour ce faire des phrases telles que « <i>parce que vous et les autres</i>

13. Plus d'information à propos du module sur la sécurité alimentaire se trouve dans *Guide to Measuring Household Food Security* par Gary Bickel, Mark Nord, Cristofer Price, William Hamilton et John Cook. U.S. Department of Agriculture, Food and Nutrition Service. mars 2000 (en ligne, à l'adresse : www.ers.usda.gov/briefing/foodsecurity).

	Nom du module/ acronyme	Source et description sommaire du module
		<p><i>adultes du ménage n'avez pas les moyens d'avoir assez de nourriture</i> » ou « <i>parce qu'il n'y avait pas assez d'argent pour la nourriture</i> ». (2) Chacune des questions demande si la situation ou le comportement signalé se sont produits au cours des 12 derniers mois. Veuillez remarquer que les questions de ce module sont posées à un membre du ménage bien informé.</p>
15	<p>Caractéristiques sociodémographiques (SDE)</p> <p><u>Univers</u> : Tous les âges</p>	<p>Source : Module sur les caractéristiques sociodémographiques du cycle 2.1 de l'ESCC, auquel les changements suivants ont été apportés :</p> <p>(1) dans l'Enquête sur la nutrition, pour répondre aux besoins de Santé Canada qui désire acquérir des unités d'échantillonnage additionnelles pour étudier la population autochtone vivant à l'extérieur des réserves, une nouvelle catégorie d'« Autochtones » a été ajoutée à la question SDE_Q7. (Les gens qui habitent au Canada ont des origines culturelles et raciales très variées. [Êtes/Est]-[vous/il/elle] :). De plus, une nouvelle question de suivi (SDE_Q7B) a été ajoutée pour les répondants qui se disent eux-mêmes « Autochtone » lorsqu'ils répondent à la question SDE_Q7. Cette question de suivi demande aux répondants autochtones s'ils sont Indiens de l'Amérique du Nord, Métis ou Inuit;</p> <p>(2) les questions sur l'orientation sexuelle et la participation aux services de guerre n'ont pas été posées.</p> <p>Aperçu : Ce module comporte des questions sur le statut d'immigrant, le pays de naissance, l'origine ethnique, la langue et la fréquentation scolaire ou universitaire.</p>
16	<p>Population active (LBF)</p> <p><u>Univers</u> : 15 à 75 ans</p>	<p>Source : Module sur la population active du cycle 2.1 de l'ESCC (certaines questions qui ne sont pas nécessaires pour la production de variables calculées ayant trait à population active n'ont pas été posées).</p>

	Nom du module/ acronyme	Source et description sommaire du module
		Aperçu : Le module sur la population active comporte des questions sur les activités du répondant dans la population active au cours de l'année écoulée, notamment sur ses liens avec le marché du travail, la recherche d'emploi et les heures et les semaines pendant lesquelles il a travaillé.
17	Revenus (INC) <u>Univers</u> : Tous les âges	Source : Module sur le revenu du cycle 2.1 de l'ESCC. Aperçu : Ce module comporte une série de questions sur le revenu du ménage (c.-à-d. le revenu total avant impôts et déductions de tous les membres du ménage) et le revenu personnel.
18	Administration (ADM) <u>Univers</u> : Tous les âges	Source : Module sur l'administration du cycle 2.1 de l'ESCC (certaines questions de l'Enquête sur la nutrition n'ont pas été posées). Aperçu : Dans ce dernier module, on a demandé aux répondants le numéro de leur carte santé provinciale afin de coupler les renseignements de l'enquête aux renseignements provinciaux sur la santé qui les concernent. On a également demandé aux répondants la permission de partager les renseignements communiqués avec Santé Canada et les ministères provinciaux de la santé.

4.2 Volet « rappel alimentaire de 24 heures »

Le rappel alimentaire de 24 heures est la principale composante de l'ESCC 2.2. Cette composante utilise un outil innovateur d'interview assistée par ordinateur appelé Automated Multiple Pass Method (AMPM), qui aide les répondants à se rappeler et à déclarer les aliments qu'ils ont consommés au cours de la période de 24 heures précédant l'interview, soit de minuit à minuit. L'AMPM comprend un logiciel spécialisé développé originellement par le United States Department of Agriculture (USDA) pour recueillir des données sur la consommation d'aliments. On a modifié cette application, en collaboration avec Santé Canada, pour tenir compte des différences entre les aliments à la disposition des consommateurs canadiens et des consommateurs américains, ainsi qu'entre les préparations alimentaires et les aliments ethniques consommés dans les deux pays.

La version adaptée de l'AMPM, qui a été utilisée dans le cadre de la composante du rappel alimentaire de 24 heures de l'ESCC 2.2, comporte cinq étapes. Chaque étape a été conçue pour soutenir l'intérêt des répondants et susciter leur participation à l'interview, ainsi que pour les aider

à se rappeler tous les aliments et boissons qu'ils ont consommés au cours de la période précédente de 24 heures. L'interview en cinq étapes comprend les éléments suivants.

1. **Liste rapide** : La liste rapide amène le répondant à faire état rapidement des aliments, y compris les collations et l'eau, dont il se souvient facilement. À cette étape, le répondant donne la liste de tous les aliments et de toutes les boissons qu'il a consommés au cours d'une période de 24 heures, le jour ayant précédé l'interview, selon l'ordre qu'il préfère (sans être interrompu par l'intervieweur). C'est donc dire que le répondant n'a pas à déclarer les aliments en ordre chronologique, ce qui libère son esprit pour se concentrer sur ce qu'il a consommé.
2. **Aliments oubliés** : Cette étape est conçue pour recueillir des données sur les aliments qui auraient pu être oubliés à l'étape de la « liste rapide ». On pose au répondant une série de questions visant à découvrir les aliments oubliés appartenant à neuf catégories : boissons non alcoolisées, boissons alcoolisées, sucreries, collations non sucrées, fruits, légumes, fromages, pain et petits pains et tout autre type d'aliments.
3. **Heure et occasion** : Le répondant précise à quelle heure il a commencé à manger ou à boire les aliments et les boissons qu'il a mentionnés et le nom qu'il donne (p. ex., déjeuner, dîner) – selon sa propre définition – à cette occasion de consommation. Ces questions sont conçues pour aider le répondant à se souvenir et à faire état de ses habitudes alimentaires au cours de la dernière période de 24 heures, ainsi qu'à regrouper les aliments qui ont été consommés au même moment.
4. **Cycle détaillé** : Cette étape est réservée aux questions plus difficiles et détaillées et est menée soigneusement par l'intervieweur, une fois que le contact est bien établi avec le répondant. On pose au répondant des questions normalisées en vue d'obtenir des renseignements détaillés au sujet de chaque aliment et boisson déclaré aux étapes antérieures. Les détails recueillis portent entre autres sur la description des aliments, les quantités d'aliments, les ajouts aux aliments et les méthodes de préparation. Un livret intitulé « *Livret de modèles de portions* » sert de guide principal pour la mesure des aliments; il aide le répondant à décrire la portion ou la quantité d'aliments et de boissons consommés. Le livret comprend des images de différentes tailles de verres, de tasses, de bols, de montagnes/carrés/tartinades et de cercles. Il comprend en outre une grille, deux pointes et une page de formes et de morceaux de poulet. Un examen des occasions de consommation et des périodes entre ces occasions fait également partie de cette étape, afin de favoriser le rappel des aliments oubliés. Une fois que les données sur l'occasion de consommation sont complètes, on demande au répondant où le repas a été préparé.
5. **Révision finale** : Une dernière question supplémentaire est posée pour recueillir des données sur les autres aliments qui auraient pu être consommés, mais oubliés ou non mentionnés précédemment dans l'interview.

5. Plan d'échantillonnage

5.1 Population cible

Le cycle 2.2 de l'ESCC vise la population de tous les âges vivant à domicile dans les dix provinces. Sont exclus du champ de l'enquête les résidents des trois territoires, des réserves indiennes et des terres de la Couronne, les résidents des établissements, les membres à temps plein des Forces canadiennes et les habitants de certaines régions éloignées. Le cycle 2.2 de l'ESCC couvre environ 98 % de la population des dix provinces.

5.2 Taille et répartition de l'échantillon

Compte tenu du budget accordé à cette enquête et pour répondre aux objectifs de produire des estimations des distributions des apports alimentaires courants pour des domaines d'intérêt particuliers pour chaque province et compte tenu du budget de l'enquête, il a été établi que cette composante devait être réalisée auprès d'un échantillon de 29 000 personnes. Une stratégie en deux étapes a été utilisée pour répartir l'échantillon entre les provinces. Premièrement, et afin d'estimer les distributions des apports alimentaires courants, 80 unités d'échantillonnage ont été attribuées à chaque domaine d'intérêt (14 groupes âge-sexe) dans chaque province (il convient de signaler que les distributions des apports alimentaires courants pour le groupe d'âge <1 étaient requises à l'échelle nationale seulement). Ainsi, 1 120 unités ont été attribuées à chaque province dans un premier temps pour un total de 11 200 unités. Les autres 17 800 unités ont été réparties entre les provinces selon un schéma de répartition de puissance où $q=0.7$.¹⁴ Une seule exception : on n'a pas attribué d'unités d'échantillonnage à la province de l'Île-du-Prince-Édouard à la deuxième étape. La taille de l'échantillon total pour une province donnée s'obtient en faisant la somme des tailles obtenues aux deux étapes. Le tableau 5.1 donne les détails de la répartition provinciale pour le cycle 2.2 de l'ESCC.

Tableau 5.1 Tailles visées d'échantillon selon la province

Province	1 ^{re} étape 80/domaine	2 ^e étape Puiss. $q=0.7$	Échantillon total
Terre-Neuve-et-Labrador	1 120	542	1 662
Île-du-Prince-Édouard	1 120	0	1 120
Nouvelle-Écosse	1 120	837	1 957
Nouveau-Brunswick	1 120	713	1 833
Québec	1 120	3 744	4 864
Ontario	1 120	5 620	6 740
Manitoba	1 120	1 050	2 170
Saskatchewan	1 120	856	1 976
Alberta	1 120	1 996	3 116
Colombie-Britannique	1 120	2 442	3 562
Canada	11 200	17 800	29 000

De plus, dans le but d'avoir une bonne représentation urbaine et rurale dans chaque province, l'échantillon provincial a ensuite été réparti selon le type de strate : urbain ou rural. Ainsi, l'échantillon provincial a été réparti selon la strate urbaine et la strate rurale proportionnellement

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

au nombre de logements estimés dans chacune des deux strates. Puis, l'effectif des échantillons a été augmenté avant la collecte des données afin de tenir compte des logements hors du champ de l'enquête ou vacants, et du taux prévu de non-réponse.

5.3 Achats provinciaux d'unités d'échantillon

Avant le début de la collecte, les provinces du Manitoba, de l'Ontario et de l'Île-du-Prince-Édouard ont fourni des fonds supplémentaires afin que l'on puisse sélectionner un échantillon plus important de logements. Ces achats d'unités d'échantillon supplémentaires avaient pour objet d'obtenir un échantillon de taille suffisante pour produire des estimations fiables pour des régions infraprovinciales pour plusieurs domaines d'intérêt pour les provinces du Manitoba et de l'Ontario. Pour l'Île-du-Prince-Édouard, l'achat avait pour but d'augmenter les tailles visées de répondants pour 11 domaines d'intérêt clés (groupes âge-sexe : 4-8 tous, 9-13*H/F, 14-18*H/F, 19-30*H/F, 31-50*H/F et 51-70*H/F). Le Manitoba a ajouté 1 500 unités, l'Ontario a ajouté 4 360 unités tandis que l'Île-du-Prince-Édouard a ajouté 300 unités.

La province du Manitoba a été divisée en 4 régions. La répartition des 3 670 unités a été effectuée de façon à garantir un minimum de 80 unités d'échantillon pour 6 domaines d'intérêt clés (groupes âge-sexe : 1-3 tous, 4-8 tous, 9-13*H/F et 14-18*H/F) pour chaque région. (Deux de ces régions ont été combinées sur le fichier de microdonnées à grande diffusion à cause de leur petite taille de population. Voir la page 3 de l'Annexe E.) La province de l'Ontario a été divisée en 7 régions et la répartition des 11 100 unités a été effectuée de façon à garantir un minimum de 80 unités d'échantillon pour 10 domaines d'intérêt clés (groupes âge-sexe : 1-3 tous, 4-8 tous, 9-13*H/F, 14-18*H/F, 19-70*H/F et 71+*H/F) pour chaque région. Le tableau 5.2 présente la répartition finale de l'échantillon pour les provinces du Manitoba et de l'Ontario suite à l'achat.

Tableau 5.2 Répartition par région du Manitoba et de l'Ontario incluant les achats

Région infra-provincial	Taille d'échantillon
Manitoba	3 670
Burntwood-Norman-Churchill	665
Assiniboine-Parkham-Brandon	825
N.Eastman-S.Eastman-Interlake-Central	955
Winnipeg	1 225
Ontario	11 100
Sud-Ouest	1 600
Centre-Sud	1 500
Centre-Ouest	1 700
Centre-Est	1 600
Toronto	1 600
Est	1 600
Nord	1 500

5.4 Échantillonnage des ménages à partir des différentes bases de sondage

Plusieurs bases de sondage ont été utilisées pour sélectionner l'échantillon de ménages du cycle 2.2 de l'ESCC pour lesquelles les stratégies de sélection différaient. Ces bases de sondage sont :

- une base aréolaire,
- une base liste d'adresses de logement provenant de l'échantillon du cycle 2.1 de l'ESCC, et
- une base liste d'adresses de logement provenant des registres de soins de santé provinciaux du Manitoba et de l'Île-du-Prince-Édouard.

5.4.1 Échantillonnage des ménages à partir de la base aréolaire

L'échantillon du cycle 2.2 de l'ESCC provenait principalement de la base aréolaire conçue pour l'Enquête sur la population active (EPA) du Canada. Le plan d'échantillonnage de l'EPA est un plan d'échantillonnage en grappes stratifié à plusieurs degrés où le logement représente l'unité finale d'échantillonnage.¹⁵ À la première étape, on a formé des strates homogènes et sélectionné un échantillon indépendant de grappes, dans chaque strate. À la deuxième étape, on a dressé la liste des logements pour chaque grappe, puis on a sélectionné les logements, ou les ménages, d'après les listes.

Pour les besoins du plan d'échantillonnage, chaque province est répartie en trois catégories de région, à savoir: les grands centres urbains, les villes et les régions rurales. Des strates géographiques ou socioéconomiques sont formées à l'intérieur de chaque grand centre urbain. Dans les strates, des grappes sont formées par regroupement de 150 à 250 logements. Dans certains centres urbains, des strates distinctes sont créées pour les immeubles à appartements ou les secteurs de dénombrement (SD) pour lesquels le revenu moyen du ménage est élevé. Dans chaque strate, on sélectionne six grappes ou immeubles résidentiels (pouvant compter de 12 à 18 appartements) par une méthode d'échantillonnage aléatoire avec probabilité proportionnelle à la taille (PPT), cette dernière correspondant au nombre de ménages. Le nombre 6 est utilisé pour l'ensemble du plan d'échantillonnage afin de permettre le renouvellement mensuel d'un sixième de l'échantillon de l'EPA.

Les autres villes et régions rurales de chaque province sont stratifiées, en premier lieu, en fonction de données géographiques, puis selon des caractéristiques socioéconomiques. Dans la plupart des strates, on sélectionne six grappes (habituellement des SD) selon la méthode PPT. Pour les strates où la densité de population est faible, on suit un plan en trois étapes en vertu duquel on sélectionne deux ou trois unités primaires d'échantillonnage (UPE), qui correspondent normalement à des groupes de SD, puis on les répartit en grappes dont six sont sélectionnées pour faire partie de l'échantillon. La sélection est faite à chaque étape à l'aide de la méthode PPT.

Après avoir établi la liste des nouvelles grappes, on obtient l'échantillon par échantillonnage systématique des logements. Le tableau 5.3 donne un aperçu des catégories d'UPE utilisées pour l'échantillon complet de l'EPA. Le *rendement* correspond au nombre de ménages sélectionnés dans le cadre de l'EPA pour un mois donné. Comme les taux d'échantillonnage sont prédéterminés, il existe souvent un écart entre la taille prévue d'échantillon et les chiffres obtenus. Ainsi, le rendement de l'échantillon est parfois excessif. Cette situation se présente surtout pour les secteurs où le nombre de logements a augmenté à la suite de nouveaux projets de

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

construction, par exemple. Pour réduire le coût de la collecte des données, on corrige la production excessive par élimination, dès le départ, d'une partie des unités sélectionnées et en modifiant la pondération appliquée dans le plan d'échantillonnage. Cette opération, habituellement réalisée au niveau agrégé, porte le nom de *stabilisation de l'échantillon*. En plus, on augmente la taille requise de l'échantillon de ménages pour tenir compte du fait qu'en général, environ 12 % de logements ne sont pas occupés par des ménages faisant partie du champ d'observation (certains logements sont vacants ou occupés de façon saisonnière, et d'autres sont occupés par des ménages non visés par l'enquête).

Tableau 5.3 Unité primaire d'échantillonnage, taille et rendement

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

Afin de répondre aux exigences particulières à l'ESCC (cycle 2.2), certaines modifications ont dû être apportées à cette stratégie d'échantillonnage. Pour obtenir un échantillon de base de 23 400 répondants pour l'ESCC (cycle 2.2), il faut sélectionner 31 000 logements de la base aréolaire (pour tenir compte des logements vacants et des ménages non-répondants). Chaque mois, le plan d'échantillonnage de l'EPA fournit environ 68 000 logements répartis entre les diverses régions économiques des dix provinces, alors que, pour l'ESCC (cycle 2.2), il faut obtenir un échantillon total de 31 000 logements répartis entre les strates urbaines et rurales des provinces ou régions (au Manitoba et en Ontario), dont les limites géographiques diffèrent de celles des régions économiques de l'EPA. Globalement, l'ESCC (cycle 2.2) nécessite la sélection d'un nombre inférieur de logements que produit le mécanisme de sélection de l'EPA, ce qui correspond à un *facteur de redressement* de 0,45 (31 000/68 000). Toutefois, comme ce facteur de redressement varie de 0,25 à 3,0 au niveau urbain/rural dans les provinces ou les régions, certains ajustements sont nécessaires.

Les modifications apportées au processus de sélection dans les régions varient selon le facteur de redressement. Pour les régions où le facteur est égal ou inférieur à 1, le nombre d'UPE choisi est réduit si nécessaire. Par exemple, si le facteur est de 0,5, alors seulement 3 UPE sont choisies dans chaque strate de l'EPA au lieu du nombre habituel de 6 UPE. Pour les régions avec un facteur supérieur à 1, on répète le processus d'échantillonnage des logements à l'intérieur d'une UPE pour un sous-ensemble des UPE sélectionnées appartenant à la région en question. Par exemple, si le facteur est de 1,6, alors la sélection des logements est répétée dans 4 des 6 UPE pour chaque strate de la région. Lorsque la répétition de la sélection de logements à l'intérieur d'une UPE est nécessaire mais qu'aucun autre logement n'est disponible dans cette UPE, alors une autre UPE est choisie. Finalement, dans les cas où la méthode choisie produit un excédent inutile de logements, on procède à la stabilisation de l'échantillon.

5.4.2 Échantillonnage des ménages de la base liste créée à partir de l'ESCC (cycle 2.1)

Afin d'assurer une taille d'échantillon minimum dans chaque groupe âge-sexe, particulièrement pour les groupes d'âge plus jeunes, une base de sondage secondaire a été utilisée pour compléter la base aréolaire dans toutes les provinces sauf le Manitoba et l'Île-du-Prince-Édouard. La base secondaire était nécessaire parce qu'il était difficile de trouver un nombre suffisant de ménages comprenant des jeunes personnes dans la base aréolaire puisque les renseignements à l'échelle du ménage n'étaient pas disponibles avant la collecte des données. On a créé cette base secondaire en utilisant les renseignements sur les ménages fournis par les répondants à la composante régionale de l'ESCC (cycle 2.1) dont la période de collecte des données sur 12 mois a commencé en janvier 2003. Les ménages comprenant au moins une personne âgée de 18 ans ou moins au moment de la collecte des données du cycle 2.1 de l'ESCC ont fait partie d'une liste d'adresses de logement. On a ensuite utilisé cette liste d'adresses de logement pour compléter l'échantillon de ménages faisant partie de la base aréolaire afin d'accroître la probabilité de trouver des ménages comprenant des personnes âgées de 18 ans ou moins. Il est d'intérêt de mentionner que la majorité des adresses de logement provenait de la partie « base aréolaire » de l'échantillon de l'ESCC (cycle 2.1); seule une petite proportion d'adresses provenait de la partie « base téléphonique ». La liste a par la suite été stratifiée par province et par région urbaine/rurale (l'Ontario a été stratifiée par région et non par un indicateur urbain/rural) et un échantillon de municipalités et (ou) de villes a été sélectionné à la première étape à l'aide d'une méthode d'échantillonnage aléatoire avec probabilité proportionnelle à la taille. On a ensuite sélectionné un échantillon systématique d'adresses de logement à la deuxième étape. Pour obtenir un échantillon de base de 11 000 ménages répondants, on a sélectionné environ 16 500 logements dans la liste en vue de compenser pour les logements vacants et la non-réponse.

5.4.3 Échantillonnage des ménages de la base liste d'adresses de logement créée à partir des registres de soins de santé provinciaux

Étant donné les achats d'échantillon des provinces du Manitoba et de l'Île du Prince Édouard et afin d'assurer une taille d'échantillon minimum dans chaque groupe âge-sexe, particulièrement pour les groupes d'âge plus jeunes, une base de sondage secondaire différente a été utilisée pour compléter la base aréolaire au Manitoba et à l'Île-du-Prince-Édouard. Cette base secondaire était nécessaire parce qu'il n'y avait pas suffisamment de ménages comprenant des jeunes personnes dans la partie « base aréolaire » de l'échantillon du cycle 2.1 de l'ESCC pour atteindre l'objectif des achats d'échantillon. On a créé cette base secondaire en utilisant les renseignements sur les ménages des détenteurs de cartes d'assurance-maladie tels que transmis à Statistique Canada par les Ministères de la santé de ces deux provinces. Ces deux ministères ont fourni une liste d'adresses de logement ainsi que la composition des ménages y résidant en termes de nombres d'individus par groupe d'âge clé. Les ménages comprenant au moins une personne âgée de 18 ans ou moins ont formé la base liste d'adresses de logement. On a ensuite utilisé cette liste d'adresses de logement pour compléter l'échantillon de ménages faisant partie de la base aréolaire afin d'accroître la probabilité de trouver des ménages comprenant des personnes âgées de 18 ans ou moins. La liste a par la suite été stratifiée par région pour le Manitoba et par indicateur urbain/rural pour l'Île-du-Prince-Édouard, et un échantillon de municipalités et (ou) de villes a été sélectionné à la première étape à l'aide d'une méthode d'échantillonnage aléatoire avec probabilité proportionnelle à la taille. On a ensuite sélectionné un échantillon systématique d'adresses de logement à la deuxième étape. Pour obtenir un échantillon de base de 1 900 ménages déclarants, on a sélectionné environ 3 000 logements dans la liste en vue de compenser pour les logements vacants et la non-réponse.

5.5 Échantillonnage des répondants

La stratégie de sélection des répondants a été conçue de façon à atteindre les nombres visés de répondants pour chaque domaine d'intérêt pour chaque province et (ou) région.¹⁶ La stratégie d'échantillonnage adoptée tient compte des besoins des utilisateurs de données, du coût, de l'efficacité du plan d'échantillonnage, du fardeau de réponse et des contraintes opérationnelles. Pour le cycle 2.2 de l'ESCC, il a été décidé de sélectionner une personne par ménage en utilisant diverses probabilités de sélection variant selon l'âge et selon la base de sondage. On a simulé plusieurs scénarios reposant sur divers paramètres dans le but de déterminer l'approche optimale qui garantissait le nombre minimum de personnes dans chaque domaine d'intérêt dans chaque province et (ou) région sans générer de poids d'échantillonnage extrêmes en bout de ligne. Le tableau 5.4 donne les facteurs multiplicatifs de poids de sélection utilisés pour déterminer les probabilités de sélection des personnes dans les ménages échantillonnés selon l'âge et selon la base de sondage. À titre d'exemple, dans un ménage de trois personnes constitué de deux adultes de 31 ans ou plus et d'un jeune de 15 ans, et sélectionné à partir de la base aréolaire, l'adolescent avait 3 fois plus de chances d'être sélectionné.

La stratégie d'échantillonnage à l'échelle de la personne décrite ci-dessus conjuguée à la stratégie d'échantillonnage à l'échelle des ménages à partir de deux bases a permis d'atteindre le nombre minimum de 80 unités déclarantes par domaine d'intérêt pour presque toutes les provinces et (ou) régions.

Tableau 5.4 Probabilités relatives de sélection pour la stratégie d'échantillonnage à l'échelle de la personne selon l'âge et selon la base de sondage

Base	Facteur multiplicatif de poids de sélection						
	< 1	1-3	4-8	9-13	14-18	19-30	31+
Aréolaire	1	3	3	3	3	3	1
Liste	1	2	1	1	1	0	0

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

Afin d'équilibrer la charge de travail des intervieweurs et d'éliminer les possibles effets saisonniers pouvant être observés chez certaines caractéristiques, l'échantillon initial de logements a été réparti au hasard, dans chaque région, sur les 11 mois de collecte des données (le douzième mois de collecte est habituellement utilisé pour tenter, une dernière fois, d'encourager les unités non-répondantes à participer à l'enquête). Pour des raisons opérationnelles, tous les logements d'une même UPE ont été assignés à la même période de collecte (T1 : janvier à mars 2004, T2 : avril et mai 2004, T3 : juin à août 2004 et T4 : septembre à novembre 2004). Il est également important de mentionner que la collecte des données s'est poursuivie jusqu'à la fin janvier 2005 afin d'améliorer les taux de réponse.

5.7 Achat spécial d'unités d'échantillon de la part de Santé Canada

En parallèle aux trois achats provinciaux d'unités d'échantillon tels que décrits précédemment, Santé Canada a fourni des fonds supplémentaires afin de sélectionner un échantillon plus important d'autochtones vivant à l'extérieur des réserves pour que l'on puisse estimer les

16. Pierre, F. et Béland, Y. (2004). Sélection des personnes dans l'Enquête Canadienne sur la Nutrition. *Recueil des présentations de la section des méthodes d'enquêtes 2004*. Hamilton : Société Statistique du Canada.

distributions des apports alimentaires courants à l'échelle nationale pour 4 groupes âge-sexe (19-30*H/F et 31-50*H/F). Pour atteindre le nombre minimum recommandé de 80 répondants par domaine d'intérêt, un échantillon supplémentaire de 980 unités d'échantillon a été réparti parmi les 10 provinces et ajouté à l'échantillon de base de l'ESCC (cycle 2.2). Ces 980 unités d'échantillon supplémentaires ont été sélectionnées à partir d'une autre base liste d'adresses de logement créée à l'aide des renseignements sur les ménages fournis par les répondants à la composante régionale de l'ESCC (cycle 2.1) dont la collecte s'est tenue de janvier à décembre 2003. Les ménages pour lesquels la personne sélectionnée avait déclaré être autochtone lors de la collecte de l'ESCC (cycle 2.1) ont fait partie d'une liste d'adresses de logement (cette liste contenait en fait 980 adresses de logements). On a ensuite utilisé cette liste d'adresses de logement afin d'accroître la probabilité de trouver des ménages comprenant des autochtones vivant à l'extérieur des réserves. La stratégie de sélection de la personne parmi les ménages échantillonnés est celle décrite à la section 5.5.

Le lecteur devrait toutefois prendre note que la variable « statut d'autochtone » n'est pas présente sur ce fichier de microdonnées à grande diffusion pour des raisons de confidentialité. Cette variable ne peut être accédée que par le biais du fichier maître localisé dans les bureaux de Statistique Canada et dans les centres de données de recherche ou par le biais du fichier partagé localisé dans les Ministères provinciaux de la santé, à Santé Canada et à l'Agence de santé publique du Canada.

5.8 Évaluation de la variabilité intra-individuelle

Dans le but de déterminer la variabilité intra-individuelle des données sur les apports alimentaires courants d'une personne qui se produit naturellement étant donné la variété d'aliments consommés quotidiennement dans les pays industrialisés, une deuxième mesure du rappel alimentaire de 24 heures a été recueillie auprès d'un sous-échantillon de 10 150 personnes. Cette deuxième mesure permet de rajuster les distributions des apports alimentaires courants. Ce sous-échantillon a été réparti parmi les 10 provinces de façon à garantir un nombre minimum de deuxièmes rappels dans chacun des 14 domaines d'intérêt clés pour permettre les ajustements. Ces personnes ont été re-contactées 3 à 10 jours après la première entrevue pour une entrevue téléphonique d'environ 30 minutes (seul le rappel alimentaire de 24 heures était administré lors de cette entrevue).

Nota : La variabilité intra-individuelle augmente la variance de la distribution des apports alimentaires courants des différents éléments nutritifs,¹⁷ ce qui nuit à l'estimation de la population « à risque » (queues des distributions). Cela tient à ce qu'elle comprend à la fois la variabilité intra-individuelle (d'un jour à l'autre) et la variabilité entre individus, de sorte que les estimations des apports alimentaires des populations à risque peuvent être inadéquats ou excessifs et sont probablement supérieures à la prévalence véritable. Comme il importe pour cette enquête d'obtenir des estimations exactes de la prévalence des populations à risque en ce qui concerne plusieurs éléments nutritifs pour fins d'études plus approfondies, il faut corriger les distributions des apports alimentaires courants observés de manière à ce qu'elles reflètent plus exactement seulement la variabilité des apports alimentaires entre une personne et une autre.

17. Nusser, S.M., Carriquiry, A.L., Dodd, K.W. et Fuller, W.A. (1996). A Semiparametric Transformation Approach to Estimating Usual Daily Intake Distributions. *Journal of American Statistical Association*. 91: 1440-1449.

6. Collecte des données

Les données du cycle 2.2 de l'ESCC ont été recueillies du 14 janvier 2004 au 21 janvier 2005. La collecte des données a été divisée en quatre trimestres :

1 ^{er} trimestre	du 1 ^{er} janvier au 31 mars 2004
2 ^e trimestre	du 1 ^{er} avril au 31 mai
3 ^e trimestre	du 1 ^{er} juin au 31 août
4 ^e trimestre	du 1 ^{er} septembre au 21 janvier 2005

6.1 Conception du questionnaire

Les questions du cycle 2.2 ont été posées par la méthode d'interview assistée par ordinateur (IAO), qui offre un certain nombre d'avantages en ce qui concerne la qualité des données par rapport à d'autres méthodes de collecte. Premièrement, le texte des questions, y compris les périodes de référence et les pronoms, est personnalisé automatiquement d'après des facteurs comme l'âge et le sexe du répondant, la date de l'interview et les réponses aux questions précédentes. Deuxièmement, des règles de vérification sont appliquées automatiquement pour repérer les réponses incohérentes ou non incluses dans la tranche de valeurs permises, et des messages-guides apparaissent à l'écran quand une entrée invalide est enregistrée. De cette façon, l'intervieweur reçoit une rétroaction immédiate et peut corriger toute incohérence. Troisièmement, l'application saute automatiquement les questions qui ne sont pas applicables à la personne interviewée.

6.2 Essai sur le terrain

Le questionnaire provisoire a été mis à l'essai sur le terrain du 1^{er} juin au 7 juillet 2003. L'essai a été réalisé dans les provinces Atlantiques, au Québec et en Colombie-Britannique auprès d'un échantillon de 700 unités. Les objectifs principaux de l'essai sur le terrain étaient les suivants :

- évaluer la durée globale de l'enquête et des modules individuels du questionnaire;
- évaluer les réactions des répondants à toutes les questions de l'enquête;
- déterminer si les répondants étaient prêts à participer par téléphone à un deuxième rappel alimentaire de 24 heures;
- évaluer les protocoles et les réponses en ce qui a trait à la collecte de mesures physiques du poids et de la taille;
- évaluer l'efficacité des procédures opérationnelles sur le terrain, de la formation des intervieweurs et de l'application d'interview assistée par ordinateur.

6.3 Personnel de collecte (formation, supervision et contrôle)

Les intervieweurs chargés de la collecte des données du cycle 2.2 de l'ESCC ont reçu une formation de 3,5 jours visant à assurer qu'ils comprenaient bien les concepts et les procédures de l'enquête. La formation a été donnée par le gestionnaire de la collecte des données de l'enquête ou par un intervieweur principal. Elle comprenait la révision du Manuel de l'intervieweur, l'exécution d'interviews simulées, la réponse à des questions test posées par les intervieweurs principaux et la discussion de toute question concernant la collecte des données.

Durant la collecte des données, les intervieweurs étaient sous la supervision d'intervieweurs principaux auxquels il incombait de s'assurer que les intervieweurs maîtrisaient bien les concepts

et les procédures de l'enquête. Les intervieweurs principaux ont veillé à ce que des mesures de suivi soient prises promptement en cas de refus de participation ou d'autres formes de non-réponse. Au besoin, les cas de non-réponse ont été transférés à un intervieweur principal et réaffectés. À leur tour, les intervieweurs principaux relevaient des questionnaires de la collecte établis dans les bureaux régionaux de Statistique Canada.

Les intervieweurs ont transmis quotidiennement les interviews achevées, directement de leur résidence au bureau central de Statistique Canada au moyen d'une ligne téléphonique sécurisée. Au bureau central, on a vérifié l'exactitude des interviews achevées et communiqué systématiquement toute incohérence aux intervieweurs.

6.4 Méthode de collecte des données

La collecte des données a débuté en janvier 2004 et s'est étendue sur 12 mois, ce qui a permis d'équilibrer la charge de travail des intervieweurs et de réduire au minimum les effets saisonniers sur certaines caractéristiques liées à la santé, comme l'activité physique et les différences saisonnières de disponibilité de certains aliments.

À l'étape du premier contact, les intervieweurs se sont rendus sur le lieu de chaque logement sélectionné. La première partie de l'interview comprenait l'énumération de tous les membres du ménage, y compris la collecte des renseignements sur les liens qui les unissaient (c.-à-d. mère, père, frère, soeur, etc.), ainsi que des caractéristiques démographiques de base, comme l'âge, le sexe, la date de naissance, l'état matrimonial et des renseignements sur le logement. Après avoir recueilli l'information de base sur le ménage, l'intervieweur a sélectionné au hasard un membre du ménage pour participer à l'enquête. Dans chaque ménage, un seul membre a été choisi.

Après avoir sélectionné un répondant dans le ménage, les intervieweurs avaient pour instructions de lui demander de répondre directement au questionnaire. La première interview avec le répondant sélectionné comportait deux volets : le « rappel alimentaire de 24 heures » et le « questionnaire général sur la santé ». Les intervieweurs avaient pour instructions de réaliser la première interview sur place au domicile du répondant.

À environ 30 % des répondants sélectionnés, on a également demandé de participer par téléphone à un deuxième rappel alimentaire de 24 heures, de 3 à 10 jours après la première interview. L'interview sur place pour ce deuxième rappel a été acceptée dans de rares circonstances où le répondant n'avait pas le téléphone ou préférait que l'interview ait lieu sur place. Au Québec et en Colombie-Britannique, un sous-ensemble de deuxièmes interviews a été réalisé intentionnellement sur place pour permettre d'analyser le biais éventuel dû au mode de collecte lors de la deuxième interview.

La durée globale de la première interview, y compris le volet du rappel alimentaire de 24 heures, était de 60 minutes en moyenne. Celle de la deuxième interview de rappel était, en moyenne, de 30 minutes.

Il convient de souligner que, si le répondant sélectionné n'était pas disponible au moment de la visite, l'intervieweur avait l'instruction de retourner à une date ultérieure. Tous les efforts raisonnables ont été faits pour réaliser sur place la première interview de rappel alimentaire de 24 heures. L'interview par téléphone n'a été autorisée que si les frais de déplacement étaient excessifs ou que le répondant refusait catégoriquement de procéder à l'interview sur place. La variable ADMD_N09 indique si l'interview a été réalisée sur place, par téléphone ou par une

combinaison des deux techniques. En dernière analyse, 7 % de répondants ont participé par téléphone à la première interview de rappel alimentaire de 24 heures.

6.5 Interviews par procuration

Le questionnaire était conçu pour être posé aux répondants sélectionnés. Cependant, une interview par procuration a été nécessaire quand le répondant avait 11 ans ou moins. Pour les enfants de moins de 6 ans, le parent était la seule personne qui pouvait fournir l'information. On a estimé que, pour les enfants de ce groupe d'âge, les parents avaient beaucoup plus de contrôle que toute autre personne sur ce que leurs enfants mangeaient. Dans les cas où les parents ne pouvaient fournir des précisions, par exemple, pour les repas consommés à la garderie, on leur a demandé de communiquer avec la personne responsable afin qu'elle leur fournisse autant de détails que possible. Pour les enfants de 6 à 11 ans, les interviews ont été réalisées avec l'aide du parent (voir le tableau qui suit pour un résumé du type d'interview selon l'âge du répondant). La variable ADMD_PRX indique si l'interview a été réalisée par procuration ou non.

Tableau 6.1 Type d'interview par procuration

Âge du répondant sélectionné	Type d'interview
De 0 à 5 ans	<p>Par procuration</p> <ul style="list-style-type: none"> • On a demandé au parent ou au gardien de fournir les renseignements nutritionnels concernant l'enfant. • L'enfant n'avait pas à être présent durant l'interview. • Si l'enfant était présent, l'intervieweur a demandé au parent ou au gardien de l'aider à mesurer la taille et le poids de l'enfant.
De 6 à 11 ans	<p>Par procuration avec l'aide du parent</p> <ul style="list-style-type: none"> • On a demandé à l'enfant de fournir ses propres renseignements avec l'aide de l'un de ses parents ou de son gardien ou sa gardienne. • L'enfant n'avait pas à être présent durant toute l'interview. Cependant, il devait être présent pour le volet du rappel alimentaire de 24 heures. • L'intervieweur a mesuré la taille et le poids de l'enfant, après avoir obtenu la permission du parent.
12 ans et plus	<p>Pas de procuration</p> <ul style="list-style-type: none"> • On a demandé aux répondants de fournir eux-mêmes les renseignements les concernant. • Pour les répondants de 12 à 17 ans, le parent ou le gardien a répondu aux modules sur le revenu, la sécurité alimentaire et l'autorisation de partager et de coupler les données.

6.6 Minimiser la non-réponse

Avant la première prise de contact par l'intervieweur, une lettre d'introduction et une brochure ont été envoyées par la poste à chaque logement sélectionné pour lequel une adresse postale valide existait. Celles-ci expliquaient aux répondants l'importance de l'enquête et leur fournissait des exemples de la façon dont les données seraient utilisées.

Les intervieweurs ont reçu l'instruction de faire tous les efforts raisonnables pour réaliser les interviews. Si le moment de la visite de l'intervieweur était mal choisi, ce dernier a pris rendez-vous pour une autre visite à un moment plus opportun. En cas d'absence au moment de la visite, de nombreuses autres tentatives ont été faites pour reprendre contact avec le ménage. Aux personnes faisant partie de l'échantillon qui ont refusé au départ de participer à l'enquête, le bureau régional de Statistique Canada les plus près a envoyé une lettre pour souligner l'importance de l'enquête et de la coopération du ménage. Ensuite, un intervieweur principal, ou un autre intervieweur a fait un deuxième appel (ou une deuxième visite) pour essayer de convaincre le répondant sélectionné de l'importance de sa participation à l'enquête.

Durant les derniers mois de la collecte des données, les ménages pour lesquels aucune réponse n'avait été obtenue et les personnes sélectionnées qui avaient refusé antérieurement de répondre ont de nouveau été approchés et encouragés à participer à l'enquête. Ces contacts assidus ont produit des résultats d'enquête plus robustes grâce à la maximisation du taux de réponse.

Pour limiter les problèmes de langue susceptibles de nuire à la réalisation des interviews, tous les bureaux régionaux de Statistique Canada ont recruté des intervieweurs parlant plusieurs langues. Dans les cas où l'intervieweur ne pouvait parler la même langue que la personne sélectionnée pour participer à l'enquête, un membre du ménage (si cela était possible) a reçu l'autorisation de traduire les questions de l'enquête et les réponses.

Un système de surveillance a été mis en place pour assurer le maintien de la qualité des données durant la collecte. Divers aspects du processus d'interview ont été surveillés au niveau de l'intervieweur, comme la durée moyenne de l'interview et les non-réponses partielles. Le bureau central a transmis régulièrement des commentaires en retour aux bureaux régionaux pour les aider à corriger les problèmes à mesure qu'ils survenaient. Un processus de validation a également été mis en place sur le terrain pour surveiller la qualité du travail des intervieweurs.

À la fin de la collecte des données, le taux national de réponse était de 76,5 %. Des renseignements détaillés sur les taux de réponse figurent à la section intitulée « Qualité des données ».

6.7 Circonstances spéciales durant la collecte des données

6.7.1 Mesure de la taille et du poids

Durant les cycles précédents de l'ESCC, les renseignements sur la taille et le poids du répondant ont été fournis par ce dernier. Or, plusieurs études ont montré que, par cette méthode, les gens ont tendance à ne pas fournir des renseignements exacts. Pour la première fois, durant le cycle 2.2, après avoir obtenu la permission de le faire, les intervieweurs ont mesuré la taille et le poids du répondant selon un protocole normalisé au moyen d'une balance et d'un mètre ruban.

Les intervieweurs ont reçu la formation nécessaire pour expliquer toutes les procédures au répondant avant de commencer à prendre les mesures. Aux répondants de 12 ans et plus qui ont

refusé de se laisser mesurer par l'intervieweur, on a demandé de déclarer eux-mêmes leur poids et leur taille. Quand le répondant n'était pas physiquement apte à participer (par exemple parce qu'il était dans un fauteuil roulant ou qu'il avait besoin d'une aide de locomotion pour se tenir debout), on a recueilli les mesures qu'il a lui-même déclarées. En ce qui concerne les enfants de 2 à 11 ans qui ont refusé, ou dont le parent a refusé, que l'intervieweur prenne leurs mesures, on a accepté que le parent déclare la taille et le poids de l'enfant. Il convient de souligner que l'information sur la taille et le poids n'a pas été recueillie pour les enfants de moins de 2 ans.

Plusieurs procédures ont été mises en place pour assurer l'exactitude et la cohérence des mesures de la taille et du poids. Premièrement, on a utilisé des balances de grande qualité, ne nécessitant aucun étalonnage, faciles à utiliser et incroyablement précises étant donné qu'elles sont portables (50 grammes). La procédure de mesure de la taille était un peu plus compliquée. Comme les intervieweurs n'étaient pas des professionnels de la santé dûment formés, la procédure devait être non technique et non invasive. Elle a été mise au point par des experts du domaine afin d'atteindre cet objectif. Une bande vidéo de formation a été produite pour assurer l'uniformité de la procédure suivie par les intervieweurs dans toutes les régions du pays. En outre, pour confirmer la cohérence des mesures faites par les intervieweurs, un test durant lequel les intervieweurs ont dû mesurer le même sujet a été réalisé après chaque séance de formation. Toute divergence entre les résultats obtenus par les divers intervieweurs a été examinée à ce moment-là et une nouvelle formation a été donnée au besoin. Toutes les mesures de la taille et du poids ont également été surveillées durant la période de collecte afin de déterminer s'il était nécessaire de fournir une formation supplémentaire aux intervieweurs.

Le tableau qui suit donne une description complète des protocoles suivis par les intervieweurs pour mesurer la taille et le poids.

Procédures pour la mesure de la taille et du poids

Procédures générales

- 1- Toutes les procédures doivent être expliquées au répondant avant de commencer.
- 2- Ne lisez pas la mesure à haute voix. Si le répondant demande la mesure, vous pouvez alors la lui communiquer.
- 3- Quand vous prenez les mesures, tenez-vous sur le côté et légèrement à l'écart du répondant. Penchez-vous pour prendre la mesure, en veillant toujours à ne pas empiéter sur l'« espace personnel » du répondant. Ne vous tenez jamais directement en face du répondant, ni derrière lui.

Procédures de mesure du poids

Instrument(s) : Balance et panneau en plastique solide d'un quart de pouce d'épaisseur

Placer la balance sur le panneau en plastique sur une surface uniforme pour la mesure du poids. Dans la mesure du possible, la surface ne devrait pas être recouverte de tapis. Remettez toujours la balance à zéro en appuyant sur le bouton situé sur le devant de celle-ci. Cela peut se faire du bout du pied. Attendez que l'affichage soit 0.000 et qu'un petit « 0 » apparaisse dans le coin supérieur gauche de la balance. Quand les deux chiffres sont visibles, vous êtes prêt pour peser le répondant.

Nota : Si la balance affiche n'importe quelle autre valeur que « 0.000 » au moment du réglage, cela signifie que les piles sont presque vides. Un jeu supplémentaire de piles sera remis à chaque intervieweur au cas où le jeu original s'userait.

Procédures de mesure de la taille

Instrument(s) : Mètre ruban, panneau en plastique solide d'un quart de pouce d'épaisseur, panneau de marquage triangulaire, papillon adhésif et crayon/stylo.

- 1- Mesurez la taille dans une pièce où la surface est lisse (pas de tapis ou tapis le plus fin possible).
- 2- Placez le panneau en plastique solide contre un mur plat. Évitez les plinthes dans la mesure du possible.
- 3- Placez un feuillet adhésif sur le mur à la hauteur approximative du répondant.
- 4- Demandez au répondant (de préférence sans qu'il porte de chaussures ni de chapeau) de :
 - se tenir droit en gardant les bras le long du corps;
 - garder les pieds joints;
 - garder les talons et le dos contre le mur;
 - regarder droit devant lui;
 - se tenir debout en étant aussi grand que possible;
 - rentrer le menton;
 - inspirer profondément pendant que la mesure est faite.
- 5- Placez le panneau de marquage contre le mur et faites-le glisser vers le bas jusqu'à ce qu'il s'enfonce dans les cheveux et soit fermement en contact avec la tête du répondant.
- 6- Veillez à ce que la base du panneau de marquage soit parallèle au sol.
- 7- Avec un crayon, faites une marque sur le feuillet adhésif au point le plus bas où le panneau de marquage est en contact.
- 8- Demandez au répondant de s'écartier du mur afin que vous puissiez mesurer la distance entre le sol et la marque sur le papillon adhésif.

- 9- Étendez le mètre ruban le long du panneau en plastique et faites la mesure au niveau de la marque à 0,5 cm près.
- 10- Assurez-vous que le mètre ruban soit vertical et ne se replie pas au contact du sol.

6.7.2 Collecte de données sur les suppléments de vitamines et de minéraux

Le module portant sur les suppléments de vitamines et de minéraux du volet général sur la santé a permis de recueillir des renseignements détaillés sur les suppléments consommés par le répondant (jusqu'à un maximum de dix). On a demandé aux répondants de fournir la bouteille ou l'emballage de chacun de leurs suppléments et, lorsque c'était possible, le numéro d'identification du médicament (DIN) figurant sur l'emballage a été enregistré. L'application d'interview assistée par ordinateur (IAO) était dotée d'une copie de la Base de données sur les produits pharmaceutiques (BDPP) de septembre 2003, de sorte que le DIN entré par l'intervieweur pouvait être vérifié immédiatement dans la BDPP. Si le DIN entré figurait dans la base de données, le nom de marque associé à ce DIN était entré automatiquement à la question NSD_N2 pour être confirmé par l'intervieweur. Si le DIN ne figurait pas dans la base de données, l'intervieweur inscrivait le nom du produit ainsi que deux ingrédients au maximum et leurs concentrations.

Pour chaque supplément, on a également demandé aux répondants la fréquence de consommation au cours du dernier mois (nombre de fois par jour, par semaine ou par mois), le nombre de fois qu'ils avaient pris un supplément chaque jour où ils l'ont consommé (dans le cas des répondants déclarant une consommation hebdomadaire ou mensuelle) et la quantité du supplément consommé chaque fois (p. ex., trois comprimés chaque fois).

6.7.3 Application informatisée pour le rappel alimentaire de 24 heures

Le rappel alimentaire de 24 heures est la principale composante de l'ESCC 2.2. Cette composante utilise une application d'interview informatisée appelée Automated Multiple Pass Method (AMPM). Cet outil informatisé de collecte de données sur les aliments a été développé au départ par le Département de l'Agriculture des États-Unis (USDA). Par suite de consultations auprès d'experts en nutrition de Santé Canada, l'AMPM a été traduite en français et adaptée pour recueillir des données concernant les aliments consommés au Canada. L'AMPM a été rédigée au moyen du langage de programmation Blaise, qui est largement utilisé en Amérique du Nord pour l'élaboration d'instruments de collecte de données.

L'AMPM comprend une compilation exhaustive de questions uniformisées propres à des aliments et d'options de réponses possibles. Elle comporte un enchaînement automatisé des questions en fonction des réponses précédentes. L'application d'interview repose sur un système en cinq étapes, afin de guider les répondants au sujet de leur consommation d'aliments sur une période de référence de 24 heures. L'application offre diverses possibilités aux répondants pour qu'ils se rappellent tous les aliments, y compris les boissons, qu'ils ont consommés au cours de la période de 24 heures précédant l'interview, soit de minuit à minuit. On recueille aussi des détails qui ont des répercussions sur la composition en éléments nutritifs, comme les ajouts aux aliments, les quantités d'aliments et les méthodes de cuisson. On recueille aussi les marques de commerce pour des aliments particuliers. On recueille en outre des données concernant l'heure à laquelle les aliments ont été consommés, l'occasion de consommation et l'endroit où ils ont été préparés¹⁸.

18. Voir la section 4.2 pour des renseignements additionnels sur l'AMPM.

L'un des principaux avantages de l'utilisation de cette application est qu'il n'est pas nécessaire qu'un nutritionniste qualifié fasse l'interview. Cependant, il s'agit de l'une des applications informatiques les plus complexes jamais utilisées sur le terrain par Statistique Canada. Étant donné cette complexité, la plupart de la formation des intervieweurs s'est concentrée sur l'utilisation de l'application et la réalisation d'exercices pratiques comprenant de nombreux scénarios susceptibles de se produire sur le terrain. Les intervieweurs ont reçu une formation de 3,5 jours avant de commencer le travail sur le terrain.

6.7.3.1 Caractéristiques de l'AMPM servant à recueillir des détails sur les aliments et sur les quantités consommés

Au cours de la collecte des données, lorsque l'on utilise l'application d'interview AMPM, les aliments déclarés par les répondants sont sélectionnés à partir d'un fichier appelé « Liste principale d'aliments » (LPA) et inscrits dans un tableau appelé « Liste d'aliments du répondant » (LAR). La LPA est une liste principale de plus de 2 500 aliments qui s'ouvre automatiquement lorsque l'intervieweur entre au moins trois lettres du nom d'un aliment. Des données sur la fréquence tirées d'enquêtes antérieures sur la nutrition menées aux États-Unis et dans les provinces canadiennes ont servi à déterminer les aliments couramment consommés à inscrire dans la LPA. Étant donné le nombre important et la diversité d'aliments au Canada, l'AMPM comprend de nombreuses questions détaillées différentes sur les aliments, ainsi qu'un nombre très important d'options de réponses possibles à chaque question. Chaque réponse est programmée pour que l'application passe automatiquement à la question appropriée suivante¹⁹.

Les aliments ont été regroupés en 132 catégories dans l'AMPM, afin de faciliter le couplage des 2 500 aliments de la LPA et des questions détaillées sur les aliments. Un code de catégorie d'aliment attribué à chaque aliment fait en sorte que l'AMPM retrouve l'ensemble approprié de questions à poser pour obtenir des descriptions détaillées des aliments et des boissons déclarés, ainsi que les quantités consommées. Par exemple, l'application comporte des questions comme les suivantes : « Quelle sorte était-ce? », « Quels étaient les ingrédients? », « Était-ce un plat frais, en conserve, surgelé, séché, cuit, ou autre chose? » etc.

Autrement dit, selon la catégorie d'aliments, l'application AMPM recueille automatiquement tous les renseignements requis auprès du répondant. Par exemple, si le répondant déclare avoir mangé une « rôtie pour déjeuner », l'application passera à la question suivante pour recueillir des détails : « *S'agissait-il de pain blanc, de blé, de blé entier, de seigle, pumpernickel, multigrains ou autre chose?* ». Des questions sur la quantité, comme « *Combien de tranches de pain avez-vous vraiment mangé?* » font suite à ces questions d'approfondissement. Il convient de souligner que la question sur la quantité porte sur la quantité réelle consommée, et non pas sur la quantité servie.

Un répondant peut déclarer des quantités d'aliments de nombreuses façons. Les unités servant à estimer la quantité d'aliments consommée varient selon le type d'aliment. Les options de réponses peuvent comprendre les suivantes²⁰ :

- Mesures de volume : cuillère à table, tasse, once liquide, pinte, millilitre, litre, etc.
- Mesures de poids : once, gramme, livre, etc.

19. Raper, N., Perloff, B., Ingwersen, L., Steinfeldt, L., Anand, J., An overview of USDA's Dietary Intake Data System. *Journal of Food Composition and Analysis* 17 (2004) 545-555.

20. Raper, N., Perloff, B., Ingwersen, L., Steinfeldt, L., Anand, J., An overview of USDA's Dietary Intake Data System. *Journal of Food Composition and Analysis* 17 (2004) 545-555.

- Dimensions : longueur, largeur, épaisseur, diamètre
- Mesures générales : grandeurs relatives (petit, moyen, grand) contenant (bouteille, canette, boîte, berlingot de jus)
- Mesures particulières pour des marques d'aliments
- Mesures types d'aliments comprises dans le Livret sur les modèles de portions²¹.

6.7.4 Interview des enfants et des jeunes

Lors des cycles précédents de l'ESCC, la population cible était les personnes de 12 ans et plus. Le cycle 2.2 est le premier qui inclut les enfants de moins de 12 ans. Par conséquent, des lignes directrices ont été établies pour faciliter l'interview des enfants et des jeunes.

Avant d'interviewer les jeunes (de 12 à 17 ans), une lettre d'introduction a été remise aux parents ou aux gardiens. Cette lettre expliquait la raison de recueillir l'information auprès du jeune et énumérait les sujets couverts durant l'enquête. Les intervieweurs ont reçu l'instruction d'obtenir du parent/gardien l'autorisation verbale d'interviewer le jeune sélectionné.

Quand il était difficile d'interviewer le jeune en privé, l'intervieweur avait l'instruction de proposer au parent/gardien de lui laisser lire les questions à haute voix et de demander à l'enfant d'entrer ses réponses directement dans l'ordinateur. Les cas pour lesquels il était impossible de respecter le caractère privé et la confidentialité des renseignements ont été codés comme un refus et une note permanente a été faite indiquant qu'il n'avait pas été possible d'assurer le respect de la vie privée et de la confidentialité.

L'avis de consentement au partage et au couplage des données a été donné à tous les répondants. Cependant, pour les enfants de moins de 18 ans, c'est au parent/gardien qu'on a demandé de donner le consentement.

Dans le cas des jeunes enfants, l'interview a été réalisée auprès du parent uniquement (enfant de 0 à 5 ans) ou simultanément auprès du parent et de l'enfant (enfants de 6 à 11 ans). Des renseignements supplémentaires sur l'interview par procuration figurent à la section 6.5.

6.8 Respect de la vie privée

Afin d'assurer la qualité des données recueillies, on s'est efforcé par tous les moyens de réaliser les interviews en privé. Dans certaines situations, le répondant a autorisé une autre personne à assister à l'interview. Dans le fichier de données, des indicateurs²² montrent si une autre personne que le répondant était présente durant l'interview (ADMD_N11) et si l'intervieweur a eu le sentiment que la présence de cette personne a influencé les réponses du répondant (ADMD_N11).

21. Au cours de l'interview, au moment de la collecte des données du rappel alimentaire de 24 heures, on utilise un « *Livret sur les modèles de portions* » comme guide principal de mesure pour aider le répondant à décrire la portion ou la quantité de nourriture consommée. Le livret comprend des images de diverses tailles de verres, de tasses, de bols, de monticules/plaquettes/tartinades, et de cercles. En outre, il comprend une grille, deux quartiers et une page de formes et de morceaux de poulet.

22. Ces indicateurs sont disponibles sur les deux fichiers de données suivants : HS (aussi appelé fichier « Général sur la santé et rappel alimentaire de 24 heures ») et HS_SIDE.

7. Traitement des données

7.1 Vérification

L'application d'interview de l'ESCC 2.2 comprend deux volets distincts, mais complémentaires, soit : 1) le volet **État de santé général** (y compris les suppléments de vitamines et de minéraux), et 2) le volet **Rappel alimentaire de 24 heures**.

7.1.1 Vérification des données du volet « général sur la santé »

Comme son nom l'indique, le volet « général sur la santé » sert à recueillir des données au sujet de l'état de santé général du répondant, y compris les problèmes de santé chroniques, de même que les comportements liés à la santé, comme les activités physiques et sédentaires, l'usage du tabac et la consommation d'alcool. Elle sert en outre à recueillir des données concernant la taille et le poids du répondant, ainsi que ses caractéristiques sociodémographiques.

L'application d'interview assistée par ordinateur (IAO) a servi à la collecte des données sur le volet « général sur la santé ». Cette application a été programmée en Blaise au moyen d'une approche similaire à l'application d'interview développée pour les autres cycles de l'ESCC. La majeure partie de la vérification des données a été faite au moment de l'interview, l'application d'IAO empêchant les intervieweurs d'entrer des valeurs situées en dehors des fourchettes permises. L'enchaînement des questions était contrôlé par les instructions « Passez à » programmées. Par exemple, l'application d'IAO assurait que toute question ne s'appliquant qu'aux enfants ne soit pas posée à des adultes.

En réponse à certains types de déclarations incohérentes ou inhabituelles, l'application affichait un message d'avertissement, mais aucune mesure corrective n'était prise au moment de l'interview. Au besoin, on a élaboré des règles de vérification qui ont été appliquées après la collecte des données, au bureau central de Statistique Canada. En général, les incohérences ont été corrigées en attribuant le code de « non déclaré » à l'une des variables en cause, ou aux deux.

7.1.2 Vérification des données sur les suppléments de vitamines et de minéraux

Vérification du DIN

La vérification des DIN des suppléments de vitamines et de minéraux s'est faite en deux étapes. À la première étape, les intervieweurs ont procédé à l'appariement sur le terrain du DIN et du nom du produit figurant sur l'emballage au DIN et au nom du produit dans la version de la BDPP dans l'ordinateur portable. Dans un petit nombre de cas, les DIN non valides ont été retournés au bureau central. Ces DIN ont ensuite été soumis à un deuxième cycle de vérification en vue de corriger le DIN original. Dans bon nombre de ces cas, les chiffres composant le DIN avaient été entrés dans un ordre incorrect ou bien des 0 avaient été entrés à la place de zéros. En outre, les produits dont il a été déterminé qu'ils étaient des suppléments à base de plantes médicinales et ceux contenant des vitamines et/ou des minéraux autres que ceux dont le profil était établi, ont été éliminés du fichier de données. Dans le cas d'une vérification du DIN, l'indicateur VMDDFDIN dans le fichier de détail des vitamines et des minéraux (VMD) est fixé à 2.

Vérifications autres que du DIN

Le deuxième groupe de vérifications était celui des vérifications autres que du DIN. Environ 3 % des enregistrements ont dû être corrigés pour des raisons autres que le DIN. On a d'abord examiné les cas pour déterminer les valeurs aberrantes. Les cas où la quantité de chaque nutriment consommé sous forme de supplément était au moins le double de la moyenne calculée dans les enquêtes provinciales sur la nutrition menées en Colombie-Britannique (pour les adultes) et au Québec (pour les enfants) ont été examinés par des experts en nutrition.

Ces vérifications autres que du DIN ont entraîné des modifications des variables des habitudes de consommation (VMDD_3, VMDD_N3, VMDD_4 et VMDD_5). Dans certains cas, ces enregistrements comprenaient des réponses répétitives probablement attribuables au fait que le répondant avait mal interprété les questions du module. Par exemple, certains répondants ont déclaré des chiffres très élevés au nombre d'unités de suppléments prises chaque fois (VMDD_5) et au nombre de fois par jour où ils prenaient le supplément (VMDD_3). On a examiné ces enregistrements de façon plus détaillée pour déterminer si la quantité consommée était raisonnable. Les enregistrements pour lesquels le profil de consommation semblait non raisonnable ont été corrigés. Dans le cas d'une vérification autre que du DIN, l'indicateur VMDDFEDT dans le fichier de détail des vitamines et des minéraux (VMD) est fixé à 2.

7.1.3 Calcul de la concentration du supplément

Le fichier de description des vitamines et des minéraux (VDC) comprend, pour chaque supplément, un enregistrement qui indique la concentration de chaque nutriment contenu dans le supplément. On calcule ces concentrations en multipliant la quantité de chaque ingrédient figurant dans la BDPP par un facteur de contenu. Le facteur de contenu décrit quelle quantité du nutriment contient cet ingrédient selon des facteurs tels que l'assimilabilité du nutriment et son absorption. Par exemple, les facteurs de contenu représentant la quantité de fer contenue dans le gluconate de fer et la quantité de fer contenue dans le sulfate de fer seront différents. Dans certains cas, les nutriments contenus dans des suppléments peuvent provenir de différentes sources. En pareil cas, on calcule la concentration séparément pour chaque ingrédient, puis on additionne les résultats. D'après ce calcul, les nutriments sont déclarés en unités normalisées (par exemple, le fer est toujours déclaré en mg). Les utilisateurs doivent savoir qu'à cause de ce calcul, les concentrations d'ingrédients figurant à l'endos de l'emballage ou dans la BDPP ne correspondent pas toujours aux concentrations des nutriments figurant dans le VDC.

La quantité de chaque nutriment présente dans un supplément donné est indiquée dans le VDC selon l'unité posologique utilisée dans la BDPP. L'unité posologique est une combinaison de la dose (la valeur de mesure, p. ex., 1, 5, 10) et de la forme posologique (p. ex., comprimé, mg, etc.). Par exemple, l'unité posologique d'un supplément de fer pourrait être deux comprimés par jour. L'unité posologique indiquée dans la BDPP sert à calculer les concentrations des nutriments et peut ne pas correspondre à l'unité posologique recommandée indiquée sur l'emballage du produit (qui pourrait varier selon l'âge du consommateur).

Les valeurs des nutriments dans le fichier de détail des vitamines et des minéraux (VMD) et dans le fichier sommaire des vitamines et des minéraux (VSD) représentent des moyennes quotidiennes. On trouvera plus de renseignements sur ces calculs dans la documentation sur les variables dérivées.

7.1.4 Vérification des données du « rappel alimentaire de 24 heures »

Le rappel alimentaire de 24 heures a été conçu pour recueillir des données sur tous les aliments et boissons consommés au cours de la période de 24 heures précédant l'interview, soit de minuit à minuit. Cette composante utilise aussi une application d'IAO. Cette dernière est toutefois différente à certains égards, étant donné qu'elle a été adaptée à partir d'une application d'interview informatisée appelée Automated Multiple Pass Method (AMPM), développée par le Département de l'Agriculture des États-Unis (USDA). L'AMPM est un outil informatisé de collecte de données sur les aliments, qui est largement utilisé dans les enquêtes sur la nutrition aux États-Unis. Pour l'ESCC 2.2, l'AMPM a été mise à jour pour mieux rendre compte des normes alimentaires et de la disponibilité des aliments au Canada.

L'AMPM comprend une compilation exhaustive de questions uniformisées propres à des aliments et d'options de réponses possibles. Au cours de la collecte des données, les entrées incorrectes ont été réduites au minimum en raison de l'enchaînement automatisé des questions et des vérifications intégrées dans l'AMPM. Les contrôles suivants ont aussi été inclus dans l'AMPM, afin d'assurer la qualité et la cohérence logique des données sur les aliments²³ :

- une « vérification de l'occasion de consommation » a été incluse, afin d'éviter la déclaration de deux occasions de consommation différentes en même temps (sauf dans le cas des aliments pour lesquels une « consommation élargie »²⁴ est déclarée);
- des vérifications de « combinaison impossible » ont été incluses pour prévenir les combinaisons de réponses illogiques, dans les cas où plusieurs réponses étaient acceptées. Par exemple, dans le cas de la question concernant le type de croustilles, les contrôles n'ont pas permis certaines combinaisons de réponses, comme « à faible teneur en matières grasses » et « à teneur régulière en matières grasses ».
- Des contrôles des données manquantes servent à rappeler à l'intervieweur que certaines données doivent être entrées avant de passer à la question suivante. Par exemple, si un répondant déclare avoir mangé une salade verte, mais que l'intervieweur n'a entré aucun légume vert (p. ex., laitue, épinards, etc.) dans les ingrédients, un contrôle lui rappelle d'entrer des « légumes verts ».

7.2 Exigences minimales de réponse

L'une des premières étapes du traitement des données du cycle 2.2 de l'ESCC a été de définir les critères permettant de déterminer ce qui pouvait être considéré comme un ménage répondant. Dans certains cas, aucune information n'a été recueillie auprès d'un ménage échantillonné. Cela s'est produit, par exemple, quand l'intervieweur n'a pu prendre contact avec le membre sélectionné du ménage pendant la période entière de collecte. Dans d'autres cas, le ménage a refusé de participer à l'enquête à cause de circonstances spéciales, comme une maladie ou un décès dans la famille, ou des conditions météorologiques extrêmes qui ont empêché l'interview d'avoir lieu. Dans ce genre de situation où l'on ne disposait pas d'information suffisante, le ménage a été éliminé de l'échantillon de répondants et traité comme un cas de non-réponse.

23 .Raper, N., Perloff, B., Ingwersen, L., Steinfeldt, L., Anand, J., An overview of USDA's Dietary Intake Data System. *Journal of Food Composition and Analysis* 17 (2004) 545-555.

24. Pour la collecte des données, on a demandé aux intervieweurs d'encourager les répondants à déclarer les occasions de consommation à des moments distincts s'ils le pouvaient. Toutefois, on a qualifié de « consommation élargie » la consommation à peu près continue d'aliments, sans début ou fin distincte, OU les cas où le répondant pouvait se rappeler plus aisément la quantité totale consommée que les moments particuliers (p. ex., le nombre de tasses de café bues à partir de la même cafetière de huit tasses).

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 pu reprendre contact avec le répondant. Dans ces situations, il a fallu définir des critères pour décider du sort à réserver à ces interviews « partielles ». Pour le cycle 2.2 de l'ESCC, pour qu'une « interview partielle » soit acceptée, le répondant devait avoir participé entièrement au rappel alimentaire de 24 heures et avoir répondu à un nombre minimal de questions du questionnaire général sur la santé allant jusqu'à la fin du module sur la sécurité alimentaire. 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 le cas a été achevé partiellement ou entièrement.

7.3 Codage des questions ouvertes

7.3.1 Questions non liées aux aliments

Dans le volet général sur la santé du questionnaire, des catégories de réponse précodées ont été fournies pour toutes les variables appropriées. Dans les cas où la réponse donnée par le répondant ne pouvait être assignée facilement à une catégorie existante, l'intervieweur pouvait poser plusieurs questions lui permettant d'entrer une réponse en toute lettre dans la catégorie « Autre – précisez ». Les réponses à toutes ces questions ont été examinées attentivement lors du traitement des données au bureau central.

Par exemple, la question PSS_N2 (qui a trait aux raisons pour lesquelles l'intervieweur n'a pu mesurer la taille et le poids du répondant) contenait une liste de catégories de réponse dont la dernière était « Autre (précisez) ». Ces réponses en toutes lettres ont été examinées au bureau central où elles ont été recodées ou maintenues dans la catégorie « Autre ». Lors du recodage, on a donné aux réponses le code d'une catégorie existante ou celui d'une catégorie créée spécialement. Pour la question susmentionnée, les nouvelles catégories de réponse MHWD_N2I à MHWD_N2L ont été créées.

7.3.2 Codage des suppléments de vitamines et de minéraux

Le DIN était manquant dans le cas d'environ le tiers de tous les enregistrements des suppléments de vitamines et de minéraux déclarés et ces suppléments ont dû être codés manuellement au bureau central. Une proportion beaucoup plus petite d'enregistrements comprenaient des DIN invalides et ont dû être envoyés à la vérification. En tout, près de 35 % des enregistrements ont dû être vérifiés ou codés (DIN) manuellement.

La base de données utilisée pour coder les enregistrements des suppléments où le DIN était manquant était différente de la base de données BDPP utilisée par les intervieweurs durant la collecte des données. Elle représentait une combinaison de la BDPP du printemps 2005 et de la base de données du Nutrition Survey System (NSS). Les deux bases de données ont été combinées de manière à créer une base de données plus complète et plus exacte pour le codage des suppléments dont le DIN était manquant.

Dans les cas où le DIN n'était pas fourni par le répondant, la description du produit a été utilisée pour appairer le supplément à un DIN similaire. Les DIN ont été attribués en fonction de l'âge et du sexe du répondant. Par exemple, si le répondant était un enfant, un DIN approprié à son âge

était attribué. Si un appariement unique n'était toujours pas possible, alors le DIN le plus souvent déclaré pour ce type de supplément était attribué. Dans les cas où le DIN a été attribué durant le traitement au bureau central, l'indicateur VMDDFDIN dans le fichier de détail des vitamines et des minéraux (VMD) est fixé à 2.

Deux autres suppléments, qui ne figuraient pas dans la version de septembre 2003 de la BDPP, sont néanmoins inclus dans ces données, soit l'huile de saumon et une multivitamine pour enfants. Ces deux produits ont été déclarés assez souvent pour mériter d'être inclus. Un DIN « 1 » a été attribué à l'huile de saumon, tandis que le DIN de la multivitamine pour enfants a été tiré d'une version plus récente de la BDPP.

7.3.3 Questions liées aux aliments

Le volet rappel alimentaire de 24 heures du questionnaire comprend une question sur l'endroit où le repas a été préparé. Si la réponse était autre chose qu'« à la maison », on demandait au répondant d'indiquer l'endroit. Un certain nombre de catégories de réponses précodées permettaient aux répondants de sélectionner des endroits comme « *Restaurant avec serveur/serveuse* », « *Comptoir de commandes à emporter* », « *Cafétéria* », etc.²⁵

Dans les cas où la réponse donnée par le répondant ne pouvait être assignée facilement à une catégorie existante, l'intervieweur pouvait entrer la réponse en toutes lettres dans la catégorie « *Autre – précisez* ». Après la collecte des données, ces réponses en toutes lettres ont été examinées au bureau central de Statistique Canada où elles ont été recodées ou maintenues dans la catégorie « *Autre* ». Lors du recodage, on a donné aux réponses le code d'une catégorie existante ou celui d'une catégorie créée spécialement. Par conséquent, les nouvelles catégories de réponse « *Épicerie/dépanneur/autres types de magasins* » (FIDD_LOT = 13) et « *Au travail* » (FIDD_LOT = 14) ont été créées pour cette question.

7.4 NSS et le codage des aliments

Cette section donne une vue d'ensemble du NSS et du processus de codage des aliments du cycle 2.2 de l'ESCC. Durant la collecte de données du cycle 2.2 de l'ESCC, les répondants ont déclaré au total 742 080 aliments. Près de 55 % (411 782) de ces aliments déclarés ont été autocodés et les autres 45 % (329 577) ont été codés manuellement. 621 aliments n'ont pu être codés.

7.4.1 Codage automatisé des aliments

Le volet rappel alimentaire de 24 heures du questionnaire d'enquête du cycle 2.2 de l'ESCC visait à recueillir des renseignements supplémentaires sur tout ce que le répondant avait mangé ou bu au cours d'une période de 24 heures. Les fichiers de données des interviews sur les aliments consommés ont été importés dans un système informatique maison élaboré par Statistique Canada pour coder automatiquement les aliments déclarés par les répondants et leur attribuer un code d'aliment du Nutrition Survey System (NSS). Le NSS est un système de base de données sur les aliments et les nutriments élaboré par Santé Canada qui contient des descriptions des aliments, les codes d'aliment qui y sont associés et les profils nutritionnels de chaque aliment (pour plus de renseignements sur le NSS, voir la sous-section 7.4.5 ci-dessous).

25. La question sur l'endroit n'a pas été posée pour les nourrissons et les enfants allaités au sein puisque le concept d'endroit où les aliments sont préparés n'est pas pertinente dans le cas de l'allaitement naturel.

Le système automatisé maison de codage des aliments a été appliqué aux réponses données aux questions détaillées sur les aliments. Il a été déterminé qu'un code d'aliment du NSS pouvait être attribué automatiquement à un ensemble de réponses aux questions liées aux aliments. La correspondance entre l'ensemble de réponses et le code d'aliment du NSS attribué a fait l'objet de consultation et a été vérifiée auprès de Santé Canada avant d'être mise en œuvre. L'attribution automatique des codes d'aliment permet de faire des économies de temps et améliore la qualité des données en éliminant les incohérences de codage éventuelles.

7.4.2 Codage manuel des aliments

Dans le cycle 2.2 de l'ESCC, tous les aliments déclarés n'ont pu être codés automatiquement par le système automatisé maison de codage des aliments. Par conséquent, des opérateurs à la saisie des données expérimentés ont été recrutés pour coder manuellement, sous la supervision d'un nutritionniste, les aliments ou boissons non codés automatiquement. Le codage manuel des aliments est l'étape qui a pris le plus de temps dans cette enquête. Elle a commencé en mai 2004 et s'est achevée vers la fin de 2005.

Pour les aliments qui n'étaient toujours pas codés, Statistique Canada a élaboré un autre système informatique maison pour procéder au codage manuel des aliments. Ce système, fondé sur les fichiers de données sur les aliments du NSS, a permis d'examiner les réponses aux questions sur les aliments et de trouver un appariement dans la base de données sur les aliments du NSS. Le codage manuel des aliments non codés s'est déroulé en plusieurs étapes : un codeur tapait un nom clé pour un aliment ou une recette de base dans la fonction recherche. Le système affichait alors toutes les descriptions d'aliments dans le NSS contenant des termes appariés. Une fois que le codeur avait sélectionné l'aliment apparié le plus approprié, le système maison attribuait automatiquement le code d'aliment du NSS correspondant à l'aliment déclaré. Par exemple, si la description détaillée de l'aliment donnée par le répondant indiquait qu'il avait bu un verre de jus d'orange enrichi de calcium, non fait de concentré, le commis aux données utilisait la fonction recherche, puis sélectionnait la description « JUS D'ORANGE, GLACÉ, ENRICHÉ EN CALCIUM ». Une fois cette description sélectionnée, le système attribuait automatiquement le code d'aliment 500691 du NSS à l'aliment déclaré.

Si l'aliment déclaré était un « aliment de base », comme le lait et le fromage, qui ne serait pas modifié par le répondant, alors le codeur sélectionnait un aliment dans la base de données du NSS étroitement apparié à l'aliment déclaré. Si l'aliment déclaré était un « aliment-recette »²⁶, alors une recette dans la base de données des recettes du NSS était utilisée. Pour les aliments déterminés comme étant des recettes, une fonction modification des recettes permettait au codeur de visualiser les recettes prédéfinies comprenant une liste des ingrédients et de leurs quantités. Le codeur modifiait ensuite les recettes de manière à ce qu'elles correspondent à l'aliment consommé par le répondant, en substituant des ingrédients, en corrigeant la quantité d'ingrédients dans la portion ou en supprimant un ingrédient existant.

Durant le codage manuel des aliments, lorsqu'un aliment nouveau ou inhabituel était déclaré qui ne figurait pas dans la base de données du NSS ou ne pouvait être apparié étroitement à un autre aliment, le système permettait de créer un indicateur d'aliment *inconnu*. Pour le cycle 2.2 de

26. En règle générale, un aliment est une « recette » lorsqu'il comprend plus d'un ingrédient. Par exemple, une pomme, un bifteck ou un morceau de fromage sont tous des aliments de base, tandis qu'un ragoût de bœuf, un sandwich et une crêpe sont tous des aliments-recettes.

l'ESCC, on a consulté les nutritionnistes du Bureau des sciences de la nutrition (BSN) de Santé Canada. Ils ont mené des recherches sur chaque « aliment inconnu » pour déterminer la différence entre la composition nutritionnelle de l'aliment déclaré et celle d'aliments similaires codés actuellement dans le NSS. Selon leurs recommandations, on a ensuite résolu le problème de codage des « aliments inconnus » en leur attribuant le même code que celui d'un aliment similaire existant dans la base de données du NSS, en sélectionnant et en modifiant une recette existante ou en ajoutant un aliment nouveau à la base de données s'il s'agissait d'un aliment nouveau sur le marché canadien.

Des spécialistes du Bureau des sciences de la nutrition de Santé Canada ayant des connaissances en nutrition ont assuré une formation aux opérateurs à la saisie des données de Statistique Canada. Le Bureau a pris des arrangements spéciaux avec Santé Canada pour qu'un de ses nutritionnistes soit disponible sur les lieux à Statistique Canada pour superviser les codeurs et assurer la qualité et l'exhaustivité des données. Une fois les données sur l'apport alimentaire codées, Statistique Canada a engagé des nutritionnistes de Santé Canada pour examiner et régler les problèmes de codage difficiles. On a consulté également l'équipe de recherche sur la nutrition de Santé Canada au Bureau des sciences de la nutrition et au Bureau des statistiques biologiques et des applications informatiques aux fins de l'examen de certaines situations de codage complexes, de l'approbation des solutions apportées et du calcul des apports en nutriments.

7.4.3 Calcul des tailles des portions

Statistique Canada a élaboré un système informatique maison pour convertir systématiquement la taille des portions d'aliments en poids en grammes, le volume liquide, en millilitres et le volume solide, en centimètres cubes. Ces unités ont été sélectionnées pour faciliter l'exportation des données sur les aliments de l'ESCC dans le NSS une fois les aliments déclarés codés. Le système maison, utilisant la base de données du NSS, a attribué les valeurs correspondantes dans les unités de mesure appropriées à chaque aliment selon les réponses aux questions sur les quantités d'aliments consommées. On a consulté l'équipe de recherche sur la nutrition de Santé Canada aux fins de l'examen des situations complexes concernant les tailles de portions d'aliments et de l'approbation de la solution apportée. Comme le montre le tableau ci-dessous, les répondants pouvaient déclarer la quantité consommée de cinq façons différentes et chacune était traitée différemment par le système maison.

Tableau 7.1 Méthode de conversion des portions

Type de déclaration de la quantité de nourriture	Méthode de conversion
1 - Unités de mesure générales (y compris grammes, millilitres, onces)	Simple conversion en grammes ou en millilitres
2 - Portions standards (y compris cuillerées à thé, tasses, etc.)	Conversion fondée sur des quantités standards
3 - Modélisation (toutes les tailles de portions d'aliments dans le « Livret sur les modèles de portions » ²⁷)	Conversion fondée sur des quantités prédéterminées
4 - Détails de portions d'aliments particuliers (p. ex., un(e) de taille moyenne, une carotte, un petit pain)	Conversion fondée sur les valeurs du NSS ou selon les quantités déterminées par Santé Canada
5 - Autres tailles de portions précisées (p. ex., une bouchée, une portion)	Conversion fondée sur les quantités déterminées par Santé Canada

Une fois les valeurs de mesure dans les unités appropriées attribuées aux tailles de portions déclarées pour chaque aliment, les données sur les aliments ont été importées du système maison dans le NSS. Dans le NSS, tous les volumes (liquides et solides) ont été convertis en quantités en grammes selon les facteurs de conversion figurant dans la base de données du NSS.

7.4.4 Importation des données sur les aliments dans le NSS

Une fois les aliments déclarés codés et les tailles de portions d'aliments converties en quantités exprimées en grammes, les données sur les aliments ont été importées dans le NSS. On a alors utilisé le NSS aux fins de l'examen de la qualité et de la vérification des données (p. ex. pour déterminer les quantités de nutriment démesurées et les tailles de portions démesurées) et du calcul de la composition nutritionnelle des aliments consommés.

7.4.5 Vue d'ensemble : Nutrition Survey System (NSS)

Le NSS est un système informatique et personnalisé de base de données sur les aliments et les nutriments élaboré par l'équipe de recherche du Bureau des sciences de la nutrition (BSN) et du Bureau des statistiques biologiques et des applications informatiques de Santé Canada²⁸. Le système de base de données du NSS comprend des descriptions des aliments, les portions d'aliments et leurs poids ainsi que les valeurs nutritives pour chaque aliment. En outre, un code d'aliment unique est attribué à chaque aliment dans la base de données du NSS.

Le système de base de données NSS peut être utilisé à diverses fins, y compris le calcul des tailles des portions d'aliments et de leurs poids en grammes (selon la portion consommée déclarée), l'examen de la qualité et la vérification des données (p. ex., pour déterminer les

27. Pour l'interview de collecte de l'information du rappel alimentaire de 24 heures, on a utilisé un « Livret sur les modèles de portions » comme principal guide de mesure pour aider le répondant à décrire la taille ou la quantité de nourriture consommée. Le livret comprend des images de différentes tailles de verres, de tasses, de bols, de monticules/plaquettes/tartinades et de cercles. Il comprend en outre une grille, deux quartiers et une page de formes et de morceaux de poulet.

28. B.C. Ministry of Health Services. British Columbia Nutrition Survey: Report on Energy and Nutrient Intakes. Mars 2004. Disponible à l'adresse suivante : <http://www.healthservices.gov.bc.ca/prevent/nutrition>.

quantités de nutriments démesurées et les tailles de portions démesurées) ainsi que pour le calcul de la composition nutritionnelle des aliments consommés. Le contenu de la base de données du NSS a été créé à partir d'une combinaison de plusieurs bases de données sur les aliments et les nutriments, y compris :

- le Fichier canadien sur les éléments nutritifs (FCÉN)
- la base de données de recettes
- les aliments visés par l'enquête

Fichier canadien sur les éléments nutritifs : Le FCÉN est une base de données sur la composition des aliments bilingue et informatisée qui comprend les valeurs moyennes des nutriments dans les aliments disponibles au Canada. Les données du FCÉN sont compilées au BSN de Santé Canada. Bon nombre des données dans le FCÉN proviennent de la Nutrient Database for Standard Reference du United States Department of Agriculture (USDA). Les aliments inclus dans la base de données de l'USDA qui n'existent pas sur le marché canadien sont exclus du FCÉN. D'autres modifications apportées au FCÉN comprennent les niveaux canadiens de normes réglementaires et d'enrichissement ainsi que l'ajout d'aliments disponibles au Canada seulement et de certaines marques de commerce particulières au Canada²⁹. La base de données du FCÉN est mise à jour sur une base permanente. La version du FCÉN utilisée aux fins du cycle 2.2 de l'ESCC était la version augmentée *CNF supplemented 2001b*, qui était la neuvième édition du FCÉN depuis 1981. Cette version du FCÉN comprenait jusqu'à 112 nutriments décrits pour 4 923 aliments.

Base de données de recettes : Outre le FCÉN, le système de base de données NSS comprend une « base de données de recettes » dans laquelle figurent plus de 3 000 recettes, adaptée de la base de données de recettes du United States Department of Agriculture (USDA). La base de données de recettes a été adaptée en fonction de l'approvisionnement alimentaire canadien et des recettes canadiennes (p. ex., la poutine). Dans le cadre du cycle 2.2 de l'ESCC, on a utilisé la base de données de recettes pour saisir les données sur les nutriments contenus dans les mets composés déclarés par les répondants.

Aliments visés par l'enquête : La base de données des « aliments visés par l'enquête » comprend les données nutritionnelles pertinentes sur les nouveaux aliments ou les aliments locaux qui ne sont pas inclus dans le FCÉN ou dans la « base de données de recettes ».

7.5 Vérification/validation des données sur les aliments

Avec l'aide des experts en nutrition de Santé Canada, plusieurs procédures ont été mises en application afin d'assurer l'exactitude des données sur les aliments. La vérification incluait la détection de données aberrantes (p. ex. les quantités de nutriment démesurées et les tailles de portions démesurées) et l'identification des mets incomplets et des interviews. Ci-dessous se trouve une liste résumée d'exemples de stratégies appliquées à l'ESCC 2.2 pour déterminer, dans l'ensemble, l'acceptabilité des données de chaque apport alimentaire de 24 heures :

- Révision des aliments ayant un total en grammes de plus de 1 000 grammes ou un total ≤ 1 gramme.
- Révision de toutes les recettes où un ingrédient représentait moins de 1 % du poids total.

29. Santé Canada, Fichier canadien sur les éléments nutritifs – *Guide de l'utilisateur*, 2005. Disponible à l'adresse suivante : www.hc-sc.gc.ca/fn-an/nutrition/fiche-nutri-data/cnf_downloads-telechargement_fcen_f.html (consulté 2006).

- Procéder à une analyse des observations aberrantes au niveau aliments pour différents profils tels que : énergie, lipides, protéines, sodium, calcium, fer, vitamine C, folate, vitamine A et caféine.
- Procéder à une analyse des observations aberrantes axée sur l'énergie totale.
- Procéder à une analyse des combinaisons anormales de renseignements démographiques et d'aliments. Par exemple :
 - Consommation de lait maternel rapporté par des adultes
 - Répondants âgés d'un an et plus déclarant seulement du lait maternel ou de préparation
 - Répondants âgés de < 12 ans buvant des boissons alcoolisées
 - Répondants âgés de < 10 ans buvant du thé ou du café
 - Repas rapportés à des heures inhabituelles
 - Quantités d'aliments extrêmement grandes
- Révision des données sur l'apport alimentaire pour les enfants âgés de moins d'un an (environ 290 cas)

7.6 Vérification/validation des données sur les suppléments de vitamines et de minéraux

Avec l'aide des experts en nutrition de Santé Canada, de nombreuses procédures ont été mises en œuvre pour garantir l'exactitude des données sur les suppléments de vitamines et de minéraux. Les procédures de vérification visaient, entre autres, à assurer la validité de tous les DIN, la détection des valeurs aberrantes et l'analyse des habitudes de consommation inhabituelles. Voici une liste des stratégies appliquées aux données sur les suppléments de vitamines et de minéraux du cycle 2.2 de l'ESCC pour en déterminer l'exactitude :

- examen des DIN des suppléments pour s'assurer de leur exactitude (par exemple, un examen plus poussé a révélé que certains DIN qui initialement semblaient être corrects comprenaient de petites erreurs, par exemple un « O » au lieu d'un « 0 », ou deux chiffres inversés);
- veiller à ce que les DIN les plus appropriés soient attribués à tous les enregistrements devant être codés manuellement;
- examen des suppléments pour s'assurer qu'ils contiennent des vitamines et(ou) des minéraux (tous les suppléments homéopathiques ou à base de plantes médicinales ont été supprimés);
- examen des suppléments pour s'assurer qu'ils contiennent les vitamines et(ou) les minéraux visés par la présente enquête (tous ceux contenant seulement des nutriments ne faisant pas l'objet d'un profil ont été supprimés);
- examen des enregistrements comprenant des valeurs nutritives représentant au moins le double de l'apport moyen selon les enquêtes sur la nutrition menées en Colombie-Britannique ou au Québec;
- analyse des enregistrements comprenant des habitudes de consommation inhabituelles. Ces cas comprenaient ceux pour lesquels le répondant avait déclaré prendre un nombre élevé de doses relativement peu de fois au cours de l'année, ou bien ceux pour lesquels le nombre de fois où le supplément était pris par jour et le nombre de doses prises chaque fois étaient les mêmes;
- examen des suppléments pris en paquets pour s'assurer que la dose et la forme posologique étaient exactes;
- examen de tous les enregistrements sur les enfants et les nourrissons codés manuellement pour s'assurer qu'ils ne sont pas codés comme suppléments pour adultes.

7.7 Création de variables dérivées et de variables regroupées

Pour faciliter l'analyse des données et réduire au minimum le risque d'erreurs, un certain nombre de variables figurant dans le fichier ont été dérivées en combinant des variables contenues dans le questionnaire de l'enquête. Dans certains cas, les variables dérivées (VD) sont simplement obtenues en regroupant des catégories de réponse. Dans d'autres, plusieurs variables sont combinées pour en former une nouvelle. Le document sur les variables dérivées fournit des détails sur la façon dont sont créées ces variables plus complexes. Habituellement, le cinquième caractère du nom des variables dérivées est un « D », un « G » ou un « F ».

7.8 Pondération

Le principe sur lequel s'appuie l'estimation dans le cas d'un échantillon probabiliste est que, outre elle-même, chaque personne faisant partie de l'échantillon en « représente » plusieurs autres qui ne font pas partie de l'échantillon. Par exemple, dans un échantillon aléatoire simple au 1/50 (2 %) de la population, chaque personne en représente 50. Conformément à la terminologie utilisée ici, chaque personne a un facteur de pondération (ou poids) de 50.

La phase de pondération est l'étape où l'on calcule le poids appliqué à chaque personne échantillonnée. Le poids figure dans le fichier de microdonnées et doit être utilisé pour produire des estimations significatives d'après les résultats de l'enquête. Par exemple, si l'on veut estimer le nombre de personnes qui fumaient quotidiennement, on choisit les enregistrements se rapportant aux personnes qui possèdent cette caractéristique et on additionne les poids inscrits dans ces enregistrements.

La méthode employée pour calculer ces poids est décrite en détail au chapitre intitulé Pondération. Il convient de souligner qu'en raison du taux élevé de non-réponses partielles aux questions sur la mesure de la taille et du poids, un poids spécial WTSD_MHW a été calculé spécialement pour effectuer l'analyse quand sont utilisées les valeurs mesurées de la taille et du poids.

Il est important de souligner qu'un poids ménage a aussi été créé afin d'analyser des questions relatives aux ménages concernant la nutrition, telle l'insécurité alimentaire. Ce poids est disponible dans un fichier de données distinct et les renseignements pertinents peuvent être retrouvés dans un document intitulé « *Enquête sur la santé dans les collectivités canadiennes (ESCC) cycle 2.2 – Nutrition – Poids ménage* » (Statistique Canada. Juin 2004).

7.9 Suppression de renseignements confidentiels

Il convient de souligner que le fichier de microdonnées à grande diffusion (FMGD) diffère en plusieurs points importants du « fichier-maître » de l'enquête gardé par Statistique Canada. Ces différences sont dues aux mesures prises pour protéger l'anonymat des participants à l'enquête. La confidentialité des renseignements fournis par les répondants est assurée grâce à la suppression de valeurs individuelles, au regroupement de variables et au plafonnement des valeurs des variables dans le fichier de microdonnées à grande diffusion. Les utilisateurs des données qui ont besoin de l'information exclue des fichiers de microdonnées à grande diffusion ont trois options : acheter des totalisations personnalisées, communiquer avec les Centres de données de recherche³⁰ ou utiliser le service de télé-accès.

30. Pour des renseignements à jour sur les Centres de données de recherche, consultez : http://www.statcan.ca/francais/rdc/index_f.htm.

8. Structure des fichiers de données

Au total, 35 107 personnes ont répondu à l'ESCC 2.2. La première interview des répondants sélectionnés portait sur les deux composantes du questionnaire de l'enquête : le « volet général sur la santé » et le « rappel alimentaire de 24 heures ». Une deuxième interview comprenant uniquement le rappel alimentaire de 24 heures a été menée auprès d'environ 30 % de ces répondants. Elle devait avoir lieu de 3 à 10 jours après la première interview, et se tenir un jour différent de la semaine, afin d'évaluer les variations dans l'alimentation d'une personne d'un jour à l'autre.

Les données du cycle 2.2. de l'ESCC ont été diffusées en trois vagues. La première diffusion a eu lieu en juillet 2005 et comprenait les données du volet « général sur la santé », y compris les problèmes de santé chroniques, l'indice de masse corporelle (IMC), ainsi que les comportements liés à la santé, comme les activités physiques et sédentaires, l'usage du tabac et la consommation d'alcool. La diffusion des données du 6 juillet 2006 (la deuxième vague) a permis d'intégrer les données de la composante du rappel alimentaire de 24 heures et celles de la première diffusion (sauf les données sur les suppléments de vitamines et de minéraux). La diffusion actuelle, celle des données de avril 2008 (la troisième vague), comprend des fichiers de données sur les suppléments de vitamines et de minéraux. Cette troisième diffusion comprend également de nouvelles variables dérivées de revenu et de taille de population par géographie.

Afin que toutes les données alimentaires et non alimentaires soient faciles à consulter, les données du cycle 2.2. de l'ESCC ont été diffusées dans plusieurs fichiers, avec des couplages appropriés de variables clés pour apparier les fichiers les uns aux autres.

8.1 Description des fichiers de données

Les données du cycle 2.2. de l'ESCC figurent dans dix fichiers de données qui fournissent des renseignements du volet général sur la santé, y compris les données sur les suppléments de vitamines et de minéraux, et de la composante du rappel alimentaire de 24 heures de l'enquête. Une description sommaire des dix fichiers de données figure ci-après (voir aussi le Diagramme des fichiers de données).

Fichier 1 : Santé générale, suppléments de vitamines et de minéraux et rappel alimentaire de 24 heures – HS.txt

- Le fichier HS (santé générale, suppléments de vitamines et de minéraux et rappel alimentaire de 24 heures) est une combinaison de la première diffusion des données du volet « général sur la santé », de la deuxième diffusion du « rappel alimentaire de 24 heures » et de la troisième diffusion des « données sur les suppléments de vitamines et de minéraux ».

Les trois composantes comprennent :

- Santé générale – les résultats des réponses au questionnaire
- Suppléments de vitamines et de minéraux – des indicateurs qui montrent si un répondant a déclaré avoir consommé un nutriment particulier provenant d'un supplément de vitamines ou de minéraux. Il importe de souligner que les données sur les apports en nutriments de chaque supplément figurent dans le fichier VMD et que la somme des apports en nutriments de tous les suppléments de vitamines et de minéraux est indiquée dans le fichier VSD.
- Rappel alimentaire de 24 heures – la somme des apports en nutriments de sources alimentaires consommés chaque jour pour le premier rappel alimentaire de 24 heures. Autrement dit, ce fichier contient toutes les données de la première diffusion, c'est-à-dire un

enregistrement par répondant, auquel s'ajoute un cumul des apports en nutriments du répondant tirés du premier rappel alimentaire. Cet ensemble de données vise à répondre aux besoins de base de la majorité des utilisateurs.

- Un dictionnaire de données, des index alphabétiques et par sujet, ainsi que des clichés d'enregistrements en formats SAS et SPSS, sont disponibles.

Fichier 2 : Rappel de 24 heures - R24.txt

- Le fichier Rappel de 24 heures (R24) comprend un sommaire des apports de nutriments de sources alimentaires consommés chaque jour pour chacun des premier et deuxième rappels alimentaires.
- Ce fichier comprend en outre les données générales sur les aliments recueillies au moyen du rappel alimentaire de 24 heures, par exemple, les données sur la consommation de sel (c.-à-d. le type de sel consommé et la quantité de sel ajoutée au moment de la cuisson), si les apports sont habituels ou inhabituels pour le répondant, et le nombre d'aliments consommés et déclarés dans le premier et le deuxième rappels alimentaires.
- Un dictionnaire de données, ainsi que des index par sujets et alphabétiques et des clichés d'enregistrements en formats SAS et SPSS, sont disponibles.

Fichier 3 : Détails des aliments et ingrédients - FID.txt

- Le fichier Détails des aliments et ingrédients (FID) comprend les valeurs nutritives pour les niveaux aliment de base³¹ et les niveaux ingrédient³² de chacun des premier et deuxième rappels alimentaires.
- Dans le cadre du rappel alimentaire de 24 heures, au moment de recueillir l'information, les répondants pouvaient sélectionner la taille des portions d'aliments à partir d'une liste préétablie générale (p. ex., une cuillère à soupe), utiliser un modèle de portions (p. ex., un morceau de 2 cm par 2 cm par 2 cm) ou choisir une quantité préétablie fondée sur l'aliment sélectionné (p. ex., une banane moyenne). Les valeurs nutritives disponibles dans ce fichier FID sont fondées sur les quantités d'aliments déclarées converties en grammes (pour plus d'information concernant le calcul des quantités d'aliments, veuillez consulter la section 7.4.3).
- Chaque niveau ingrédient comporte un numéro de recette (RECIPEID), une heure (FIDD_TIM), un endroit (FIDDDLOC) et une occasion (FIDD_OCC) dans le fichier FID. Il convient de souligner que les données au niveau recette se trouvent dans le fichier Niveau recette (FRL).
- Les fichiers FID et FRL comportent exactement les mêmes clichés d'enregistrements. Ils peuvent être annexés pour combiner les aliments de base, les recettes et leurs ingrédients.
- Un dictionnaire de données, ainsi que des index par sujet et alphabétiques et des clichés d'enregistrements en formats SAS et SPSS, sont disponibles.

Fichier 4 : Niveau recette - FRL.txt

- Le fichier Niveau recette (FRL) comprend les valeurs nutritives des niveaux recette principale³³ et des niveaux sous-recette³⁴. Par exemple, si une recette principale comporte

31. Un aliment peut être classé dans la catégorie « niveau aliment de base » dans deux situations. Premièrement, lorsque l'aliment ne peut être divisé en d'autres aliments (p. ex., une pomme, du lait 2 %). Deuxièmement, lorsque l'aliment est une recette, mais qu'aucune information nutritionnelle (c.-à-d., contenu nutritionnel et taille des portions) n'est actuellement disponible sur les ingrédients qui composent la recette (p. ex., certaines barres granola).

32. Les « niveaux ingrédient » correspondent aux aliments qui composent une recette. Par exemple, le fromage, le pain commercial et la margarine font partie des niveaux ingrédient d'un « sandwich au fromage ».

des ingrédients, ceux-ci se retrouvent dans le fichier Détails des aliments et ingrédients (FID), mais le cumul des valeurs nutritives de ces ingrédients se trouve dans le fichier FRL. Cela est aussi vrai aux niveaux sous-recette. Par exemple, lorsqu'il y a un troisième niveau sous-recette, les ingrédients s'y rapportant sont regroupés au niveau sous-recette 3. Pour le deuxième niveau sous-recette, les ingrédients sont regroupés jusqu'au deuxième sous-niveau et comprennent les ingrédients du niveau sous-recette 3.

- Chaque niveau recette comporte un numéro de recette (RECIPEID), une heure (FIDD_TIM), un endroit (FIDDDLOC) et une occasion (FIDD_OCC) dans le fichier FRL.
- Les fichiers FID et FRL comportent exactement les mêmes clichés d'enregistrements. Ils peuvent être annexés pour combiner les aliments de base, les recettes et leurs ingrédients.
- Un dictionnaire de données, ainsi que des index par sujets et alphabétiques et des clichés d'enregistrements en formats SAS et en SPSS, sont disponibles.

Fichier 5 : Fichier sommaire des vitamines et des minéraux – VSD.txt

- Le fichier VSD est un fichier des totaux cumulatifs qui comprend un enregistrement pour chaque répondant, que celui-ci ait déclaré ou non avoir consommé un supplément.
- Le fichier VSD comprend les variables pertinentes selon que le répondant a ou n'a pas consommé de suppléments et le nombre de différents suppléments consommés (VSDD_01 et VSDD_13, respectivement), ainsi que les variables pour la consommation quotidienne moyenne de chaque nutriment pour l'ensemble des suppléments consommés.
- Un dictionnaire de données, ainsi que des index alphabétiques et par sujet et des clichés d'enregistrements en formats SAS et SPSS, sont disponibles.

Fichier 6 : Fichier de données détaillées sur les suppléments de vitamines et de minéraux – VMD.txt

- Le fichier VMD comprend un seul enregistrement pour chaque supplément pris par chaque répondant. Il n'y aura pas d'enregistrement dans le fichier VMD pour les répondants qui n'ont pas déclaré consommer un supplément. Étant donné qu'il était permis aux répondants de déclarer jusqu'à concurrence de dix suppléments différents, le fichier peut contenir dix enregistrements pour un seul répondant.
- Chaque enregistrement dans le fichier VMD comprend les habitudes de consommation du répondant et l'apport nutritionnel quotidien moyen de ce supplément.
- Un dictionnaire de données, des index alphabétiques et par sujet, ainsi que des clichés d'enregistrements en formats SAS et SPSS, sont disponibles.

Fichier 7 : Description des aliments - FDC.txt

- Le fichier Description des aliments (FDC) comprend les valeurs nutritives des niveaux aliment de base et des niveaux recette. Les valeurs nutritives disponibles dans ce fichier correspondent à un poids d'un gramme pour un aliment de base ou une recette.

33. Le « niveau recette principale » fait référence aux recettes qui peuvent être divisées en ingrédients ET pour lesquels des informations nutritionnelles (c.-à-d., contenu nutritionnel et taille des portions) sont disponibles sur TOUS les ingrédients qui composent la recette. Les aliments au niveau recette principale comportent un ensemble connexe de niveaux ingrédient et peuvent aussi comporter des niveaux sous-recettes connexes. Un hamburger, une pizza ou du pain fait maison sont des exemples d'aliments de la catégorie « niveau recette principale ».

34. Les « niveaux sous-recette » correspondent aux recettes intégrées au niveau recette principale et comportent leur propre ensemble de niveaux ingrédient connexes. Par exemple, les « œufs à la bénedictine sauce hollandaise » correspondent au niveau recette principale, et la sauce hollandaise, au niveau sous-recette.

- Le fichier FDC comprend en outre des regroupements et descriptions d'aliments, ainsi que des codes d'aliments tirés d'une gamme variée de bases de données sur les aliments et les nutriments. Par exemple, le FDC comprend :
 - a) les codes d'aliments et les descriptions d'aliments du Nutrition Survey System (NSS), qui ont été utilisés dans le cadre de l'ESCC 2.2 pour le codage des aliments déclarés;
 - b) les codes de groupes d'aliments du Fichier canadien sur les nutriments (FCEN);
 - c) les groupes et les codes d'aliments qui ont été établis par le Bureau des sciences de la nutrition (BSN) de Santé Canada, au début des années 90, selon les systèmes britanniques et américains de groupes d'aliments³⁵.
- Un dictionnaire de données, ainsi que des index par sujets et alphabétiques et des clichés d'enregistrements en formats SAS et SPSS, sont disponibles.

Fichier 8 : Fichier de description des vitamines et des minéraux –VDC.txt

- Pour chaque DIN, le fichier VDC comprend une description du produit ainsi que les codes ATC et AHFS. Il comprend également l'unité posologique (la dose et la forme posologique) ainsi que le profil des nutriments par unité posologique.
- Les renseignements sur chaque nutriment sont indiqués en unités normalisées (par exemple, la teneur en fer est donnée en mg).
- Contrairement au FDC, seuls les suppléments qui ont été consommés par les répondants sont inclus dans le VDC.
- Un dictionnaire de données, des index alphabétiques et par sujet, ainsi que des clichés d'enregistrements en formats SAS et SPSS, sont disponibles.

Fichier 9 : Données du cycle 2.2 pour le logiciel SIDE – HS_SIDE.txt

- Le fichier HS_SIDE a été créé pour adapter le fichier HS (Volet général et rappel alimentaire de 24 heures), en vue de son utilisation avec le logiciel SIDE. Ce dernier sert à estimer les distributions d'apports alimentaires³⁶.
- Le logiciel SIDE nécessite un fichier d'entrée qui comporte une ligne par répondant, selon le nombre maximum de rappels alimentaires de 24 heures recueillis auprès du répondant. Étant donné que les répondants pouvaient fournir jusqu'à deux rappels alimentaires et que l'ESCC 2.2 comptait 35 107 répondants, le fichier HS_SIDE comporte 70 214 lignes (35 107 x 2). Dans le cas des répondants qui n'ont pas fourni de deuxième rappel alimentaire, les données de la première interview du volet général sur la santé ont été reprises, mais les valeurs du rappel alimentaire de 24 heures ont été indiquées comme « manquantes ».
- Dans le fichier HS_SIDE, la variable SUPPID sert à déterminer si la ligne correspond à un premier ou à un deuxième rappel alimentaire.
- Pour le fichier HS_SIDE, une touche numérique SIDEID, qui remplace la variable alphanumérique SAMPLEID, a été créée.

35. Les « groupes d'aliments et de recettes du BNS » permettent : a) de catégoriser et de résumer les données détaillées sur les aliments et les recettes recueillies dans le cadre des enquêtes sur la nutrition, b) de faciliter les analyses de la composition des aliments consommés. Santé Canada, les ministères provinciaux de la santé et les universités ont aussi utilisé les groupes d'aliments et de recettes du BSN pour évaluer la contribution des catégories d'aliments aux apports de certains nutriments, selon l'âge et le sexe, le revenu, la scolarité, ainsi que l'endroit où les aliments ont été consommés, notamment.

36. Le logiciel SIDE a été développé par le Center for Survey Statistics and Methodology de l'Iowa State University, de concert avec l'Agriculture Research Service de l'USDA, en 1996. Pour plus de renseignements au sujet de SIDE et du fichier HS_SIDE, voir le document de Statistique Canada intitulé « Enquête sur la santé dans les collectivités canadiennes (ESCC), cycle 2.2 Nutrition – Documentation sur Software for Intake Distribution (SIDE) » Juin 2006.

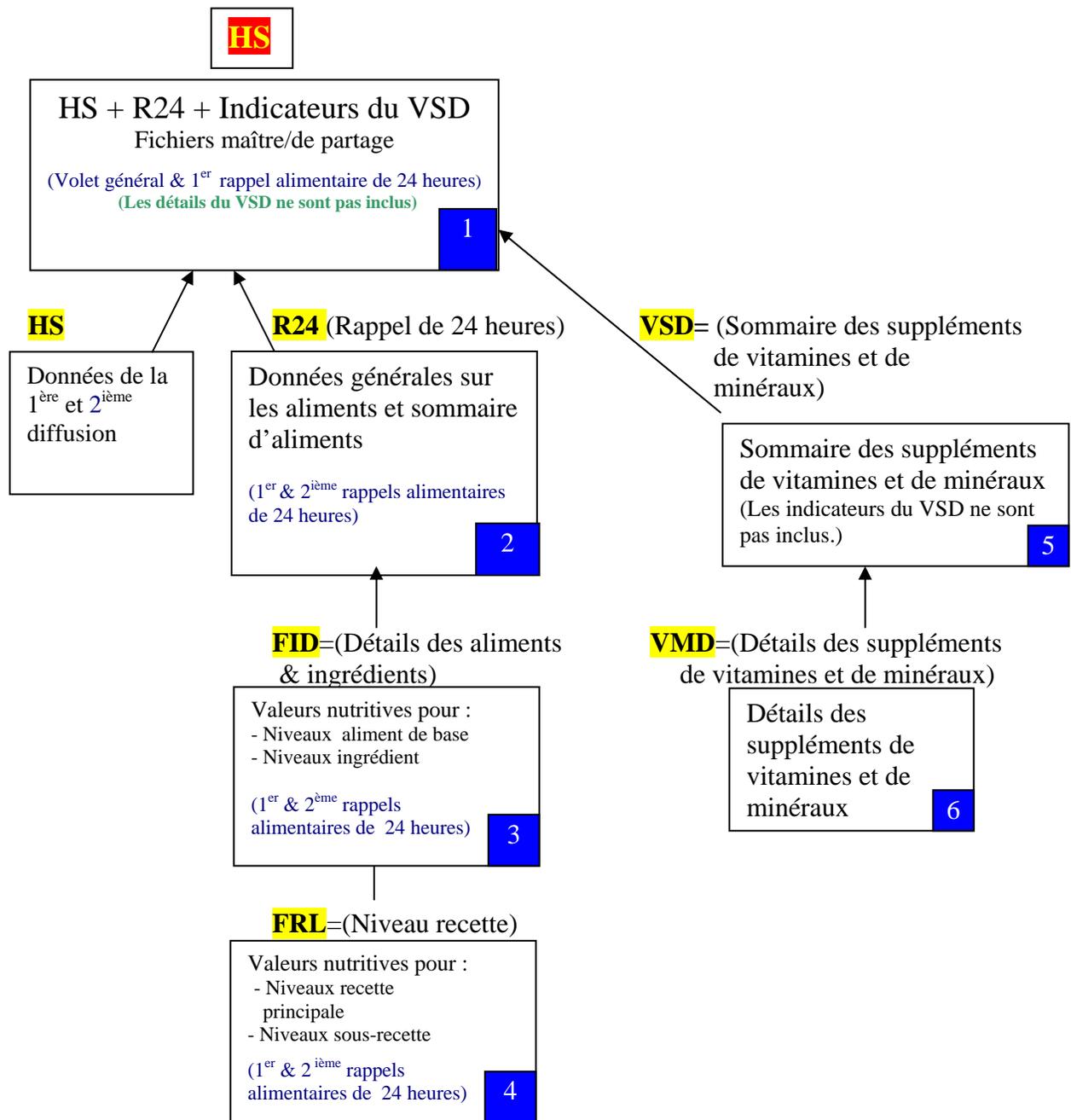
- Dans le fichier HS (Volet général et rappel alimentaire de 24 heures), lorsque les valeurs nutritives n'étaient pas disponibles, le code « non disponibles actuellement » a été attribué. Toutefois, étant donné que le logiciel SIDE ne peut utiliser ce code, il a été converti à « données manquantes » dans le fichier HS_SIDE, afin de permettre l'analyse des données au moyen de SIDE.
- Le dictionnaire de données du fichier HS (Volet général et rappel alimentaire de 24 heures), ainsi que les clichés d'enregistrements, peuvent être utilisés comme documents d'appui pour le fichier HS_SIDE. Les variables additionnelles SIDEID et SUPPID qui ont été ajoutées à la fin du fichier HS_SIDE constituent la seule différence entre le fichier HS et le fichier HS_SIDE du point de vue des variables.

Fichier 10 : Métadonnées

- Il s'agit du dictionnaire électronique de données sous forme de base de données.

L'ESCC 2.2

Troisième diffusion - Schéma des fichiers de données



Fichiers de support

FDC (Description des aliments)

VDC (Description des suppléments de vitamines et de minéraux)

HS_SIDE

Métadonnées

Description d'aliment
- Regroupements et codes d'aliments/de recettes
- Valeur nutritive d'un gramme d'un aliment de base/d'une recette

7

Description des suppléments de vitamines et de minéraux (Seulement les DIN codés)

8

Fichiers HS maîtres/partagés (Fichier de données adapté pour le logiciel SIDE)
(Volet général & 1^{er} & 2^{ième} rappels alimentaires de 24 heures)

9

Métadonnées du dictionnaire électronique de données

10

8.2 Variables clés importantes et définitions

Plusieurs variables incluses dans des fichiers de données particuliers nécessitent d'être expliquées, par exemple, celles énumérées ci-dessous. Ces renseignements seront utiles pour mieux comprendre les fichiers de données et la façon de les coupler. (Pour une description complète de toutes les variables comprises dans chaque fichier de données, voir les rapports des dictionnaires de données appropriés.)

SAMPLEID : Identificateur unique de ménage attribué à chaque répondant de l'enquête.

SUPPID : Identificateur unique de rappel alimentaire de 24 heures attribué au rappel de chaque répondant. Les répondants peuvent fournir jusqu'à deux rappels.

SEQID : Numéro séquentiel unique d'un rappel alimentaire de 24 heures, qui sert à déterminer l'aliment (niveau aliment de base, niveau ingrédient ou niveau recette) déclaré par un répondant. SEQID sert aussi à coupler les ingrédients et les recettes, lorsqu'on annexe un fichier de nutriments à un autre fichier.

RECIPEID : Numéro unique de recette d'un rappel alimentaire de 24 heures qui pourrait servir à coupler une recette et un ensemble d'ingrédients connexe. Les recettes se trouvent dans le fichier Niveau recette (FRL). Les ingrédients liés à un numéro particulier RECIPEID dans un rappel alimentaire de 24 heures se trouvent dans le fichier Détails des aliments et ingrédients (FID).

SIDEID : Identificateur unique de ménage attribué à chaque répondant de l'enquête. SIDEID doit être utilisé avec le logiciel SIDE. Ce logiciel sert à estimer les distributions d'apports alimentaires.

FIDD_CDE : Variable correspondant au code d'aliment du Nutrition Survey System (NSS). Il s'agit d'un code unique pour chaque aliment. Chaque code NSS comporte un profil nutritif particulier. Cette variable figure dans les fichiers FID, FRL et FDC.

FIDD_FID : Variable correspondant à un identificateur d'aliment de base, d'ingrédient ou de recette. Elle permet de faire une distinction entre les niveaux aliment de base, les niveaux ingrédient, les niveaux recette principale et les niveaux sous-recette. Cette variable se trouve dans les fichiers FID, FRL et FDC.

VMDDCDIN: Variable correspondant au numéro d'identification de drogue (DIN) Il s'agit d'un code unique pour chaque supplément de vitamine et/ou de minérale. Chaque DIN comporte un profil nutritif. Cette variable figure dans les fichiers VMD et VDC.

8.3 Variables clés pour le couplage des fichiers de données

Un certain nombre de variables clés sont incluses dans les fichiers de données et devraient servir à les coupler ou à les annexer. Le tableau qui suit fait état de la ou des variables clés servant à coupler ou à annexer des fichiers de données. À noter que tous les autres types de couplages de fichiers ne sont pas possibles ou ne sont pas recommandés.

Coupler le fichier	au fichier	variable(s) clé
HS	R24	SAMPLEID
HS	FID	SAMPLEID

HS	FRL	SAMPLEID
HS	VSD	SAMPLEID
HS	VMD	SAMPLEID
R24	FID	SAMPLEID, SUPPID
R24	FRL	SAMPLEID, SUPPID
R24	VSD	SAMPLEID
R24	VMD	SAMPLEID
FID	FRL	Pas de variable clé*
FID	FDC	FIDD_CDE
FRL	FDC	FIDD_CDE
VSD	VMD	SAMPLEID
VMD	VDC	VMDDCDIN

* Dans ce cas, un fichier est annexé à l'autre fichier pour créer un fichier unique des nutriments. Voir la sous-section suivante pour plus de renseignements.

8.4 Principaux fichiers de données sur les aliments (FID et FRL) et variables clés

Les fichiers Détails des aliments et ingrédients (FID) et Niveau recette (FRL) sont les deux principaux fichiers de données sur les aliments, qui comprennent les valeurs nutritives des aliments de base, des recettes et de leurs ingrédients. La présente sous-section porte sur le rapport entre les fichiers FID et FRL et le rôle des variables clés (p. ex., FIDD_FID, SEQID et RECIPEID), lorsque l'on utilise les deux fichiers. Un exemple est aussi fourni pour illustrer le rapport entre les fichiers FID et FRL, en ce qui a trait aux aliments aux divers niveaux (p. ex. aliment de base, recette et ingrédient), et l'utilisation des variables clés de FIDD_FID, SEQID et RECIPEID.

Rôle de FIDD_FID et détermination des niveaux aliment

FIDD_FID est un identificateur d'aliment de base, de recette ou d'ingrédient qui figure dans les fichiers FID et FRL. Cette variable comporte des codes différents, afin de permettre une distinction entre les niveaux aliment de base, les niveaux ingrédient et les niveaux recette principale et sous-recette. Le tableau qui suit donne un exemple des codes FIDD_FID dans les fichiers FID et FRL.

Emplacement du fichier	Valeur de FIDD_FID*				Type d'aliment
FID	00				Niveau aliment de base
FRL	01				Niveau recette principale
FID		10			Niveaux ingrédient pour le niveau recette principale
FRL		12			Niveau sous-recette 1
FID			20		Niveaux ingrédient pour le niveau sous recette 1
FRL			23		Niveau sous-recette 2
FID				30	Niveaux ingrédient pour le niveau sous recette 2
FRL				34	Niveau sous-recette 3
FID				40	Niveaux ingrédient pour le niveau sous recette 3

*Les aliments du fichier FID comportent une valeur FIDD_FID de 00, 10, 20, 30 ou 40. Les aliments du fichier FRL comportent une valeur FIDD_FID de 01, 12 or 23 ou 34.

Pour mieux comprendre les codes FIDD_FID précédents, une description des aliments aux différents niveaux figure ci-après.

Niveau aliment de base (00) – Un aliment peut être classé dans la catégorie « niveau aliment de base » dans deux situations. Premièrement, lorsque l'aliment ne peut être divisé en d'autres aliments (p. ex., une pomme, du lait 2 %). Deuxièmement, lorsque l'aliment est une recette, mais qu'aucune information nutritionnelle (c.-à-d., contenu nutritionnel et taille des portions) n'est actuellement disponible sur les ingrédients qui composent la recette (p. ex., certaines barres granola). Les données au niveau aliment de base se trouvent dans le fichier de données FID.

Niveau recette principale (01) – Le « niveau recette principale » fait référence aux recettes qui peuvent être divisées en ingrédients ET pour lesquels des informations nutritionnelles (c.-à-d., contenu nutritionnel et taille des portions) sont disponibles sur TOUS les ingrédients qui composent la recette. Les aliments au niveau recette principale comportent un ensemble connexe de niveaux ingrédient et peuvent aussi comporter des niveaux sous-recette. Un hamburger, une pizza ou du pain fait maison sont des exemples d'aliments de la catégorie « niveau recette principale ». Les données au niveau recette principale se trouvent dans le fichier de données FRL.

Niveaux ingrédient (10, 20, 30, 40) – Les niveaux ingrédient correspondent aux aliments qui composent une recette. Par exemple, le fromage, le pain commercial et la margarine font partie des niveaux ingrédient pour un « *sandwich au fromage* ». Les données au niveau ingrédient se trouvent dans le fichier de données FID.

Niveaux sous-recette (12, 23, 34) – Les niveaux sous-recette correspondent à des recettes intégrées dans le niveau recette principale et comportent leur propre ensemble de niveaux ingrédient connexes. Par exemple, les « *œufs à la bénedictine sauce hollandaise* » correspondent à la recette principale, et la sauce hollandaise, au niveau sous-recette. Les données au niveau sous-recette se trouvent dans le fichier de données FRL.

Exemple de la façon d'utiliser les variables clés avec les fichiers FID et FRL

Le tableau qui suit illustre le rapport qui existe entre les fichiers FID et FRL, en ce qui a trait aux aliments des divers niveaux, et l'utilisation des variables FIDD_FID, SEQID et RECIPEID. Cet exemple représente UN répondant et QUATRE aliments consommés (à noter que le niveau principal pour chaque aliment consommé est numéroté et figure en caractères **GRAS**). Lorsqu'on examine le tableau ci-dessous, on note que :

- les valeurs attribuées à SEQID sont uniques à l'intérieur d'un rappel alimentaire de 24 heures dans les fichiers FID et FRL; ces valeurs demeurent dans l'ordre pour le répondant;
- tous les niveaux ingrédient pour un niveau recette donné, y compris les niveaux sous-recette, comportent la même valeur RECIPEID;
- les niveaux ingrédient connexes des fichiers FID et FRL partagent la même heure (FIDD_TIM), la même occasion (FIDD_OCC) et le même endroit (FIDDLOC).

Ex.	Aliment	Fichier FID			Fichier FRL		
		FIDD_FID	SEQID	RECIPEID	FIDD_FID	SEQID	RECIPEID
* 1	FROMAGE, FONDU, PREPARATION, CHEDDAR	00	1	0			
* 2	Fromage, sandwich au				01	2	1
	FROMAGE, FONDU, PREPARATION, CHEDDAR	10	3	1			
	<i>PAIN, BLANC</i>	10	4	1			
	<i>MARGARINE, MOLLE</i>	10	5	1			
* 3	Croissant, avec bacon, oeuf et fromage				01	6	2
	<i>PORC, SALE, BACON</i>	10	7	2			
	<i>CROISSANT, BEURRE</i>	10	8	2			
	<i>FROMAGE, FONDU, PREPARATION, CHEDDAR</i>	10	9	2			
	<i>Omelette ou oeufs brouillés</i>				12	10	2
	<i>LAIT, LIQUIDE</i>	20	11	2			
	<i>MARGARINE, MOLLE</i>	20	12	2			
	<i>Sel</i>	20	13	2			
	<i>OEUF, POULE</i>	20	14	2			
* 4	Croissant avec Jambon, pane				01	15	3
	<i>CROISSANT, BEURRE</i>	10	16	3			
	<i>FROMAGE, FONDU, PREPARATION, CHEDDAR</i>	10	17	3			
	<i>Omelette ou oeufs brouillés, avec Jambon, pane</i>				12	18	3
	<i>LAIT, LIQUIDE</i>	20	19	3			
	<i>MARGARINE, MOLLE</i>	20	20	3			
	<i>Sel</i>	20	21	3			
	<i>OEUF, POULE</i>	20	22	3			
	<i>Jambon, pane, frit</i>				23	23	3
	PORC, SALE, JAMBON	30	24	3			
	OEUF, POULE	30	25	3			
	MIE DE PAIN	30	26	3			
	HUILE VEGETALE	30	27	3			

* Exemple #1: « **Fromage fondu, préparation, cheddar** » est un aliment au niveau aliment de base qui se trouve dans le fichier FID seulement. Il ne comporte pas de niveaux ingrédient ni de niveaux sous-recette.

* Exemple #2: « **Fromage, sandwich au** » est un aliment au niveau recette principale qui se trouve dans le fichier FRL. Les niveaux ingrédient de cette recette se trouvent dans le fichier FID. Il n'y a pas de niveaux sous-recette.

* Exemple #3: « **Croissant, avec bacon, œuf et fromage** » est un aliment au niveau recette principale comportant un niveau sous-recette qui se trouve dans le fichier FRL. Tous les niveaux ingrédient des niveaux recette principale et sous-recette se trouvent dans le fichier FID. À noter que les aliments en italiques correspondent aux ingrédients de la sous-recette « *Omelette ou œufs brouillés* ».

* Exemple #4: « **Croissant avec jambon, pané** » est un aliment au niveau recette principale qui comporte deux niveaux sous-recette qui se trouvent dans le fichier FRL. Tous les niveaux ingrédient des deux niveaux recette principale et tous les niveaux sous-recette se trouvent dans le fichier FID. À noter que les aliments en italiques correspondent aux niveaux ingrédient du niveau sous-recette « *Omelette ou œufs brouillés, avec jambon, pané* ». Les aliments en caractères gras représentent les niveaux ingrédient du niveau sous-recette 2, « **Jambon, pané, frit** ».

9. Pondération

Pour que les estimations produites à partir de données d'enquête soient représentatives de la population couverte, et non pas seulement représentatives de l'échantillon comme tel, l'utilisateur doit incorporer les facteurs de pondération, appelés ici les poids d'enquête, dans ses calculs. Un poids d'enquête est attribué à chaque personne incluse dans l'échantillon final, c'est-à-dire dans l'échantillon de personnes ayant répondu à l'enquête. Ce poids correspond au nombre de personnes représentées par le répondant dans l'ensemble de la population.

Tel que décrit dans la section 5, l'ESCC (cycle 2.2) a eu recours à cinq bases de sondage pour la sélection de son échantillon : une base aréolaire de logements agissant comme base principale, trois bases listes visant principalement les personnes de 18 ans et moins et une base supplémentaire utilisée pour sur-échantillonner les autochtones hors-réserve. Les trois types de bases (principale, liste et supplémentaire) ont pu être combinés pour satisfaire les besoins spécifiques d'une province. Noter toutefois qu'une seule des bases listes pouvait être utilisée à l'intérieur d'une province donnée.

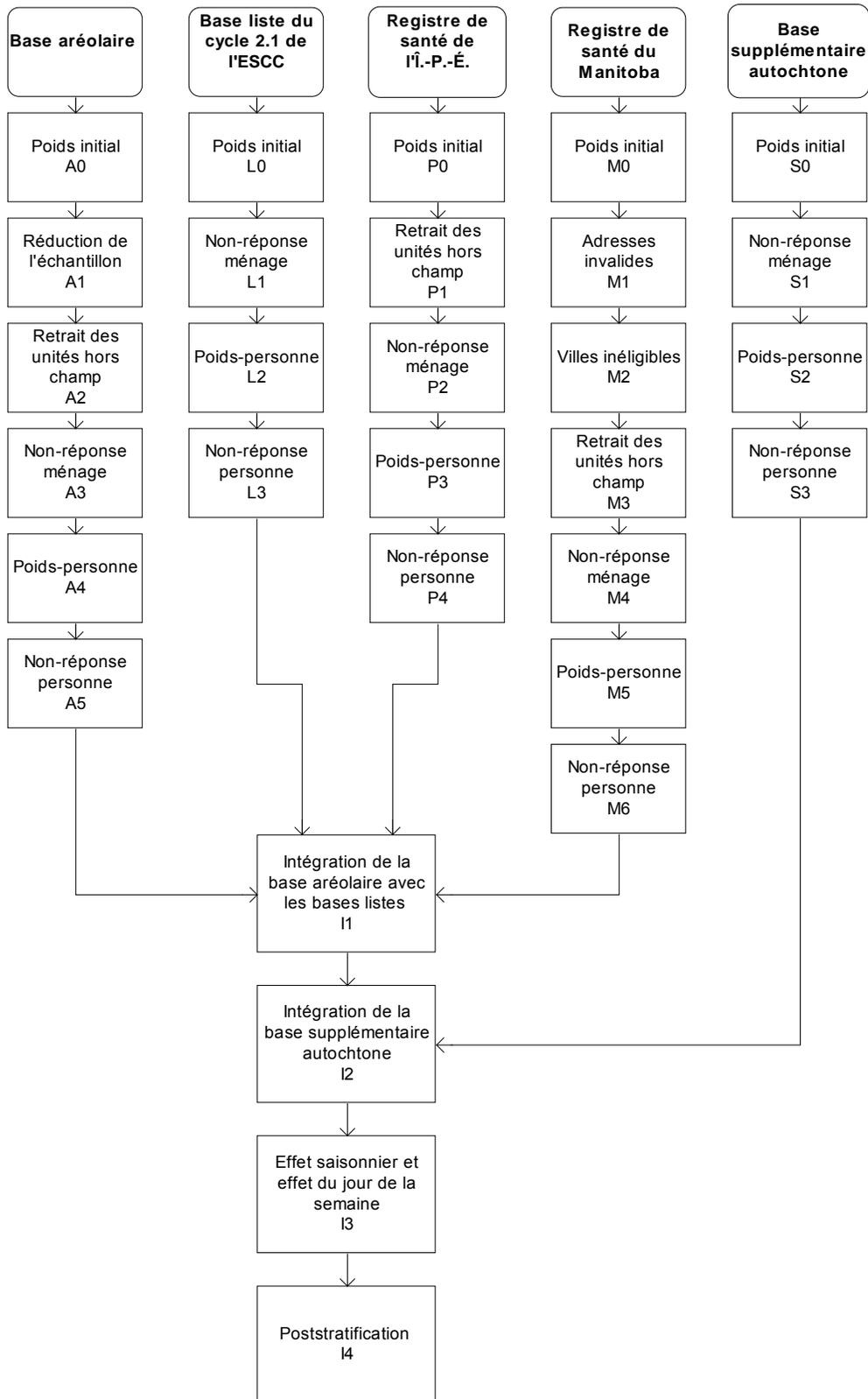
La stratégie de pondération a été développée en traitant séparément les cinq bases. Les poids résultant de ces bases sont ensuite intégrés en un seul ensemble de poids lors des étapes appelées "*intégration*". Suite à quelques ajustements, ce poids intégré devient le poids final.

9.1 Pondération de l'échantillon

Tel que mentionné auparavant, les unités des cinq bases sont traitées séparément jusqu'à l'étape d'intégration. La sous-section 9.1.1 fournit les détails de la stratégie de pondération pour la base aréolaire alors que la sous-section 9.1.2, ceux pour la base liste du cycle 2.1. Les sous-sections 9.1.3 et 9.1.4 traitent des deux autres bases listes et décrivent respectivement la pondération du registre de santé de l'Île-du-Prince-Édouard (Î.-P.-É.) et celle du registre du Manitoba. De son côté, la sous-section 9.1.5 présente la pondération de la base utilisée pour le supplément sur les autochtones hors réserve. L'intégration de la base aréolaire avec les bases listes est traitée en 9.1.6 alors que l'intégration des cas provenant de la base utilisée pour le supplément d'autochtones est expliquée à la sous-section 9.1.7. Les deux étapes finales de la pondération, c'est-à-dire l'ajustement pour contrôler la saisonnalité et l'effet du jour de la semaine, puis la poststratification, sont expliquées dans les sous-sections 9.1.8 et 9.1.9 respectivement. Finalement, la section 9.2 présente la stratégie utilisée pour créer le poids supplémentaire attribué au sous-échantillon de personnes pour lesquelles les mesures exactes de taille et de poids sont disponibles.

Le diagramme A présente un sommaire des différents ajustements faisant partie de la stratégie de pondération dans l'ordre qu'ils sont appliqués. Un système de numérotation est utilisé pour identifier chaque ajustement apporté au poids et sera utilisé tout au long de la section. Les lettres *A*, *L*, *P*, *M* et *S* sont utilisées comme préfixes pour référer aux ajustements appliqués respectivement aux unités de la base **A**réolaire, de la base **L**iste du cycle 2.1, du registre de santé de l'Î.-P.-É., du registre de santé du **M**anitoba et de la base **S**upplémentaire autochtone. Le préfixe *I* est quant à lui utilisé pour identifier les ajustements d'*I*ntégration de même que ceux qui suivent.

Diagramme A: Sommaire de la stratégie de pondération



9.1.1 Pondération de l'échantillon provenant de la base aréolaire

A0 – Poids initial

Puisque le mécanisme utilisé pour sélectionner l'échantillon de la base aréolaire a été celui établi pour l'Enquête sur la population active (EPA), le poids initial a dû être calculé selon les particularités de cette enquête. D'abord, à l'intérieur de chacune des strates définies par l'EPA, des grappes (unités primaires) sont sélectionnées avec probabilités proportionnelles à la taille (selon les comptes de recensement de 1991). À l'intérieur de chacune des grappes sélectionnées, un échantillon de logements est ensuite choisi à l'aide d'un échantillonnage systématique. Le produit des probabilités de chacune de ces sélections représente la probabilité de sélection du logement et son inverse représente le poids initial du logement. Pour plus de détails sur le mécanisme de sélection, de même qu'une définition plus complète des strates et des grappes, se référer à Statistique Canada (1998).³⁷

A1 – Réduction de l'échantillon

Certaines modifications ont dû être faites au mécanisme standard de l'EPA lors de la sélection de l'échantillon pour le cycle 2.2 de l'ESCC. Le plan de l'EPA peut fournir un échantillon d'environ 68 000 logements à l'échelle du pays alors que les besoins du cycle 2.2 de l'ESCC sont plus petits. Le processus de sélection des logements est donc modifié afin d'éliminer les logements superflus et ainsi obtenir les tailles visées dans chaque province et région. Cette réduction de l'échantillon doit être considérée dans la pondération afin de correctement représenter la probabilité de sélection. Un facteur d'ajustement, A1, représentant le taux de réduction de l'échantillon a donc été calculé. Cependant, cette réduction de l'échantillon n'a pas été observée dans certaines strates de l'Île-du-Prince-Édouard, de l'Ontario et du Manitoba où les besoins en échantillon étaient supérieurs à ce que le mécanisme de l'EPA pouvait fournir. En effet, pour certaines strates, le plan de l'EPA conduit à des tailles d'échantillon inférieures à celles requises par le cycle 2.2 de l'ESCC. Pour ces strates, le taux de réduction de l'échantillon qui est calculé représente plutôt un taux d'accroissement. Le poids initial calculé en A0 est multiplié par ce facteur d'ajustement, ce qui permet d'obtenir le poids A1.

A2 – Retrait des unités hors champ

Parmi tous les logements échantillonnés, une certaine proportion de ceux-ci est, lors de la collecte, identifiée comme étant hors du champ de l'enquête. Des logements détruits ou en construction, des logements vacants, saisonniers ou secondaires, de même que des établissements, sont tous des exemples de cas hors champ pour l'ESCC. Ces logements sont tout simplement retirés de l'échantillon, ne laissant plus que les logements faisant partie du champ de l'enquête. Ces derniers conservent donc le même poids qu'à l'étape précédente que l'on appelle maintenant poids A2.

37. STATISTIQUE CANADA. 1998. *Méthodologie de l'enquête sur la population active du Canada*. Produit no 71-526-XPB au catalogue de Statistique Canada. Ottawa.

A3 – Non-réponse ménage

Lors de la collecte, une certaine proportion des logements visités a inévitablement résulté en non-réponse. Ceci survient habituellement lorsque le ménage habitant le logement refuse de participer à l'enquête, fournit des données inutilisables, ou encore, ne peut être rejoint pour réaliser l'interview. Le poids des ménages non-répondants est redistribué aux répondants à l'aide de classes de réponse. L'algorithme CHAID (Chi-Square Automatic Interaction Detector), disponible dans Knowledge Seeker,³⁸ permet d'identifier les caractéristiques qui divisent le mieux l'échantillon en groupes selon leurs propensions à répondre. Noter que ces groupes sont formés de façon indépendante à l'intérieur de chaque province ou région (Ontario et Manitoba). Puisque l'information disponible auprès des non-répondants est très limitée, seules quelques caractéristiques telles que la période de collecte (avec 4 périodes: janvier à mars / avril à juin / juillet à septembre / octobre à décembre) et un indicateur du milieu rural/urbain ont été utilisées pour la création des classes. Un facteur d'ajustement a donc été calculé à l'intérieur de chaque classe de la façon suivante:

$$\frac{\text{Somme des poids A2 pour tous les ménages}}{\text{Somme des poids A2 pour tous les ménages répondants}}$$

Le poids A2 des ménages répondants a donc été multiplié par ce facteur d'ajustement pour produire le poids A3. Les ménages non-répondants sont éliminés du processus de pondération à partir de ce point.

A4 – Création du poids-personne

Puisque l'unité d'échantillonnage finale est la personne, le poids-ménage calculé jusqu'ici doit être converti en un poids-personne. Celui-ci est obtenu en multipliant le poids A3 par l'inverse de la probabilité de sélection de la personne choisie dans le ménage. Nous obtenons ainsi le poids A4. Rappelons que la probabilité de sélection d'une personne à l'intérieur d'un ménage est définie en utilisant les facteurs multiplicatifs de poids de sélection (FMPS) et dépend de la composition en nombre et en âge du ménage (voir section 5.5 pour plus de détails). La probabilité de sélection d'une personne peut même varier selon la province et la période de collecte pendant laquelle le répondant a été sélectionné.

A5 – Non-réponse personne

Dans le cadre du cycle 2.2 de l'ESCC, une interview peut être vue comme un processus en deux étapes. Dans un premier temps, l'intervieweur obtient la liste complète des personnes vivant dans le ménage, puis par la suite interviewe la personne sélectionnée dans le ménage. Dans certains cas, les intervieweurs ne réussissent qu'à compléter la première étape, soit parce qu'ils ne peuvent entrer en contact avec la personne sélectionnée, ou encore parce que la personne sélectionnée refuse d'être interviewée. De tels cas sont définis comme étant des non-réponses à l'échelle de la personne, et un facteur d'ajustement doit être appliqué aux poids des personnes répondantes pour compenser cette non-réponse. Tout comme pour la non-réponse à l'échelle du ménage, l'ajustement est appliqué à l'intérieur de classes définies à partir des caractéristiques disponibles à la fois pour les répondants et non-répondants. Toutes les caractéristiques recueillies lors du listage des membres du ménage étaient en fait disponibles pour créer ces classes.

38. ANGOSS Software. 1995. *Knowledge Seeker IV for Windows - User's Guide*. ANGOSS Software International Limited.

L'algorithme CHAID a été utilisé pour obtenir la définition des classes. Notez que des groupes ont été formés indépendamment à l'intérieur de chaque province (ou région en Ontario et au Manitoba). Selon la région, les caractéristiques suivantes ont pu être utilisées pour former les classes d'ajustement : le sexe, le groupe d'âge, l'indicateur de milieu rural/urbain, le nombre de personnes dans le ménage, l'état matrimonial et la période de collecte. Un facteur d'ajustement est calculé à l'intérieur de chaque classe de la façon suivante:

$$\frac{\text{Somme des poids A4 pour toutes les personnes sélectionnées}}{\text{Somme des poids A4 pour toutes les personnes sélectionnées répondantes}}$$

Le poids A4 des personnes répondantes a donc été multiplié par ce facteur d'ajustement pour produire le poids A5. Les personnes non-répondantes sont éliminées de la pondération à partir de cette étape.

Étant donné que cet ajustement est le dernier nécessaire pour l'échantillon provenant de la base aréolaire, le poids A5 représente donc le **poids final de la base aréolaire**. Ce poids sera plus tard intégré au poids final des bases listes (section 8.1.6) pour créer le poids final du cycle 2.2 de l'ESCC.

9.1.2 Pondération de la base liste du cycle 2.1

Tel que mentionné précédemment, la base liste provenant du cycle 2.1 a été utilisée afin d'atteindre les tailles d'échantillons minimales nécessaires pour la population de 18 ans et moins (et des 71 ans et plus en Ontario). Cette base a été utilisée pour compléter la base aréolaire dans toutes les provinces, sauf au Manitoba et à l'Île-du-Prince-Édouard où des registres de santé ont été utilisés à titre de base complémentaire. Les logements présents sur la base liste du cycle 2.1 proviennent majoritairement des répondants de la base aréolaire du cycle 2.1 étant donné que l'information relative à l'adresse est disponible pour toutes ces unités. Les répondants provenant de l'étude sur l'effet du mode de collecte³⁹ et pour lesquels l'interview avait eu lieu en personne sont également présents sur la base liste du cycle 2.1 étant donné la disponibilité d'une adresse complète pour ces cas. Finalement, puisque la base aréolaire du cycle 2.1 n'arrivait pas à fournir assez de ménages avec des 18 ans et moins pour la Nouvelle-Écosse, le Nouveau-Brunswick et la région de Toronto, les répondants de la base téléphonique du cycle 2.1 ont dû ajoutés à la base liste pour ces régions.

Même si la base liste du cycle 2.1 ciblait les personnes de 18 ans et moins, il était possible d'obtenir des répondants âgés de plus de 18 ans à partir de celle-ci. Effectivement, étant donné que le ménage qui vivait à l'adresse visitée lors du cycle 2.1 pouvait être différent de celui vivant à cette adresse au cycle 2.2, il est possible que ce nouveau ménage ne comportait pas de membres de 18 ans et moins. Or, pour des raisons opérationnelles, les intervieweurs ont dû conduire l'interview même s'il n'y avait pas de 18 ans et moins dans le ménage sélectionné. Cela explique pourquoi il y a quand même des répondants provenant de la base liste du cycle 2.1 qui sont âgés de plus de 18 ans.

39. ST-PIERRE, M. et Y. BÉLAND. 2004. «Mode Effects in the Canadian Community Health Survey: a Comparison of CAPI and CATI». 2004. *Proceedings of the American Statistical Association Annual Meeting, Survey Research Methods Section*. Toronto, Canada: American Statistical Association.

L0 – Poids initial

Le poids initial pour les répondants de la base liste du cycle 2.1 a été calculé selon les particularités de cette base. L'échantillon a été sélectionné en deux temps. Pour les deux premiers trimestres du cycle 2.2, l'échantillon a été sélectionné à partir d'une base ne contenant que les personnes ayant à ce moment complété leurs interviews pour le cycle 2.1 de l'ESCC. Cette base contenait seulement les répondants obtenus au cours des huit premiers mois de collecte du cycle 2.1. Dans un deuxième temps, c'est l'échantillon complet du cycle 2.1 de l'ESCC qui a servi de base afin de sélectionner l'échantillon pour les deux derniers trimestres. Cette base contenait tous les ménages avec des personnes de 18 ans et moins (et 71 ans et plus en Ontario) de la base aréolaire du cycle 2.1 de même que ceux de l'étude sur l'effet du mode de collecte et de la base téléphonique dans les provinces où ces bases ont été utilisées.

Pour sélectionner l'échantillon à partir de la base liste du cycle 2.1, un plan d'échantillonnage stratifié en grappes à deux degrés a été utilisé. Les provinces (et les régions en Ontario) ont été stratifiées en utilisant l'indicateur rural/urbain comme variable de stratification. Au premier degré, une liste de municipalités et/ou villes a été sélectionnée selon un plan systématique avec probabilité proportionnelle à la taille. La variable de taille utilisée pour déterminer l'importance d'une municipalité est la somme des poids des ménages répondants qui avaient des membres de moins de 18 ans au cycle 2.1. De cette façon, les municipalités qui ont beaucoup de ce type de ménages ont plus de chances d'être sélectionnées. Le poids qui a été utilisé dans le calcul de taille des municipalités est le poids A4 du cycle 2.1, c'est-à-dire le poids ménage ajusté pour la non-réponse au niveau du ménage (voir Guide du fichier de microdonnées à grande diffusion du cycle 2.1⁴⁰ pour plus de détails). Ensuite, à partir des municipalités choisies, des adresses de logements ont été sélectionnées suivant un plan systématique. Les ménages demeurant aux adresses sélectionnées font partie de l'échantillon. La probabilité de sélection d'un ménage est donc le produit des probabilités de sélection à chaque degré multiplié par la probabilité de sélection du ménage au cycle 2.1. Le poids initial est donc l'inverse de la probabilité de sélection au cycle 2.2 multiplié par le poids A4 du cycle 2.1.

Pour les unités provenant de la base téléphonique du cycle 2.1, le calcul du poids initial est légèrement différent étant donné les différences dans le plan. Effectivement, pour qu'une adresse soit sélectionnée à partir de la base téléphonique, il fallait d'abord que la municipalité dans laquelle se trouve l'adresse soit sélectionnée à partir de la l'échantillon de la partie aréolaire de la base liste. Toutes les adresses ne faisant pas partie de ces villes sont exclues de la partie téléphonique de la base liste. De plus, le poids initial doit tenir compte du fait que certaines adresses sont invalides sur la base téléphonique en raison du manque d'information relative à l'adresse. Ces adresses sont exclues de la base de sondage.

L1 - Non-réponse ménage

L'ajustement fait ici pour compenser l'effet de la non-réponse ménage est identique à celui appliqué pour la base aréolaire (ajustement A3). En plus des variables qui ont été utilisées pour la base aréolaire, la variable « Type de logement » provenant du cycle 2.1 a également pu être utilisée dans la création des classes. C'est d'ailleurs cette variable qui a été la plus significative dans la plupart des provinces pour définir les classes d'ajustement. Le facteur d'ajustement calculé à l'intérieur de chaque classe a été obtenu de la façon suivante :

40. STATISTIQUE CANADA. 2005. *L'Enquête sur la santé dans les collectivités canadiennes (cycle 2.1) - Guide*, Produit no 82M0013GPF au catalogue de Statistique Canada, Ottawa.

Somme des poids L0 pour tous les ménages

Somme des poids L0 pour tous les ménages répondants

Le poids L0 des ménages répondants a donc été multiplié par ce facteur d'ajustement pour produire le poids L1. Les ménages non-répondants sont éliminés à partir de ce point.

Noter que pour la base liste du cycle 2.1, les cas hors du champ de l'enquête ne sont pas mis de côté avant le traitement de la non-réponse. Ils sont plutôt traités comme non-répondants. Cette particularité associée à la base liste est due au fait que les logements sélectionnés étaient répondants au cycle 2.1 et par le fait même dans le champ de l'enquête. La perte de leur poids aurait donc pour effet de causer une sous-estimation du nombre de logements au pays. Les ménages qui étaient dans le champ de l'enquête au cycle 2.1 et qui sont hors du champ pour le cycle 2.2 sont contrebalancés par ceux qui étaient hors du champ de l'enquête au cycle 2.1 et dans le champ au cycle 2.2.

L2 – Création du poids-personne

Tout comme l'ajustement A4, cet ajustement permet de convertir ce qui était jusqu'à cette étape-ci un poids-ménage en un poids-personne. L'algorithme de sélection de la personne à l'intérieur du ménage étant le même que pour la base aréolaire, le calcul du facteur d'ajustement est effectué de la même façon. La seule différence vient du fait que pour les ménages avec des 18 ans et moins, seulement les moins de 18 ans avaient des chances d'être sélectionnés, sauf en Ontario où les personnes âgées de 71 ans et plus avaient également une chance de sélection. Ce facteur, multiplié par le poids L1, donne le poids L2.

L3 - Non-réponse personne

Cet ajustement est similaire à l'ajustement A5 utilisé pour la base aréolaire. Il consiste à compenser pour l'effet de la non-réponse à l'échelle de la personne. Tout comme pour A5, une approche par classes d'ajustement a été utilisée. Ces classes étaient définies à partir des variables disponibles pour toutes les personnes sélectionnées, répondantes ou non (voir A5 pour la liste des variables disponibles). Le groupe d'âge et l'état matrimonial de la personne sélectionnée, le nombre de personnes dans le ménage et la période de collecte de l'interview sont toutes des variables qui ont été utilisées dans la création des classes. Un facteur d'ajustement a donc été calculé à l'intérieur de chaque classe de la façon suivante :

Somme des poids L2 pour toutes les personnes sélectionnées

Somme des poids L2 pour toutes les personnes sélectionnées répondantes

Le poids L2 des personnes répondantes a donc été multiplié par ce facteur d'ajustement pour produire le poids L3. Les personnes non-répondantes sont éliminées à partir de ce point.

Étant donné que cet ajustement est le dernier nécessaire pour l'échantillon provenant de la base liste du cycle 2.1 de l'ESCC, le poids L3 représente donc le **poids final de la base liste du cycle 2.1**. Ce poids sera plus tard intégré au poids final de la base aréolaire (section 8.1.6) pour ensuite créer le poids final du cycle 2.2 de l'ESCC.

9.1.3 Pondération du registre de santé de l'Île-du-Prince-Édouard

Tout comme la base liste du cycle 2.1, le registre de santé de l'Î.-P.-É. est une base liste d'adresses utilisée pour cibler certains groupes d'âge afin d'atteindre les tailles d'échantillons requises pour ces groupes. Pour l'Î.-P.-É., en plus des personnes de 18 ans et moins, les groupes d'âge 19-30 et 71 ans et plus s'ajoutent aux groupes ciblés. La base de sondage formée à partir du registre de santé de l'Î.-P.-É. est en fait une liste d'adresses où demeurerait au moins une personne des groupes visés au moment de la création de la base. Tout comme pour la base liste du cycle 2.1, il est possible qu'il n'y ait plus personne dans les groupes visés pour certains ménages au moment de la collecte. Pour ces ménages, l'interview a été réalisée quand même en sélectionnant une personne ne faisant pas partie des groupes cibles.

P0 - Poids initial

Le plan d'échantillonnage utilisé pour le registre de santé de l'Î.-P.-É. est plus simple que ceux utilisés pour la base aréolaire et pour la base liste du cycle 2.1. La base est divisée en deux strates, une strate urbaine et une strate rurale. Un échantillon systématique d'adresses a été sélectionné à l'intérieur de chaque strate. Le poids initial P0 est l'inverse de la probabilité de sélection de l'adresse.

P1 - Retrait des unités hors champ

La définition des unités hors champ est la même que pour la base aréolaire. Les logements détruits ou en construction, les logements vacants, saisonniers ou secondaires sont considérés comme des cas hors champ. De la même façon que pour la base aréolaire, les logements sont retirés de l'échantillon et les unités restantes conservent le même poids qu'à l'étape précédente (poids initial P0), qui est maintenant nommé poids P1.

P2 - Non-réponse ménage

L'ajustement fait ici pour compenser l'effet de la non-réponse ménage est identique à celui appliqué pour la base aréolaire (ajustement A3). Cependant, étant donné le nombre limité de variables présentes sur le registre de l'Î.-P.-É. et la petite taille d'échantillon, aucune variable n'est ressortie dans la création des classes d'ajustement. Le même ajustement a donc été appliqué à l'ensemble de l'échantillon de la province. Le facteur d'ajustement calculé à l'échelle de la province a été obtenu de la façon suivante:

$$\frac{\text{Somme des poids } P1 \text{ pour tous les ménages}}{\text{Somme des poids } P1 \text{ pour tous les ménages répondants}}$$

Le poids P1 des ménages répondants a donc été multiplié par ce facteur d'ajustement pour produire le poids P2. Les ménages non-répondants sont éliminés à partir de ce point.

P3 - Poids-personne

Tout comme l'ajustement A4, cet ajustement permet de convertir ce qui était jusqu'à cette étape-ci un poids-ménage en un poids-personne. L'algorithme de sélection de la personne à l'intérieur du ménage étant le même que pour la base aréolaire, le calcul du facteur d'ajustement est effectué de la même façon. La seule différence vient du fait que pour les ménages avec enfants (18 ans et moins), les personnes ne faisant pas partie des groupes visés par le registre (entre 31 et 70 ans)

n'avaient aucune chance d'être sélectionnées. Ce facteur, multiplié par le poids P2, donne le poids P3.

P4 - Non-réponse personne

Cet ajustement est similaire à l'ajustement A5 utilisé pour la base aréolaire. Il consiste à compenser pour l'effet de la non-réponse à l'échelle de la personne. Tout comme pour A5, une approche par classes d'ajustement a été utilisée. Ces classes étaient définies à partir des variables disponibles pour toutes les personnes sélectionnées, répondantes ou non (voir A5 pour la liste des variables disponibles). Cependant, comme c'était le cas pour la non-réponse ménage, aucune variable n'est ressortie pour définir les classes. Il y a donc une seule classe qui est définie par la province et où le facteur d'ajustement a été calculé de la façon suivante :

$$\frac{\text{Somme des poids P3 pour toutes les personnes sélectionnées}}{\text{Somme des poids P3 pour toutes les personnes sélectionnées répondantes}}$$

Le poids P3 des personnes répondantes a donc été multiplié par ce facteur d'ajustement pour produire le poids P4. Les personnes non-répondantes sont éliminées à partir de ce point.

Étant donné que cet ajustement est le dernier nécessaire pour l'échantillon provenant du registre de santé de l'Î.-P.-É., le poids P4 représente donc le **poids final du registre de santé de l'Î.-P.-É.** Ce poids sera plus tard intégré au poids final de la base aréolaire (section 8.1.6) pour ensuite créer le poids final du cycle 2.2 de l'ESCC.

9.1.4 Pondération du registre de santé du Manitoba

Comme la base liste du cycle 2.1, le registre de santé du Manitoba est une base liste d'adresses utilisée afin de cibler les individus de 18 ans et moins. La base de sondage formée à partir du registre de santé du Manitoba est en fait une liste d'adresses où demeurait au moins une personne de 18 ans et moins au moment de la création de la base. Comme c'était le cas pour la base liste du cycle 2.1, il est possible qu'il n'y ait plus personne de 18 ans et moins pour certains ménages au moment de la collecte. Pour ces ménages, l'interview était réalisée quand même en sélectionnant une personne ne faisant pas partie du groupe cible.

Le registre de santé du Manitoba comporte quelques particularités qui font en sorte que certains ménages ne pouvaient pas être sélectionnés. En effet, certaines adresses étaient invalides ou manquantes et aucun numéro de téléphone n'était fourni pour ces adresses. Pour le registre de l'Î.-P.-É., le numéro de téléphone était utilisé afin de contacter les ménages pour lesquels l'adresse était invalide ou manquante permettant ainsi de demander l'adresse du ménage en question. Il était évidemment impossible de répéter la même procédure avec le registre de santé du Manitoba. Les adresses invalides ou manquantes de même que les villes contenant une grande proportion de ces adresses ont dû être exclues du registre du Manitoba et n'avaient donc aucune chance de faire partie de l'échantillon.

M0- Poids initial

L'échantillonnage au Manitoba a été fait par région, qui sont au nombre de quatre. Pour la région de Winnipeg, le plan de sondage comporte deux degrés. D'abord, la région est divisée en strates définies par les villes. Des grappes formées en utilisant les trois premiers caractères du code postal sont alors sélectionnées de façon proportionnelle au nombre d'adresses à l'intérieur des

grappes au premier degré. Des adresses sont ensuite sélectionnées de façon systématique à l'intérieur de chaque grappe au second degré. Pour les trois autres régions, le croisement de la ville et des trois premiers caractères du code postal a été utilisé pour former des strates. Ensuite, un échantillon systématique d'adresses valides a été sélectionné à l'intérieur de chacune de ces strates. Le poids initial est l'inverse de la probabilité de sélection de l'adresse.

M1 – Adresses invalides

Tel que mentionné ci-haut, le registre de santé du Manitoba contient des adresses auxquelles il aurait été impossible pour l'intervieweur de se rendre par manque d'information. La majorité de ces adresses sont en fait des numéros de boîte postale sans numéro et nom de rue. Puisque ces adresses ne pouvaient pas être sélectionnées, un ajustement a été appliqué aux poids afin de compenser pour leur perte. L'ajustement est appliqué au niveau des grappes qui sont définies par les trois premiers caractères du code postal. L'ajustement M1 appliqué à l'intérieur des grappes est le suivant :

$$\frac{\text{Nombre total d'adresses dans la grappe}}{\text{Nombre d'adresses valides dans la grappe}}$$

Le poids M1 est obtenu en multipliant le poids M0 par l'ajustement M1.

M2 - Villes inéligibles

Afin d'éviter d'envoyer des intervieweurs dans des villes pour un très petit nombre de cas, les villes pour lesquelles le nombre d'adresses valides était trop petit ou pour lesquelles la proportion d'adresses valides était trop faible ont été exclues de la base. Pour conserver la représentativité au niveau de la région, le poids des villes inéligibles a été redistribué aux villes éligibles. L'ajustement M2 est calculé de la façon suivante à l'intérieur de chaque région :

$$\frac{\text{Nombre total d'adresses dans la région}}{\text{Nombre d'adresses dans les villes éligibles de la région}}$$

Le poids M1, multiplié par l'ajustement M2, produit le poids M2.

M3 – Retrait des unités hors champ

La définition des unités hors champ est la même que pour la base aréolaire. Les logements détruits ou en construction, les logements vacants, saisonniers ou secondaires sont également considérés comme des cas hors champ. De la même façon que pour la base aréolaire, les logements sont retirés de l'échantillon et les unités restantes conservent le même poids qu'à l'étape précédente (poids M2), qui est maintenant nommé poids M3.

M4 - Non-réponse ménage

L'ajustement fait ici pour compenser l'effet de la non-réponse ménage est identique à celui appliqué pour la base aréolaire (ajustement A3). Contrairement au registre de l'Î.-P.-É., certaines variables ont pu être utilisées pour la formation des classes d'ajustement. En effet, le nombre d'enfants et d'adultes dans le ménage (selon l'information de la base de sondage) et la période de

collecte ont été utilisés pour former les classes de certaines régions. Le facteur d'ajustement calculé à l'intérieur de chaque classe a été obtenu de la façon suivante :

$$\frac{\text{Somme des poids M3 pour tous les ménages}}{\text{Somme des poids M3 pour tous les ménages répondants}}$$

Le poids M3 des ménages répondants a donc été multiplié par ce facteur d'ajustement pour produire le poids M4. Les ménages non-répondants sont éliminés à partir de ce point.

M5 - Poids-personne

Tout comme l'ajustement A4, cet ajustement permet de convertir ce qui était jusqu'à cette étape-ci un poids-ménage en un poids-personne. L'algorithme de sélection de la personne à l'intérieur du ménage étant le même que pour la base aréolaire, le calcul du facteur d'ajustement est effectué de la même façon. Ce facteur, multiplié par le poids M4, donne le poids M5.

M6 - Non-réponse personne

Cet ajustement est similaire à l'ajustement A5 utilisé pour la base aréolaire. Il consiste à compenser pour l'effet de la non-réponse à l'échelle de la personne. Tout comme pour A5, une approche par classes d'ajustement a été utilisée. Ces classes étaient définies à partir des variables disponibles pour toutes les personnes sélectionnées, répondantes ou non (voir A5 pour la liste des variables disponibles). Il y a seulement une région parmi les quatre pour laquelle des variables étaient significatives pour expliquer la non-réponse. Ces variables sont le nombre de personnes dans le ménage et le nombre de personnes de 12 ans et moins dans le ménage. Pour les autres régions, l'ajustement a été appliqué globalement à l'échelle de la région. Les facteurs d'ajustement ont été calculés à l'intérieur de chaque classe de la façon suivante :

$$\frac{\text{Somme des poids M5 pour toutes les personnes sélectionnées}}{\text{Somme des poids M5 pour toutes les personnes sélectionnées répondantes}}$$

Le poids M5 des personnes répondantes a donc été multiplié par ce facteur d'ajustement pour produire le poids M6. Les personnes non-répondantes sont éliminées à partir de ce point.

Étant donné que cet ajustement est le dernier nécessaire pour l'échantillon provenant du registre de santé du Manitoba, le poids M6 représente donc le **poids final du registre de santé du Manitoba**. Ce poids sera plus tard intégré au poids final de la base aréolaire (section 8.1.6) pour ensuite créer le poids final du cycle 2.2 de l'ESCC.

9.1.5 Pondération des cas provenant de la base supplémentaire autochtone

Un échantillon supplémentaire s'est ajouté à l'échantillon principal afin d'obtenir des tailles minimales d'autochtones non métis vivant à l'extérieur des réserves. Pour pouvoir atteindre ces tailles, il fallait d'abord être en mesure d'identifier les ménages susceptibles de compter des autochtones. Pour y arriver, les répondants du cycle 2.1 de l'ESCC ont été utilisés de nouveau. Tous les ménages pour lesquels le répondant s'était déclaré autochtone au cycle 2.1 font partie de la base supplémentaire autochtone.

S0 – Poids initial

L'échantillon supplémentaire d'autochtones est en fait un recensement de tous les ménages de la base aréolaire du cycle 2.1 pour lesquels le répondant s'était identifié comme étant autochtone. Le poids initial S0 est égal au poids ménage compensé pour la non-réponse (poids A4) du cycle 2.1 (voir guide du FMGD du cycle 2.1).

S1 – Non-réponse ménage

L'ajustement fait ici pour compenser l'effet de la non-réponse ménage est semblable à celui appliqué aux unités des autres bases. Une des différences provient du fait que les classes d'ajustement sont créées pour certains regroupements de provinces plutôt qu'à l'échelle de la province étant donné les petites tailles d'échantillon impliquées pour certaines provinces. Les provinces ont été regroupées de la façon suivante : Est (T.-N.-L., Î.-P.-É., N.-É. et N.-B.), Centre (Qc et Ont.) et Ouest (Alb. et C.-B.). Les provinces du Manitoba et de la Saskatchewan n'ont pas nécessité de regroupement. La seule variable qui s'est révélée significative pour la création des classes est la période de collecte. Le facteur d'ajustement calculé à l'intérieur de chaque classe a été obtenu de la façon suivante :

$$\frac{\text{Somme des poids S0 pour tous les ménages}}{\text{Somme des poids S0 pour tous les ménages répondants}}$$

Le poids S0 des ménages répondants a donc été multiplié par ce facteur d'ajustement pour produire le poids S1. Les ménages non-répondants sont éliminés à partir de ce point.

Comme pour la base liste du cycle 2.1, les cas hors du champ de l'enquête ne sont pas mis de côté avant le traitement de la non-réponse mais sont plutôt traités comme non-répondants et ce pour les mêmes raisons (voir L1).

S2- Poids-personne

Comme pour les autres bases, cet ajustement permet de convertir ce qui était jusqu'à cette étape-ci un poids-ménage en un poids-personne. Un facteur d'ajustement égal à l'inverse de la probabilité de sélection de la personne à l'intérieur de son ménage est appliqué. On obtient le poids S2 en multipliant le poids S1 par ce facteur. Noter que la personne sélectionnée à l'intérieur du ménage n'est pas nécessairement autochtone. Il est possible que la personne sélectionnée soit différente de celle qui l'avait été lors du cycle 2.1 et que cette personne ne soit pas autochtone. Cette particularité est prise en considération lors de l'intégration de cet échantillon à celui des autres bases (en I2). Ce facteur, multiplié par le poids S1, donne le poids S2.

S3 – Non-réponse personne

Cet ajustement est similaire aux ajustements appliqués aux unités provenant des autres bases. Il consiste à compenser pour l'effet de la non-réponse à l'échelle de la personne. Les classes d'ajustement ont été développées au niveau des mêmes regroupements de provinces que pour la non-réponse ménage (voir S1). Elles ont été définies à partir des variables disponibles pour toutes les personnes sélectionnées, répondantes ou non (voir A5 pour la liste des variables disponibles). C'est le sexe qui a été la variable la plus utilisée dans la création des classes. Un facteur d'ajustement a été calculé à l'intérieur de chaque classe de la façon suivante :

$$\frac{\text{Somme des poids S2 pour toutes les personnes sélectionnées}}{\text{Somme des poids S2 pour toutes les personnes sélectionnées répondantes}}$$

Le poids S2 des personnes répondantes a donc été multiplié par ce facteur d'ajustement pour produire le poids S3. Les personnes non-répondantes sont éliminées à partir de cette étape.

Étant donné que cet ajustement est le dernier nécessaire pour l'échantillon provenant de la base supplémentaire autochtone, le poids S3 représente donc le **poids final de la base supplémentaire autochtone**. Ce poids sera plus tard intégré au poids final des autres bases (section 8.1.7) pour ensuite créer le poids final du cycle 2.2 de l'ESCC.

9.1.6 Intégration de la base aréolaire avec les bases listes (I1)

Cette étape consiste à intégrer les poids finaux de l'échantillon de la base aréolaire avec les poids finaux des échantillons des bases listes (base liste du cycle 2.1, registre de santé de l'Î.-P.-É. et registre de santé du Manitoba). Rappelons que la base liste du cycle 2.1 n'est pas utilisée dans les provinces où les registres de santé sont utilisés. Il y a donc une seule base liste utilisée par province. Une méthode d'intégration⁴¹ est appliquée dans le but de créer un seul poids à partir des poids précédents. Un facteur d'ajustement, compris entre 0 et 1, est déterminé de façon à représenter l'importance relative de chaque échantillon dans l'échantillon total. Cette importance relative est mesurée en termes de taille d'échantillon et d'effet de plan. Plus la proportion d'échantillon qu'une base représente dans l'échantillon total est grande, plus grande sera son importance relative dans l'échantillon total. Pour ce qui est de l'effet de plan, l'importance relative sera plus grande pour les unités provenant de la base dont l'effet de plan (ou la variabilité) est plus petit(e).

La stratégie d'intégration a dû être adaptée pour le cycle 2.2 étant donné les particularités associées aux bases listes. En effet, puisque la couverture des bases listes en terme de types de ménages est différente de celle de la base aréolaire, l'intégration doit en tenir compte. C'est pour cette raison que l'intégration se fait séparément pour des types de répondants différents. C'est la composition du ménage en termes de groupes d'âge qui définit le type de répondant. Le tableau I compare la couverture de la base aréolaire avec celle des bases listes. Noter que pour alléger le texte, le terme adulte est utilisé afin de représenter les personnes de plus de 18 ans alors que le terme enfant est utilisé pour représenter les personnes de 18 ans et moins.

Tableau 9.1 Couverture des bases aréolaire et listes selon le type de répondant

Types de répondant *	Couverture	
	Base aréolaire	Bases listes
A) Adultes provenant de ménages comptant au moins un enfant	Complète	Nulle
B) Adultes provenant de ménages sans enfant	Complète	Faible
C) Enfants	Complète	Bonne

* D'autres types de répondants s'ajoutent en Ontario et à l'Î.-P.-É.. Il s'agit des personnes âgées de 71 ans et plus de ménages avec enfants. De plus, pour l'Î.-P.-É., s'ajoutent également les

41. SKINNER, C.J. et J.N.K. Rao. 1996. « Estimation in Dual Frame Surveys with Complex Designs ». *Journal of the American Statistical Association*, 91, 433, 349-356.

personnes âgées entre 19 et 30 ans de ménages avec enfants. Tous ces types de répondants ont une couverture complète sur la base aréolaire et une bonne couverture sur les bases listes.

A) Intégration des adultes provenant de ménages avec enfants

Puisqu'il est impossible d'obtenir ce type de répondant à partir des bases listes, aucun ajustement n'est apporté aux poids de ces répondants. Ils conservent le même poids que le poids final de la base aréolaire, c'est-à-dire le poids A5, que l'on appelle maintenant I1.

B) Intégration des adultes provenant de ménages sans enfant

Ce type de répondant se retrouve sur les bases listes, mais dans une proportion beaucoup plus faible que sur la base aréolaire. Effectivement, étant donné que les bases listes étaient composées uniquement de ménages avec enfants au moment de leur création, la probabilité d'obtenir un ménage sans enfant est plus faible que sur la base aréolaire. De plus, la somme de tous les poids de niveau ménage des bases listes devrait représenter le nombre de ménages avec enfants au moment de la création des bases. Cependant, puisqu'on obtient d'autres types de ménages au moment de la collecte, une partie du poids qui devait représenter les ménages avec enfants est transférée aux ménages sans enfant. Cela implique que la somme des poids de chacun des types de ménage sous-représente la population totale de ces ménages. Afin que les adultes provenant de ménages sans enfants soient bien représentés par les bases listes, un ajustement aux poids est appliqué avant l'intégration. La base aréolaire est utilisée à titre de référence pour calculer l'ajustement suivant à l'intérieur de chaque province (ou région en Ontario et au Manitoba) :

$$\frac{\text{Somme des poids A6 des adultes de ménages sans enfant}}{\text{Somme des poids L3, P4 ou M6 (selon la province) des adultes de ménages sans enfant}}$$

Suite à cet ajustement, la somme des poids des adultes provenant de ménages sans enfant est la même pour les deux bases. Les cas des deux bases sont ensuite intégrés. Pour obtenir le facteur d'ajustement d'intégration, on calcule d'abord un facteur α_B pour chaque province (ou région en Ontario et au Manitoba), obtenu de la façon suivante :

$$\alpha_B = \frac{n_{AB}}{R} \left/ \left(\frac{n_{AB}}{R} + n_{CB} \right) \right.$$

où n_{AB} et n_{CB} représentent respectivement le nombre de répondants adultes provenant de ménages sans enfant des bases aréolaire et liste, alors que R représente le rapport médian des effets de plan estimés pour chacune des deux bases. Le poids des unités de la base aréolaire est multiplié par ce facteur α_B , alors que le poids des unités des bases listes est multiplié par $1 - \alpha_B$. Le produit du facteur d'ajustement dérivé ici, par le poids final calculé auparavant (A5, L3, P4 ou M6 dépendant de quelle base provient l'unité), procure le poids intégré I1.

C) Enfants

C'est le type de répondant visé par les bases listes. Ils sont donc présents sur ces bases dans une forte proportion. Cependant, puisqu'une partie du poids est transférée aux adultes de ménages sans enfant, les bases listes ne couvrent pas la totalité des enfants. Un ajustement semblable à celui calculé en B) est appliqué aux enfants des bases listes. Cet ajustement est

calculé à l'intérieur de chaque province (ou région en Ontario et au Manitoba) de la façon suivante :

$$\frac{\text{Somme des poids A6 des enfants}}{\text{Somme des poids L3, P4 ou M6 (selon la province) des enfants}}$$

La somme des poids des enfants est donc la même sur les bases listes que sur la base aréolaire. Les enfants provenant des deux bases sont ensuite intégrés. Comme c'est le cas pour les adultes de ménages sans enfant, un facteur α_C est obtenu de la façon suivante :

$$\alpha_C = \frac{n_{AC}}{R} \left/ \left(\frac{n_{AC}}{R} + n_{CC} \right) \right.$$

où n_{AC} et n_{CC} représentent respectivement le nombre d'enfants répondants des bases aréolaire et liste, alors que R représente le rapport médian des effets de plan estimés pour chacune des deux bases. Le poids des unités de la base aréolaire est multiplié par ce facteur α_C , alors que le poids des unités des bases listes est multiplié par $1 - \alpha_C$. Le produit du facteur d'ajustement dérivé ici, par le poids final calculé auparavant (A5, L3, P4 ou M6 dépendant de quelle base provient l'unité), procure le poids intégré I1.

Noter que pour les types de répondants qui s'ajoutent en Ontario et à l'Î.-P.-É., le traitement est le même que ceux présentés en B) et en C). La seule différence pour ces deux provinces, c'est que les répondants sont fractionnés en un nombre plus grand de types de répondants.

9.1.7 Intégration de la base supplémentaire autochtone avec les autres bases (I2)

Après avoir intégré la base aréolaire avec les bases listes, la dernière étape afin d'obtenir un seul échantillon final est l'intégration des cas provenant de la base supplémentaire autochtone aux cas des autres bases. Le principe est sensiblement le même que pour l'intégration I1. Les répondants sont divisés en deux catégories : les autochtones et les non-autochtones. Le poids des cas de la base autochtones est ajusté en utilisant les comptes de la base aréolaire et on intègre ensuite par type de répondant, autochtone ou non. Un facteur d'intégration est calculé basé sur l'importance relative de chaque base en termes de tailles d'échantillons. Le facteur suivant est calculé à l'intérieur des provinces (ou régions en Ontario et au Manitoba) pour chaque type de répondant :

$$\alpha_S = \frac{n_I}{n_I + n_S}$$

où n_I et n_S représentent respectivement le nombre de répondants (autochtones ou non-autochtones selon le cas) des bases intégrées en I1 et celui de la base autochtone. Le poids des unités des bases intégrées est multiplié par ce facteur α_S , alors que le poids des unités de la base autochtone est multiplié par $1 - \alpha_S$. Le produit du facteur d'ajustement dérivé ici, par le poids final calculé auparavant (I1 ou S3 dépendant de quelle base provient l'unité), procure le poids I2.

9.1.8 Effet saisonnier et du jour de la semaine (I3)

L'ESCC (cycle 2.2) avait initialement planifié répartir la collecte des données également sur les douze mois de l'année de référence de l'enquête afin de contrôler entre autres l'effet saisonnier des données recueillies. Certains événements ont toutefois affecté ce plan, de sorte qu'un ajustement additionnel a dû être ajouté pour assurer qu'il n'y ait pas d'effet saisonnier dans les estimations produites à l'aide des données du cycle 2.2 de l'ESCC. De plus, étant donné que le jour de la semaine a un impact sur la consommation de certains aliments, une attention particulière devait être portée au jour où se déroulait l'interview. En effet, si par exemple la consommation de certains aliments est beaucoup plus élevée durant le week-end et qu'une plus grande proportion des interviews sont conduites le week-end, la consommation de ces aliments sera inévitablement surestimée. L'ajustement appliqué en I3 a été fait de façon à faire diminuer l'effet de la saison et du jour de la semaine. L'ajustement fait en sorte que la somme des poids des unités interviewées lors d'une des quatre saisons, représente exactement 25 % de la somme des poids de l'échantillon total de la province (ou de la région en Ontario et au Manitoba). De plus, l'ajustement fait en sorte qu'à l'intérieur d'une saison, la distribution des poids pour les jours de la semaine représente 4/7 des poids de la saison et que les poids pour le week-end représentent 3/7 des poids de la saison parce que les fins de semaine ont été définies comme incluant vendredi, samedi et dimanche. Bref, après l'application de cet ajustement, la portion d'échantillon interviewée à chaque saison représente 25 % de la population totale de chaque province (ou région en Ontario et au Manitoba) et la portion d'échantillon interviewée un jour de semaine représente 4/7 de cette même population.

Les quatre saisons définies pour l'ESCC sont les périodes couvrant octobre à décembre, janvier à mars, avril à juin, puis juillet à septembre. Pour ce qui est de l'indicateur semaine/fin de semaine, les interviews qui ont été conduites de mardi à vendredi faisaient partie de la semaine alors que celles qui ont été conduites de samedi à lundi faisaient partie du week-end. Le week-end est défini de cette façon puisque les interviews sont conduites la journée suivant le rappel de 24 heures et la consommation déclarée par le répondant se rapporte donc à la journée précédant celle de l'interview. (La fin de semaine est définie ainsi car l'apport rapporté était associé avec la journée précédente, donc une interview le lundi demandant les aliments consommés le dimanche par exemple). Le facteur d'ajustement utilisé pour contrôler l'effet saisonnier de même que l'effet du jour de la semaine d'une personne interviewée lors de la saison S est défini comme suit :

$$\left\{ \begin{array}{l} \frac{3}{7} x \frac{\text{Somme des poids I2 pour l'échantillon total avec } I = 0}{4 \times \text{somme des poids I2 de l'échantillon interviewé avec } I = 0 \text{ lors de la saison } S} \quad \text{si } I = 0 \\ \frac{4}{7} x \frac{\text{Somme des poids I2 pour l'échantillon total avec } I = 1}{4 \times \text{somme des poids I2 de l'échantillon interviewé avec } I = 1 \text{ lors de la saison } S} \quad \text{si } I = 1 \end{array} \right.$$

où I = indicateur de jour de semaine (=1 pour jour de la semaine et 0 sinon)

Cet ajustement appliqué au poids I2 permet d'obtenir le poids I3.

Notez que suite à la série d'ajustements appliqués sur les poids, il est possible que certaines unités se retrouvent avec des poids se démarquant des autres poids de leur province au point même de devenir aberrants. Certains répondants peuvent effectivement représenter une proportion anormalement élevée de leur province et ainsi influencer fortement les estimations de

ces provinces. 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. De la même façon, les poids qui sont anormalement bas sont ajustés à la hausse.

9.1.9 Poststratification (I4)

La dernière étape nécessaire afin d'obtenir le poids final du cycle 2.2 de l'ESCC est la poststratification. La poststratification est appliquée afin d'assurer que la somme des poids finaux corresponde aux estimations de populations définies à l'échelle des provinces et des régions pour l'Ontario et le Manitoba pour chacun des 15 groupes d'âge-sexe d'intérêt. Les groupes d'intérêt sont définis par les groupes d'âge <1, 1-3, 4-8, sans distinction au niveau du sexe de même que par les groupes d'âges 9-13, 14-18, 19-30, 31-50, 51-70, 71+, pour chacun des deux sexes.

Les estimations de population utilisées sont basées sur les comptes du Recensement de 2001, de même que sur les comptes de naissance, décès, immigration et émigration. La moyenne des estimations mensuelles de 2004 pour chacun des croisements province-âge-sexe (région-âge-sexe en Ontario et au Manitoba) a été retenue pour calculer la poststratification. Le poids I3 a donc été ajusté afin d'obtenir le poids final I4, à l'aide du facteur d'ajustement I4 défini comme suit :

$$\frac{\text{Estimation de population pour le groupe province - âge - sexe du répondant}}{\text{Somme des poids I3 pour le groupe province - âge - sexe du répondant}}$$

Le poids I4 correspond au **poids final du cycle 2.2 de l'ESCC** que l'on retrouve dans le fichier de données portant le nom de variable WTSD_M.

9.2 Pondération du sous-échantillon ayant des mesures de tailles et de poids

Au départ, les mesures de taille et de poids devaient être prises pour tous les répondants de deux ans et plus. Cependant, certains événements ont fait en sorte qu'il a été impossible de prendre ces mesures pour environ 40 pour cent de l'échantillon. Le refus du répondant d'être mesuré, des problèmes de matériel de mesure, et les interviews réalisées au téléphone sont toutes des raisons pour lesquelles les mesures n'ont pas pu être prises sur le répondant. Étant donné le nombre élevé de répondants sans mesures de taille et de poids, il a été convenu de créer un nouveau poids appliqué seulement au sous-échantillon de répondants pour lequel les mesures ont été obtenues.

MHW0 - Poids initial

La stratégie de pondération pour créer le poids sur les mesures de taille et de poids est exactement la même que pour la pondération régulière jusqu'à l'étape I3. Le poids initial MHW0 est égal au poids intégré avant la poststratification, c'est-à-dire le poids I3.

MHW1 – Poids pour les mesures de tailles et de poids

C'est à cette étape que les répondants pour lesquels il n'y a pas de mesure de taille et/ou de poids sont retirés de l'échantillon. Le poids de ces répondants est redistribué aux répondants pour lesquels les mesures sont disponibles. Il faut noter que le poids WTSD_MHW a été créé afin d'analyser les variables se rapportant à l'IMC et que le poids WTSD_MHW est manquant pour quelques répondantes pour lesquelles les mesures de taille et de poids sont quand même

disponibles. L'IMC n'a pas été calculé pour ces répondantes étant donné qu'il est impossible de savoir si ces femmes sont enceintes ou non et que l'IMC n'est pas calculé pour les femmes enceintes.

Afin de réduire le biais qui aurait pu être introduit par la perte des répondants sans mesures, la redistribution des poids s'est faite en se basant sur des classes de répondants homogènes quant à la probabilité d'avoir les mesures ou non. Ces classes ont été créées en utilisant la même technique que pour le traitement de la non-réponse, c'est-à-dire en utilisant l'algorithme CHAID disponible dans le logiciel Knowledge Seeker. Contrairement aux étapes de traitement de la non-réponse rencontrées jusqu'ici, toutes les variables du cycle 2.2 étaient disponibles pour les répondants, avec ou sans mesures. Elles pouvaient donc toutes être utilisées dans la création des classes. L'âge, le sexe, le revenu, le niveau d'éducation et le nombre d'aliments déclarés pour le rappel de 24 heures sont les variables qui ont été les plus utilisées dans la création des classes. Plusieurs autres variables corrélées avec les variables d'intérêt ont aussi été utilisées. Les classes ont été créées indépendamment pour chaque province (ou région en Ontario et au Manitoba) et le facteur d'ajustement suivant a été calculé à l'intérieur de chaque classe :

$$\frac{\text{Somme des poids MHW0 pour tous les répondants}}{\text{Somme des poids MHW0 pour tous les répondants avec mesures personnelles}}$$

Le poids MHW0 des répondants avec mesures personnelles a donc été multiplié par ce facteur d'ajustement pour produire le poids MHW1. Les répondants pour lesquels on n'a pas de mesure personnelles de taille et de poids sont éliminés à partir de ce point.

MHW2- Poststratification

La poststratification est exactement la même que celle appliquée dans la pondération régulière. Elle est appliquée au niveau des mêmes régions et des mêmes groupes d'âge-sexe qu'en I4 sauf pour les groupes <1 an et 1-3 ans qui sont modifiés étant donné qu'on ne prenait pas de mesures sur les répondants âgés de moins de 2 ans. Pour conserver la confidentialité, le poids des femmes enceintes a ensuite été masqué sur le FMGD.

Le poids MHW2 correspond au **poids final du sous-échantillon pour lequel on a des mesures de taille et de poids pour le cycle 2.2 de l'ESCC** que l'on retrouve dans le fichier de données portant le nom de variable WTSD_MHW.

10. Qualité des données

10.1 Limites des données

10.1.1 Limites du rappel alimentaire de 24 heures

Le volet rappel alimentaire de 24 heures du questionnaire de l'ESCC 2.2 recueille des renseignements sur les types et les quantités d'aliments consommés par les répondants durant la période de 24 heures précédant l'interview, de minuit à minuit. Ce volet utilise une méthode d'interview assistée par ordinateur appelée Automated Multiple Pass Method (AMPM), élaborée initialement par le Food Surveys Research Group (FSRG) du US Department of Agriculture (USDA). De concert avec Santé Canada, l'AMPM a été adaptée pour tenir compte des différences entre les aliments à la disposition des consommateurs canadiens et des consommateurs américains, ainsi qu'entre les préparations alimentaires et les aliments ethniques consommés dans les deux pays.

Les données obtenues des interviews du rappel alimentaire de 24 heures sont utilisées pour produire des estimations des apports en énergie, nutriments et autres composants alimentaires provenant des aliments et boissons consommés par les répondants. Toutefois, les données sur les aliments consommés au cours des dernières 24 heures peuvent être entachées d'erreurs, ce qui pourrait compromettre l'exactitude des données sur les aliments. Selon le rapport de Santé Canada intitulé *The Canadian Community Health Survey (CCHS) 2.2, Nutrition Focus: A Guide to Accessing and Interpreting the Data* (July 2006), les deux principales sources d'erreurs sont liées aux deux éléments suivants⁴² :

1) les renseignements sur les aliments fournis par les participants à l'enquête. Par exemple, les aliments consommés ont peut-être été sous-estimés parce que les répondants ont oublié de les mentionner ou qu'ils étaient gênés de déclarer certains types d'aliments perçus comme étant moins bons pour la santé, comme les desserts, les matières grasses et l'alcool. Dans certains cas, la quantité déclarée n'était peut-être pas exacte, certains répondants ayant déclaré la quantité servie mais non celle réellement consommée⁴³.

2) le système de base de données sur la nutrition qui est utilisé aux fins du codage des aliments déclarés et du calcul des nutriments contenus dans les aliments. Le Nutrition Survey System (NSS) est le système de base de données sur les aliments et les nutriments utilisé aux fins du cycle 2.2 de l'ESCC pour coder les aliments et calculer les apports nutritionnels des répondants. Le NSS utilise plusieurs bases de données canadiennes et américaines sur les aliments et il est constamment mis à jour de manière à inclure les données nouvelles et révisées sur la composition des aliments au fur et à mesure qu'elles deviennent disponibles au Canada. Toutefois, cette base de données ne comprend peut-être pas de données sur tous les nutriments et sur tous les aliments consommés au Canada.

42. Health Canada. *Canadian Community Health Survey (CCHS) 2.2, Nutrition Focus: A Guide to Accessing and Interpreting the Data*. July 2006. Available at: http://www.hc-sc.gc.ca/fn-an/surveill/nutrition/commun/index_e.html.

43. Les intervieweurs ont reçu pour instruction d'inscrire la quantité de l'aliment en question consommée et non servie. Ils ont été formés pour poser les questions supplémentaires, en procédant avec soin, lorsqu'un répondant déclare avoir consommé des sandwiches, des morceaux de poulet et des viandes comme des côtelettes de porc. En effet, les répondants peuvent s'appliquer tellement à décrire la taille de la portion servie qu'ils pourraient oublier qu'ils ne l'ont pas toute mangée.

Pour plus de renseignements sur les facteurs qui influent sur l'exactitude des données du rappel alimentaire de 24 heures sur les aliments consommés, consulter le rapport de Santé Canada à l'adresse http://www.hc-sc.gc.ca/fn-an/surveill/nutrition/commun/index_f.html.

Un résumé des sources éventuelles d'erreurs est fourni ci-dessous.

Tableau 10.1 Facteurs influençant l'exactitude des données du rappel alimentaire de 24 heures

1) L'exactitude de l'information alimentaire recueillie par l'application informatisée pour le rappel de 24 heures

- L'application d'interview informatisée appelée « Automated Multiple Pass Method (AMPM) » utilise plusieurs méthodes afin d'aider les répondants à se rappeler leur consommation d'aliments et boissons au cours de 24 heures.
- Les interviews par procuration pour les nouveau-nés et jeunes enfants pourront avoir une incidence sur l'exactitude des données.
- Les adultes sous-déclarent leurs apports d'aliments systématiquement – cela peut résulter dans la sous déclaration de leurs apports des nutriments.
- La sous déclaration n'est pas unanime parmi les individus et les aliments – cela pourra avoir une implication concernant l'interprétation des données de l'apport des nutriments.
- Plusieurs méthodes peuvent être utilisées pour comprendre les implications de la sous déclaration.

2) L'exactitude des bases de données des nutriments

- Il se peut que les bases de données ne soient pas complètes pour tous les nutriments et n'incluent pas tous les aliments.
- Les erreurs aléatoires peuvent être produits à cause du désaccord entre la composition de nutriments des différents types du même aliment ; ou à cause des conditions de la préparation et du traitement de l'aliment, qu'il soit à la maison ou ailleurs.

Source: Adapté de *L'Enquête sur la santé dans les collectivités canadiennes, cycle 2.2, Nutrition (2004): Guide d'accès et d'interprétation des données*. Santé Canada: Juillet 2006. Disponible à : http://www.hc-sc.gc.ca/fn-an/surveill/nutrition/commun/index_f.html

10.1.2 Limites des données sur les suppléments de vitamines et de minéraux

Le module sur les suppléments de vitamines et de minéraux a permis de recueillir des renseignements sur la consommation de suppléments par le répondant au cours du dernier mois. Les données présentent certaines lacunes, tel qu'indiqué ci-dessous :

1. La consommation de suppléments est basée sur la consommation au cours du mois précédent dont le répondant se remémore. Il se peut que le répondant n'ait pu mémoriser tous les suppléments pris ainsi que de la quantité prise. Cela est vrai tout particulièrement pour les suppléments consommés seulement à l'occasion.
2. Les données dans la BDPP, qui a été utilisée pour calculer la teneur en nutriments des suppléments déclarés, ont été recueillies à un moment spécifique (septembre 2003) et ne portent que sur les suppléments de vitamines et de minéraux actifs. En outre, les données dans la base de données ne portaient que sur les suppléments disponibles au Canada. Par conséquent, il se peut qu'un répondant ait déclaré un supplément qui soit n'était plus actif, soit n'était pas disponible

au Canada. De même, étant donné que la collecte s'est échelonnée sur un an, certains répondants ont déclaré consommer des suppléments qui n'étaient pas inclus dans la version de septembre 2003 de la BDPP mais qui ont été ajoutés dans des versions subséquentes de cette base de données. En pareil cas, des suppléments similaires sur lesquels des données étaient disponibles dans la BDPP et(ou) le NSS ont été substitués.

3. Les données sur les suppléments portent sur la consommation au cours du dernier mois, tandis que les données sur la consommation des aliments portent sur la période de 24 heures de la journée précédente. Cette différence, en ce qui concerne la période de référence entre les deux types de données sur les apports, présentera des défis au moment de combiner les données sur les aliments et celles sur les suppléments (pour plus de détails, voir l'annexe).

10.2 Taux de réponse

Au total et après avoir retiré les unités hors du champ de l'enquête, 45 889 ménages ont été sélectionnés pour participer à l'ESCC (Cycle 2.2). De ce nombre, 38 725 ont accepté de participer à l'enquête ce qui résulte en un taux de réponse à l'échelle du ménage de 84,4 %. Parmi ces ménages répondants, 38 725 personnes ont été sélectionnées (une personne par ménage) pour participer à l'enquête parmi lesquelles 35 107 ont accepté de le faire ce qui résulte en un taux de réponse à l'échelle de la personne de 90,7 %. À l'échelle canadienne, un taux de réponse combiné de 76,5 % a donc été observé à l'ESCC (Cycle 2.2). Le tableau 10.2 donne les taux de réponse combinés ainsi que l'information pertinente au calcul de ceux-ci pour chaque province, région ou région regroupée.

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

Taux de réponse à l'échelle du ménage

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

Taux de réponse à l'échelle de la personne

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

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

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

$$\text{HHRR} = \frac{21\,382 + 17\,343}{25\,436 + 20\,453} = \frac{38\,725}{45\,889} = 0,844$$

$$\text{PPRR} = \frac{19\,312 + 15\,795}{21\,382 + 17\,343} = \frac{35\,107}{38\,725} = 0,907$$

$$\begin{aligned}\text{Taux de réponse combiné} &= 0,844 \times 0,907 \\ &= 0,765 \\ &= \mathbf{76,5 \%}.\end{aligned}$$

Table 10.2 Tableau 10.2		Area frame / Base aréolaire							Other frames / Autres bases							All cases / Tous les cas
Prov.	Health Region	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	Combined resp. rates
Prov.	Région socio-sanitaire	# mén. cibles	# mén. rép.	Taux de rép. mén.	# pers. sélect.	# rép.	Taux de rép. pers.	Taux de rép.	# mén. cibles	# mén. rép.	Taux de rép. mén.	# pers. sélect.	# rép.	Taux de rép. pers.	Taux de rép.	Taux de rép. combiné
CA	Total	25436	21382	84.1	21382	19312	90.3	75.9	20453	17343	84.8	17343	15795	91.1	77.2	76.5
NL	Total	1405	1273	90.6	1273	1189	93.4	84.6	677	604	89.2	604	545	90.2	80.5	83.3
PE	Total	769	690	89.7	690	618	89.6	80.4	1037	896	86.4	896	812	90.6	78.3	79.2
NS	Total	1275	1110	87.1	1110	1002	90.3	78.6	894	798	89.3	798	703	88.1	78.6	78.6
NB	Total	1414	1227	86.8	1227	1060	86.4	75.0	742	646	87.1	646	573	88.7	77.2	75.7
QC	Total	3741	3080	82.3	3080	2799	90.9	74.8	2569	2166	84.3	2166	1981	91.5	77.1	75.8
ON	Total	7366	5947	80.7	5947	5273	88.7	71.6	7647	6244	81.7	6244	5648	90.5	73.9	72.7
	35901	1025	859	83.8	859	783	91.2	76.4	1170	963	82.3	963	877	91.1	75.0	75.6
	35902	963	759	78.8	759	645	85.0	67.0	1017	811	79.7	811	706	87.1	69.4	68.2
	35903	1129	940	83.3	940	834	88.7	73.9	1135	961	84.7	961	878	91.4	77.4	75.6
	35904	1063	853	80.2	853	754	88.4	70.9	1118	912	81.6	912	837	91.8	74.9	72.9
	35905	975	747	76.6	747	662	88.6	67.9	1026	817	79.6	817	721	88.2	70.3	69.1
	35906	1044	827	79.2	827	727	87.9	69.6	1096	878	80.1	878	799	91.0	72.9	71.3
	35907	1167	962	82.4	962	868	90.2	74.4	1085	902	83.1	902	830	92.0	76.5	75.4
MA	Total	2694	2374	88.1	2374	2211	93.1	82.1	2375	2112	88.9	2112	1983	93.9	83.5	82.7
	46901*	705	628	89.1	628	583	92.8	82.7	1424	1272	89.3	1272	1193	93.8	83.8	83.4
	46903	846	760	89.8	760	710	93.4	83.9	555	488	87.9	488	462	94.7	83.2	83.7
	46904	1143	986	86.3	986	918	93.1	80.3	396	352	88.9	352	328	93.2	82.8	81.0
SK	Total	1582	1331	84.1	1331	1195	89.8	75.5	1064	930	87.4	930	846	91.0	79.5	77.1
AB	Total	2472	2083	84.3	2083	1894	90.9	76.6	1433	1232	86.0	1232	1127	91.5	78.6	77.4
BC	Total	2718	2267	83.4	2267	2071	91.4	76.2	2015	1715	85.1	1715	1577	92.0	78.3	77.1

* = régions socio-sanitaires regroupées

10.3 Erreurs dans les enquêtes

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

Les erreurs qui ne sont pas liées à l'échantillonnage peuvent être commises à presque toutes les étapes d'une enquête. Il est possible que les intervieweurs comprennent mal les instructions, que les répondants fassent des erreurs en complétant le questionnaire, que les réponses soient mal saisies et que des erreurs se produisent au moment du traitement et de la totalisation des données. Tous ces exemples représentent des erreurs non dues à l'échantillonnage.

10.3.1 Erreurs non dues à l'échantillonnage

Sur un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations tirées de l'enquête. Toutefois, les erreurs qui se produisent systématiquement contribueront à des biais dans les estimations de l'enquête. On a consacré beaucoup de temps et d'efforts à réduire les erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été appliquées à chaque étape du cycle de collecte et de traitement des données afin de contrôler la qualité des données. On a notamment fait appel à des intervieweurs hautement qualifiés, une formation poussée sur les méthodes d'enquête et le questionnaire et l'observation des intervieweurs afin de déceler les problèmes. La mise à l'essai de l'application IAO et les essais sur le terrain ont également été au nombre des procédures essentielles pour réduire au maximum les erreurs de collecte de données.

L'effet de la non-réponse sur les résultats de l'enquête constitue une source importante d'erreurs non dues à l'échantillonnage dans les enquêtes. L'ampleur de la non-réponse varie de non-réponse partielle (le fait de ne pas répondre à une ou plusieurs questions) à une non-réponse totale. Dans le cas du cycle 2.2 de l'ESCC, il n'y a presque pas eu de non-réponse partielle car une fois le questionnaire débuté les répondants avaient tendance à le terminer. Il y a eu non-réponse totale lorsque la personne sélectionnée pour participer à l'enquête a refusé de le faire ou que l'intervieweur a été incapable d'entrer en contact avec elle. On a traité les cas de non-réponse totale en corrigeant les poids des personnes qui ont répondu à l'enquête afin de compenser pour ceux qui n'ont pas répondu. Voir la section 8 pour avoir de plus amples détails sur la correction de la pondération pour la non-réponse.

10.3.2 Erreurs dues à l'échantillonnage

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

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

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

Statistique Canada utilise fréquemment les résultats du CV pour l'analyse des données et conseille vivement aux utilisateurs produisant des estimations à partir des fichiers de données du cycle 2.2 de l'ESCC de faire de même. Pour plus d'information sur le calcul des CVs, voir la section 11. Pour consulter les lignes directrices sur la façon d'interpréter les résultats du CV, se référer au tableau à la fin de la sous-section 11.5.

11. Lignes directrices pour la totalisation, l'analyse et la diffusion

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

11.1 Lignes directrices pour l'arrondissement

Afin que les estimations calculées d'après ce fichier de microdonnées en vue d'être publiées ou diffusées de toute autre façon correspondent à celles produites par Statistique Canada, il est vivement conseillé à l'utilisateur de les arrondir en se conformant aux lignes directrices suivantes :

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

11.2 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 des fichiers de microdonnées ne pourront être considérées comme représentatives de la population observée et ne correspondront pas à celles de Statistique Canada.

L'utilisateur ne doit pas non plus perdre de vue qu'en raison du traitement réservé au champ du poids, certains progiciels ne permettent pas d'obtenir des estimations qui coïncident exactement avec celles de Statistique Canada.

11.2.1 Définitions des catégories d'estimations : de type nominal par opposition à quantitatives

Avant d'exposer la façon de totaliser et d'analyser les données de l'enquête, il est bon de décrire les deux grandes catégories d'estimations ponctuelles des caractéristiques de la population qui peuvent être produites d'après le fichier de microdonnées de l'enquête.

Estimations de type nominal :

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

Exemple de question de type nominal :

*Actuellement, est-ce que ... fume(z) des cigarettes tous les jours, à l'occasion ou jamais?
(SMKD_202)*

- Tous les jours
- À l'occasion
- Jamais

Estimations quantitatives :

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes ou d'autres mesures de tendance centrale de quantités qui ont trait à tous les membres de la population observée ou à certains d'entre eux.

Un exemple d'estimation quantitative est le nombre moyen de cigarettes que fument par jour les personnes qui fument tous les jours. Le numérateur correspond à l'estimation du nombre total de cigarettes que fument par jour les personnes qui fument tous les jours et le dénominateur, à l'estimation du nombre de personnes qui fument tous les jours.

Exemple de question quantitative :

Actuellement, combien de cigarettes est-ce que ... fume(z) chaque jour?
(**SMKD_204**)

|_|_| Nombre de cigarettes

11.2.2 Totalisation d'estimations de type nominal

On peut obtenir, à partir des fichiers de microdonnées, des estimations du nombre de personnes qui possèdent une caractéristique donnée en additionnant les poids finals de tous les enregistrements contenant des données sur la caractéristique étudiée.

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

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

11.2.3 Totalisation d'estimations quantitatives

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

- multiplier la valeur de la variable étudiée par le poids final, puis faire la somme de cette quantité pour tous les enregistrements visés pour obtenir le numérateur (\hat{X});
- faire la somme des poids finals des enregistrements contenant la variable étudiée pour obtenir le dénominateur (\hat{Y});
- diviser l'estimation du numérateur par l'estimation du dénominateur.

Par exemple, pour estimer le nombre moyen de cigarettes que fument chaque jour les personnes qui fument tous les jours, on calcule d'abord le numérateur (\hat{X}) en calculant le produit de la variable **SMKD_204** par le poids, **WTSD_M** pour tous les enregistrements pour lesquels la valeur de la variable **SMKD_202** est « tous les jours ». On obtient ensuite le dénominateur (\hat{Y}) en additionnant le poids final de tous les enregistrements pour lesquels la valeur de la variable **SMKD_202** est « tous les jours ». Le nombre moyen de cigarettes fumées chaque jour par les personnes qui fument tous les jours est finalement obtenu en divisant (\hat{X}) par (\hat{Y}).

11.3 Lignes directrices pour l'analyse statistique

L'ESCC se fonde sur un plan de sondage complexe qui prévoit une stratification et un échantillonnage à plusieurs degrés, ainsi que la sélection des répondants avec probabilités inégales. L'utilisation des données provenant d'une enquête aussi complexe pose des difficultés aux analystes, car le choix des méthodes d'estimation et de calcul de la variance dépend du plan de sondage et des probabilités de sélection.

Nombre de méthodes d'analyse intégrées aux progiciels statistiques permettent d'utiliser des poids, mais la signification et la définition de ces poids peuvent différer de celles applicables dans le contexte d'une enquête par sondage. Par conséquent, si les estimations calculées au moyen de ces progiciels sont souvent exactes, les variances n'ont, quant à elles, pratiquement aucune signification.

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

Pour permettre à l'utilisateur d'évaluer la qualité des totalisations estimées d'après les données, Statistique Canada a produit un ensemble de tableaux de variabilité d'échantillonnage approximative (couramment appelés « Tableaux des CV ») pour l'ESCC. On peut employer ces tableaux pour obtenir des coefficients de variation approximatifs pour les estimations de type nominal et les proportions. Pour plus de détails, consulter la section 12.

11.4 Software for intake distribution (SIDE)

Il faut distinguer deux concepts en ce qui touche la consommation alimentaire : l'apport d'un jour et l'apport habituel. **L'apport d'un jour** correspond à l'apport nutritionnel total au cours d'une période spécifique de 24 heures. Ces données sont recueillies dans le cadre d'une entrevue au cours de laquelle on demande aux répondants de dresser la liste de tous les aliments consommés de minuit à minuit la veille. **L'apport habituel** correspond à une vue d'ensemble des aliments consommés pendant une journée type.

Il est impossible d'estimer directement l'apport habituel à partir de l'apport d'un jour. En effet, l'apport habituel varie d'une personne à l'autre. L'apport d'un jour diffère lui aussi d'une personne à l'autre mais, pour une personne donnée, il varie également d'une journée à l'autre. Par conséquent, pour estimer l'apport habituel, il importe de distinguer la variation de l'apport chez une personne et la variation entre les personnes. On a utilisé pour ce faire le programme *Software for Intake Distribution Estimation* (SIDE), conçu par l'Iowa State University. Ce logiciel effectue une série de transformations mathématiques pour estimer chaque composante de la variance ainsi que la répartition de l'apport habituel d'un nutriment. Dans ces calculs, le jour de la semaine a été utilisé pour enlever partiellement l'effet de classification. Comme la moyenne établie pour la première journée d'entrevue est utilisée comme repère, les moyennes calculées à partir de l'apport d'un jour et de l'apport habituel sont les mêmes. Toutefois, pour calculer les estimations

de la proportion de la population située sous ou au-dessus d'un seuil donné, il faut utiliser une répartition de l'apport habituel. Pour plus de renseignements sur le programme SIDE, veuillez consulter le document de Statistique Canada intitulé : « *Enquête sur la santé dans les collectivités canadiennes (ESCC) Cycle 2.2 (Nutrition) – Documentation sur Software for Intake Distribution Estimation (SIDE)*. »

11.5 Lignes directrices pour la diffusion

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

Table 11.1 : Lignes directrices relatives à la variabilité d'échantillonnage

Type d'estimation	c.v. (en %)	Lignes directrices
Acceptable	$0,0 \leq \text{c.v.} \leq 16,6$	On peut envisager une diffusion générale non restreinte des estimations. Aucune annotation particulière n'est nécessaire.
Marginale	$16,6 < \text{c.v.} \leq 33,3$	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	$\text{c.v.} > 33,3$	Statistique Canada recommande de ne pas publier des estimations dont la qualité est inacceptable. Toutefois, si l'utilisateur choisit de le faire, il doit alors adjoindre la lettre F (ou un autre identificateur semblable) et les diffuser avec l'avertissement suivant : « Nous avisons l'utilisateur que ... (précisez les données)... ne répondent pas aux normes de qualité de Statistique Canada pour ce programme statistique. Les conclusions tirées de ces données ne sauraient être fiables et seront fort probablement erronées. Ces données et toute conclusion qu'on pourrait en tirer ne doivent pas être publiées. Si l'utilisateur choisit de les publier, il est alors tenu de publier également le présent avertissement. »

12. Tableaux de la variabilité d'échantillonnage approximative

Afin de permettre aux utilisateurs d'avoir facilement accès à des coefficients de variation qui s'appliqueraient à une multitude d'estimations de type nominal obtenues à partir de ce fichier de microdonnées, Statistique Canada a produit un ensemble de tableaux de la variabilité d'échantillonnage approximative (aussi appelés tableaux de CV). Ces tableaux permettent aux utilisateurs d'obtenir un coefficient de variation approximatif selon la taille de l'estimation calculée à partir des données de l'enquête.

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

Les effets de plan, les tailles d'échantillon et les comptes de population qui ont servi à produire les tableaux de la variabilité d'échantillonnage approximative de même que les tableaux sont disponibles à l'annexe E. Notez qu'il y a deux séries de tableaux de CV, une pour l'échantillon total et une pour le sous-échantillon pour lequel on a les mesures exactes de l'indice de masse corporelle. La série de tableaux à utiliser dépend de la caractéristique estimée. Si la caractéristique à estimer est liée aux variables relatives au sous-échantillon avec mesures exactes de l'indice de masse corporelle (voir section 12.2 pour la liste des variables), ce sont les tableaux de CV de ce sous-échantillon qui doivent être utilisés. Si la caractéristique à estimer n'est pas liée aux variables relatives au sous-échantillon avec mesures exactes de l'indice de masse corporelle, ce sont les tableaux de CV de l'échantillon total qui doivent être utilisés. Tous les coefficients de variation sont *approximatifs* dans les tableaux de la variabilité d'échantillonnage approximative et ils ne doivent donc pas être considérés comme des valeurs exactes. Les possibilités concernant le calcul d'un coefficient de variation exact sont discutées dans la sous-section 12.7.

Rappel : Tel qu'indiqué dans les lignes directrices à la section 11.5, si le nombre d'observations sur lesquelles une estimation est basée est inférieur à 30, l'estimation pondérée ne doit pas être diffusée, quelle que soit la valeur de son coefficient de variation. Les coefficients de variation d'estimations basées sur des échantillons de petite taille sont trop imprévisibles pour être adéquatement représentés dans les tableaux.

12.1 Comment utiliser les tableaux de CV pour les estimations de type nominal

Les règles suivantes devraient permettre à l'utilisateur de calculer à partir des tableaux de la variabilité d'échantillonnage, les coefficients de variation approximatifs d'estimations relatives au nombre, à la proportion ou au pourcentage de personnes dans la population observée qui possèdent une caractéristique donnée ainsi que des rapports et des écarts entre ces estimations.

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

Le coefficient de variation dépend uniquement de la taille de l'estimation elle-même. Dans le tableau de variabilité d'échantillonnage correspondant à la région appropriée, il faut repérer l'estimation calculée dans la colonne d'extrême gauche (intitulée «Numérateur du pourcentage») et suivre les astérisques (s'il y en a) de gauche à droite jusqu'au premier nombre. Puisque toutes les valeurs possibles de l'estimation ne sont pas disponibles dans cette colonne, il faut prendre la valeur la plus petite qui s'en rapproche le plus (par exemple, si l'estimation vaut 1 700 et que les deux valeurs disponibles dans la colonne « Total » se rapprochant le plus de 1 700 sont 1 000 et 2 000, il faut choisir 1 000). Ce nombre constitue le coefficient de variation approximatif pour l'estimation en question.

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

Le coefficient de variation d'une proportion (ou d'un pourcentage) estimée dépend à la fois de l'ordre de grandeur de cette proportion et de l'ordre de grandeur du numérateur utilisé dans le calcul de la proportion. Les proportions estimées sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion lorsque celle-ci est fondée sur un sous-ensemble de la population. Cela est dû au fait que les coefficients de variation des estimations du dernier type sont basés sur le chiffre le plus élevé dans une rangée d'un tableau particulier, tandis que les coefficients de variation des estimations du premier type sont basés sur un chiffre quelconque de cette même rangée (pas nécessairement le plus élevé). (Il convient de noter que dans les tableaux, la valeur des coefficients de variation décroît de gauche à droite sur une même ligne.) Par exemple, la proportion estimative de personnes qui fument tous les jours parmi les fumeurs est plus fiable que le nombre estimatif de personnes qui fument tous les jours.

Lorsque la proportion (ou le pourcentage) est fondée sur la population totale de la région géographique à laquelle le tableau s'applique, le coefficient de variation de la proportion est égal à celui du numérateur de la proportion. Dans ce cas-ci, cela équivaut à appliquer la règle 1.

Lorsque la proportion (ou le pourcentage) est fondée sur un sous-ensemble de la population totale (p. ex., les personnes qui fument), il faut se reporter à la proportion (haut du tableau) et au numérateur de la proportion ou du pourcentage (côté gauche du tableau). Puisque toutes les valeurs possibles de la proportion et du numérateur ne sont pas disponibles, il faut, dans les deux cas, prendre la valeur la plus petite qui s'en rapproche le plus (par exemple, si la proportion est de 23 % et que les deux valeurs disponibles sur la ligne « Pourcentage estimé » s'en rapprochant le plus sont 20 % et 25 %, il faut choisir 20 %). Le coefficient de variation se trouve à l'intersection de la ligne et de la colonne appropriées. Si, à cet endroit, on retrouve des astérisques, il faut prendre le premier chiffre que l'on retrouve à la droite de ces astérisques.

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 la différence entre des estimations de sous-populations indépendantes. Dans le cas où les sous-populations ne sont pas indépendantes, ce calcul mènera à une surestimation ou une sous-estimation selon que la relation (ou corrélation) entre ces deux sous-populations est positive ou négative.

Règle 4 : Estimations de rapports

Si le numérateur est un sous-ensemble du dénominateur, il faut convertir le rapport en pourcentage et appliquer la règle 2. Ce serait le cas, par exemple, si le dénominateur est le nombre de personnes qui fument et le numérateur est le nombre de personnes qui fument tous les jours parmi celles qui fument.

Si le numérateur n'est pas un sous-ensemble du dénominateur (par exemple, le rapport du nombre de personnes qui fument tous les jours ou à l'occasion au nombre de personnes qui ne fument pas du tout), l'écart-type du rapport entre les estimations est à peu près égal à la racine carrée de la somme des carrés de chaque coefficient de variation pris séparément multipliée par \hat{R} , où \hat{R} est le rapport des estimations ($\hat{R} = \hat{X}_1 / \hat{X}_2$). L'erreur-type d'un rapport est donc :

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

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

Le coefficient de variation de \hat{R} est donné par $\sigma_{\hat{R}} / \hat{R} = \sqrt{\alpha_1^2 + \alpha_2^2}$. La formule tend à surestimer l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélés positivement et à sous-estimer l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélés négativement.

Règle 5 : Estimations des différences entre des rapports

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

12.2 Exemples d'utilisation des tableaux de CV pour des estimations de type nominal

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

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

Supposons qu'un utilisateur estime à 5 137 429 le nombre de personnes (âgées de 12 ans et plus) qui fument tous les jours au Canada. Comment l'utilisateur fait-il pour déterminer le coefficient de variation de cette estimation?

- 1) Se reporter au tableau de CV pour le CANADA.
- 2) L'agrégat estimé (5 137 429) 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 5 000 000.
- 3) Le coefficient de variation d'un agrégat estimé (exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 2,8 %.
- 4) Le coefficient de variation approximatif de l'estimation est donc 2,8 %. Par conséquent, selon les lignes directrices présentées à la section 10.4, l'estimation selon laquelle 5 137 429 personnes fument tous les jours peut être diffusée sans réserve.

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

Supposons qu'un utilisateur estime à $5\,137\,429 / 6\,256\,048 = 82,1 \%$ le pourcentage de personnes, parmi les fumeurs, qui fument tous les jours au Canada. Comment l'utilisateur fait-il pour déterminer le coefficient de variation de cette estimation?

- 1) Se reporter au tableau de CV pour le CANADA.
- 2) Parce que l'estimation est un pourcentage basé sur un sous-ensemble de la population totale (c.-à-d. les personnes qui fument tous les jours ou à l'occasion), il faut utiliser à la fois le pourcentage (82,1 %) et la partie numérateur du pourcentage (5 137 429) pour déterminer le coefficient de variation.
- 3) Le numérateur (5 137 429) 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 5 000 000. De même, l'estimation du pourcentage ne figure pas parmi les en-têtes de colonnes; il faut donc utiliser le nombre le plus petit qui s'en rapproche le plus, soit 70,0 %.
- 4) Le nombre qui se trouve à l'intersection de la ligne et de la colonne utilisées, soit 1,7 %, est le coefficient de variation (exprimé en pourcentage) à employer.
- 5) Le coefficient de variation de l'estimation est donc 1,7 %. Par conséquent, selon les lignes directrices présentées à la section 10.4, l'estimation selon laquelle 82,1 % des gens qui fument le font tous les jours peut être diffusée sans réserve.

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

Supposons qu'un utilisateur estime que, parmi les hommes (âgés de 12 ans et plus), $2\,881\,670/13\,178\,748 = 21,9\%$ fument tous les jours (estimation 1), alors que chez les femmes (âgées de 12 ans et plus), ce pourcentage est estimé à $2\,255\,759/13\,530\,930 = 16,7\%$ (estimation 2). Comment l'utilisateur fait-il pour déterminer le coefficient de variation de la différence entre ces deux estimations?

- 1) À l'aide du tableau de CV pour le CANADA, utilisé de la même façon que dans l'exemple 2, vous établissez à 4,5 % le CV de l'estimation 1 (exprimé en pourcentage) et à 4,6 % le CV de l'estimation 2 (exprimé en pourcentage).
- 2) Selon la règle 3, l'erreur-type pour une différence ($\hat{d} = \hat{X}_2 - \hat{X}_1$) est :

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

où \hat{X}_1 est l'estimation 1, \hat{X}_2 est l'estimation 2, et α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement. L'erreur-type de la différence $\hat{d} = (0,219 - 0,167) = 0,052$ est donc :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0,219)(0,045)]^2 + [(0,167)(0,046)]^2} \\ &= 0,012\end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / |\hat{d}| = 0,012/0,052 = 0,231$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 23,1 % (exprimé en pourcentage). Par conséquent, toujours selon les lignes directrices présentées à la section 10.4, cette estimation peut être publiée sans réserve.

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime à 5 137 429 le nombre de personnes qui fument tous les jours et à 1 118 619 le nombre de celles qui fument à l'occasion. L'utilisateur veut comparer ces deux estimations sous la forme d'un rapport. Comment fait-il pour déterminer le coefficient de variation de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation de rapport, où le numérateur de l'estimation (= \hat{X}_1) est le nombre de personnes qui fument à l'occasion. Le dénominateur de l'estimation (= \hat{X}_2) est le nombre de personnes qui fument tous les jours.
- 2) Se reporter au tableau de CV pour le CANADA.
- 3) Le numérateur de cette estimation de rapport est 1 118 619. Le nombre le plus petit qui se rapproche le plus de ce nombre est 1 000 000. Le coefficient de variation de cette estimation

(exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 6,9 %.

4) Le dénominateur de cette estimation de rapport 5 137 429. Le nombre le plus petit qui se rapproche le plus de ce nombre est 5 000 000. Le coefficient de variation de cette estimation (exprimé en pourcentage) est la première entrée sur cette ligne (à part les astérisques), soit 2,8 %.

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

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

c'est-à-dire,

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0,069)^2 + (0,028)^2} \\ &= 0,074\end{aligned}$$

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement. Le rapport des personnes qui fument occasionnellement à celles qui fument tous les jours est 1 118 619/5 137 429, soit 0,22:1. Le coefficient de variation de cette estimation est 7,4 % (exprimé en pourcentage); selon les lignes directrices présentées à la section 10.4, l'estimation peut donc être diffusée sans réserve.

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

Bien que les coefficients de variation soient largement utilisés, l'intervalle de confiance d'une estimation représente une mesure plus intuitive de l'erreur d'échantillonnage. Un intervalle de confiance est une façon d'énoncer la probabilité que la valeur vraie de la population se situe dans une plage de valeurs données. Par exemple, un intervalle de confiance de 95 % peut être décrit comme suit : si l'échantillonnage de la population se répète à l'infini, chacun des échantillons donnant un nouvel intervalle de confiance pour une estimation, l'intervalle contiendra la valeur vraie de la population dans 95 % des cas.

Une fois déterminée l'erreur-type d'une estimation, on peut calculer des intervalles de confiance pour les estimations en partant de l'hypothèse qu'en procédant à un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique de la population sont réparties selon une distribution normale autour de la valeur vraie de la population. Selon cette hypothèse, il y a environ 68 chances sur 100 que l'écart entre une estimation de l'échantillon et la valeur vraie de la population soit inférieur à une erreur-type, environ 95 chances sur 100 que l'écart soit inférieur à deux erreurs-types et environ 99 chances sur 100 que l'écart soit inférieur à trois erreurs-types. On appelle ces différents degrés de confiance des niveaux de confiance.

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

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

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

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

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

Note : Les lignes directrices concernant la diffusion des estimations de la section 10.4 s'appliquent aussi aux intervalles de confiance. Par conséquent, si l'estimation ne peut être diffusée, alors l'intervalle de confiance ne peut l'être lui non plus.

12.4 Exemple d'utilisation de tableaux de CV pour obtenir des limites de confiance

Voici la marche à suivre pour calculer un intervalle de confiance de 95 % pour la proportion estimée de personnes qui fument tous les jours parmi celles qui fument (d'après l'exemple 2 de la sous-section 12.2).

$$\hat{X} = 0,821$$

$$z = 2$$

$$\alpha_{\hat{X}} = 0,017 \text{ est le coefficient de variation de cette estimation selon les tableaux.}$$

$$CIX = \{0,821 - (2) (0,821) (0,017), 0,821 + (2) (0,821) (0,017)\}$$

$$CIX = \{0,793, 0,849\}$$

12.5 Comment utiliser les tableaux de CV pour effectuer un test Z

On peut aussi utiliser les erreurs-types pour effectuer des tests d'hypothèses, une technique qui permet de faire la distinction entre les paramètres d'une population à l'aide d'estimations basées sur un échantillon. Ces estimations peuvent être des nombres, des moyennes, des pourcentages, des rapports, etc. Les tests peuvent être effectués à divers niveaux de signification; un niveau de signification est la probabilité de conclure que les caractéristiques sont différentes quand, en fait, elles sont identiques.

Supposons que \hat{X}_1 et \hat{X}_2 sont des estimations basées sur un échantillon pour deux caractéristiques voulues. Supposons aussi que l'erreur-type de la différence $\hat{X}_1 - \hat{X}_2$ est $\sigma_{\hat{a}}$. Si $z = (\hat{X}_1 - \hat{X}_2) / \sigma_{\hat{a}}$ est compris entre -2 et 2, alors on ne peut tirer aucune conclusion à propos de la différence entre les caractéristiques au niveau de signification de 5 %. Toutefois, si ce rapport est inférieur à -2 ou supérieur à +2, la différence observée est significative au niveau de 0,05.

12.6 Exemple d'utilisation des tableaux de CV pour effectuer un test Z

Supposons que nous voulions tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y a pas de différence entre la proportion d'hommes qui fument tous les jours et cette même proportion chez les femmes. Dans l'exemple 3 de la sous-section 11.2, nous avons déterminé que l'erreur-type de la différence entre ces deux estimations est égale à 0,012. Par conséquent,

$$z = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{a}}} = \frac{0,219 - 0,167}{0,012} = \frac{0,052}{0,012} = 4,3$$

Puisque $z = 4,3$ est supérieur à 2, on doit conclure qu'il existe une différence significative entre les deux estimations au niveau de signification de 0,05. À noter que les deux sous-groupes comparés sont considérés comme étant indépendants, ce qui fait en sorte que le résultat du test est valide.

12.7 Variances ou coefficients de variation exacts

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

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

Le calcul de coefficients de variation (ou tout autre mesure de précision) fait à l'aide de la méthode du bootstrap nécessite toutefois l'accès à de l'information considérée confidentielle qui n'est évidemment pas disponible dans le fichier de microdonnées à grande diffusion. Le calcul doit donc se faire à l'aide du fichier maître. L'accès au fichier maître est discuté à la section 13.4.

Pour le calcul de coefficients de variation, il est conseillé d'utiliser la méthode du bootstrap. Un programme macro, appelé le "Bootvar", a été développé pour faciliter le calcul à l'aide de la méthode bootstrap. Le programme Bootvar est offert en formats SAS et SPSS, et est constitué de

macros qui calculent les variances de totaux, ratios, différences entre ratios, et pour des régressions linéaires et logistiques.

Les raisons pour lesquelles un utilisateur pourrait souhaiter connaître la précision exacte de ses estimations sont diverses. En voici quelques-unes.

Premièrement, si un utilisateur désire obtenir des estimations à un niveau géographique autre que ceux présentés dans les tableaux (par exemple, au niveau urbain ou rural), l'utilisation des tableaux de CV publiés ne convient pas parfaitement. Néanmoins, on peut obtenir les coefficients de variation de ce type d'estimations en appliquant la méthode d'estimation par domaine, au moyen du programme de calcul de la variance exacte (le "Bootvar").

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

Troisièmement, dans le cas de l'estimation de variables quantitatives, il est nécessaire d'utiliser des tableaux distincts pour déterminer l'erreur d'échantillonnage. Or, la plupart des variables de l'ESCC étant de type nominal, de tels tableaux n'ont pas été produits. Les utilisateurs qui souhaitent connaître les coefficients de variation de variables quantitatives peuvent néanmoins obtenir ces derniers grâce au programme de calcul de la variance réelle. À noter, toutefois, que le coefficient de variation d'un total quantitatif est généralement plus grand que celui de l'estimation de type nominal correspondante (c'est-à-dire, l'estimation du nombre de personnes qui contribuent à l'estimation quantitative). Si l'estimation de type nominal correspondante ne peut être diffusée, il en sera de même pour l'estimation quantitative. Par exemple, le coefficient de variation de l'estimation du nombre total de cigarettes que fument chaque jour les personnes qui fument tous les jours serait supérieur à celui de l'estimation correspondante du nombre de personnes qui fument tous les jours. Par conséquent, si on ne peut diffuser le coefficient de variation de cette dernière estimation, on ne pourra non plus diffuser celui de l'estimation quantitative correspondante.

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

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

L'annexe E présente les tableaux indiquant les seuils de diffusion des totaux selon les estimations pour le Canada, les provinces, les régions sociosanitaires ainsi que pour les différents groupes d'âges (pour le Canada seulement). Il y a une série de tables pour l'échantillon total et une autre série pour l'échantillon de mesures exactes de l'indice de masse corporelle. Les estimations inférieures à la valeur indiquée dans la colonne «Marginal» ne peuvent en aucun cas être diffusées.

13. Utilisation du fichier

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

13.1 Poids d'échantillonnage principal

D'abord la variable de poids **WTSD_M** correspond au poids d'échantillonnage principal du cycle 2.2 de l'ESCC et est relative à l'ensemble des répondants 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. Le poids d'échantillonnage doit toujours être utilisée, afin de permettre l'inférence à l'échelle de la population. La variable de poids **WTSD_M** doit être utilisée en tout temps, sauf dans les cas où l'analyse implique les variables reliées aux mesures exactes de l'indice de masse corporelle. La production de résultats non pondérés n'est pas recommandée. La répartition de l'échantillon, de même que les détails du plan de sondage, peuvent entraîner des résultats biaisés qui ne représentent pas correctement la population. Pour une description plus détaillée du calcul de ce poids, consulter la section 8 sur la pondération.

13.2 Poids d'échantillonnage pour les mesures exactes de l'indice de masse corporelle

La plupart des variables peuvent être analysées en utilisant le poids **WTSD_M**. Cependant, pour analyser les variables reliées aux mesures exactes de taille et de poids, c'est le poids **WTSD_MHW** qui doit être utilisé. Voici la liste des variables pour lesquelles cette variable de poids doit être utilisée : **MHWDGHTM**, **MHWDGWTK**, **MHWDGBMI**, **MHWDGISW** et **MHWDGCOL**. Par exemple, si on veut étudier le croisement de la classification de l'indice de masse corporelle (IMC) pour les 18 ans et plus (**MHWDGISW**) avec le niveau d'éducation du répondant (**EDUDDR04**), on doit utiliser le poids **WTSD_MHW** et ce, même si la variable **EDUDDR04** est disponible pour tous les répondants. Les tableaux I et II donnent les résultats de ce croisement en utilisant le mauvais et le bon poids respectivement.

Tableau 13.1 - Résultats en utilisant le mauvais poids (WTSD_M)

MHWDGISW	Moins qu'un diplôme d'études secondaires	EDUDDR04			Total
		Diplôme d'études secondaires, aucune étude postsecondaire	Études post-secondaires partielles	Grade/diplôme d'études post-secondaires	
Poids insuffisant	70 580	52 770	29 913	118 747	272 010
Poids normal	790 927	1 049 469	551 367	2 966 304	5 358 068
Embonpoint	962 170	819 911	439 929	2 623 916	4 845 925
Obésité – Classe I, II, III	841 368	634 189	262 474	1 439 233	3 177 264
Total	2 665 045	2 556 339	1 283 683	7 148 200	13 653 267

Tableau 13.2 Résultats en utilisant le bon poids (WTSD_MHW)

MHWDGISW	Moins qu'un diplôme d'études secondaires	EDUDDR04			Total
		Diplôme d'études secondaires, aucune étude postsecondaire	Études post-secondaires partielles	Grade/diplôme d'études post-secondaires	
Poids insuffisant	114 036	94 632	49 768	212 880	471 315
Poids normal	1 328 276	1 864 471	960 449	5 075 780	9 228 976
Embonpoint	1 692 783	1 436 898	763 445	4 662 882	8 556 007
Obésité – Classe I, II, III	1 410 899	1 113 254	436 421	2 512 821	5 473 395
Total	4 545 994	4 509 255	2 210 083	12 464 363	23 729 693

Notez que le poids WTSD_MHW a été créé afin d'analyser les variables se rapportant à l'IMC et que le poids WTSD_MHW est manquant pour quelques répondantes pour lesquelles les mesures de taille et de poids sont quand même disponibles. L'IMC n'a pas été calculé pour ces répondantes étant donné que dans ces cas particuliers et d'après les réponses données il était impossible de savoir si ces femmes étaient enceintes ou non et que l'IMC n'est pas calculé pour les femmes enceintes.

13.3 Convention appliquée pour nommer les variables

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

13.3.1 Structure élémentaire des noms des variables de l'ESCC

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

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

Par exemple, la structure du nom de la variable correspondant à la question 202, module Usage du tabac, cycle 2.1, c'est-à-dire SMKD_202 est la suivante :

Positions 1 à 3 : SMK Module sur l'usage du tabac
Position 4 : D Cycle 2.2
Position 5 : _ (_ = données recueillies)
Position 6 à 8 : 202 numéro de la question et option de réponse

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

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

ADM	Renseignements administratifs	INC	Revenu
ALC	Consommation d'alcool	LBF	Population active
CCC	Problèmes de santé chronique	MHW	Taille et poids - mesurés
CPA	Activités physiques des enfants	PAC	Activités physiques
DHH	Variables démographiques et sur les ménages	R24	Rappel alimentaire de 24 heures
EDU	Éducation	SAC	Activités sédentaires
FSC	Sécurité alimentaire	SAM	Variables de l'échantillon
FSD	Détail sommaire alimentaire	SDC	Renseignements socio-démographiques
FVC	Consommation de fruits et légumes	SMK	Usage du tabac
GEN	État de santé général	VSD	Détail sommaire des vitamines
GEO	Variables géographiques	WHC	Santé des femmes
HWT	Taille et poids - autodéclarés	WTS	Poids

13.3.3 Position 4 : Cycle

Cycle Description

- A** Cycle 1.1 : Enquête sur la santé dans les collectivités canadiennes
 : enquête à l'échelle régionale, échantillon stratifié selon la région sociosanitaire
 : contenu commun et contenu optionnel sélectionnés par les régions sociosanitaires
 : estimations à l'échelle régionale (régions sociosanitaires), (provincial, territorial et national)
- B** Cycle 1.2 : Enquête sur la santé dans les collectivités canadiennes, santé mentale et bien-être
 : enquête à l'échelle provinciale
 : contenu thématique et contenu général supplémentaire
 : estimations aux échelle provinciale et nationale
- C** Cycle 2.1 : Enquête sur la santé dans les collectivités canadiennes
 : enquête à l'échelle régionale, échantillon stratifié selon la région socio-sanitaire
 : contenu commun et contenu optionnel sélectionnés par les régions sociosanitaires
 : estimations à l'échelle régionale (régions sociosanitaires), (provincial, territorial et national)
- D** Cycle 2.2 : Enquête sur la santé dans les collectivités canadiennes, nutrition
 : enquête à l'échelle provinciale
 : contenu thématique et contenu général supplémentaire
 : estimations aux échelle provinciale et nationale

13.3.4 Position 5 : Type de variable

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

13.3.5 Positions 6 à 8 : Nom de la variable

En général, les trois dernières positions correspondent à la numérotation de la variable qui figure sur le questionnaire. On supprime la lettre « Q » utilisée pour représenter le mot “question” et on

présente tous les numéros de question au moyen d'un groupe de deux chiffres. Par exemple, la question Q01A du questionnaire devient simplement 01A et la question Q15, simplement 15.

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

13.4 Accès au fichier maître

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

Une première méthode offerte à tous les utilisateurs consiste à demander au personnel des Services personnalisés à la clientèle de la Division de la statistique de la santé de produire des totalisations personnalisées. Ce service permet aux utilisateurs qui ne savent pas se servir de logiciels de totalisation d'obtenir des résultats personnalisés. Les résultats sont filtrés pour s'assurer qu'ils sont conformes aux normes de confidentialité et de fiabilité avant d'être diffusés. Ce service est offert contre remboursement des frais.

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

En dernier lieu, le service de téléaccès aux fichiers maîtres de l'enquête est un moyen d'accéder à ces données si il est impossible de passer par un Centre. On peut fournir à l'acheteur d'un produit de microdonnées un fichier maître "fictif" d'essai et le cliché d'enregistrement correspondant. Grâce à ces outils, il peut mettre au point son propre ensemble de programmes analytiques en se servant du fichier fictif pour confirmer que les routines fonctionnent convenablement. Il ne lui reste plus qu'à envoyer le code pour les totalisations personnalisées par courrier électronique à cchs-escs@statcan.ca. Le code est transmis au réseau interne protégé de Statistique Canada et traité en regard du fichier maître approprié de données du cycle 2.2 de l'ESCC. Les estimations générées seront communiquées à l'utilisateur, sujet aux directives sur l'analyse et la communication des données tel qu'exposé dans les grandes lignes à la section 10 de ce document. Les résultats sont filtrés pour vérifier s'ils sont conformes aux normes de

confidentialité et de fiabilité, puis, les données de sortie sont renvoyées au client. Ce service est gratuit.

Annexe 1 : Questions analytiques

1. Données sur les suppléments vitaminiques et minéraux

Fichiers de données sur les suppléments de vitamines et minéraux

Avec la diffusion de la vague 3 des fichiers de l'ESCC 2.2, les données sur la consommation de suppléments de vitamines et minéraux sont maintenant accessibles. Ces données sont contenues notamment dans les fichiers suivants :

- 1. Fichier de description des vitamines et des minéraux (VDC) :** Il s'agit d'une base de données de type administratif qui ne contient pas de renseignements fournis par les répondants, mais plutôt de l'information sur les suppléments qui ont été consommés par les répondants. Le fichier comprend le numéro d'identification du médicament (DIN), la description du produit, la dose et l'unité posologique ainsi que les ingrédients et leur concentration par unité posologique. Contrairement au fichier de description des aliments (FDC), le fichier VDC renferme uniquement des enregistrements sur les suppléments de vitamines et minéraux déclarés par au moins un répondant.
- 2. Fichier de détail des vitamines et des minéraux (VMD) :** Ce fichier fournit des renseignements sur chaque supplément consommé par un répondant. Dans ce fichier, les répondants ont un enregistrement par supplément consommé (jusqu'à concurrence de 10). Les répondants qui n'ont pas consommé de suppléments ne sont pas inclus dans ce fichier. Les variables de ce fichier comprennent le DIN, les habitudes de consommation, une variable dérivée qui indique le nombre de jours de consommation du supplément dans un mois, ainsi que les apports nutritionnels quotidiens moyens de chaque supplément.
- 3. Fichier sommaire des vitamines et des minéraux (VSD) :** Ce fichier contient un enregistrement par répondant, même pour les répondants qui n'ont pas consommé de suppléments. Les variables de ce fichier comprennent une indication de la consommation ou de la non-consommation d'un supplément par le répondant, le nombre de suppléments différents consommés et l'apport nutritionnel quotidien moyen total de tous les suppléments.
- 4. Fichiers HS et HS_SIDE :** Les fichiers HS et HS_SIDE comprennent des indicateurs qui montrent si un répondant a consommé un nutriment sous forme de supplément. Ces fichiers comprennent en outre des variables qui indiquent si le répondant a consommé un supplément et le nombre de suppléments différents consommés. Le fichier HS_SIDE contient les mêmes variables que le fichier HS, mais en format compatible avec le logiciel SIDE.

Toutes les analyses n'exigeront pas l'utilisation de la totalité des fichiers. Pour certaines analyses, le fichier VSD répondra aux besoins de l'utilisateur si ce dernier ne s'intéresse qu'aux apports nutritionnels totaux des suppléments. Les analystes qui désirent combiner les données sur les aliments et les suppléments pour déterminer les niveaux d'insuffisance des nutriments dans la population auront sans doute davantage recours au fichier VMD afin d'examiner les habitudes de consommation de suppléments en vue d'identifier les consommateurs occasionnels.

Combinaison des données sur les aliments et les suppléments

Problèmes d'analyse

La diffusion de la vague 3 comprend des apports nutritionnels distincts pour les aliments et les suppléments. Statistique Canada ne combinera pas les données sur les aliments et les suppléments pour les utilisateurs. La combinaison de ces données est compliquée puisqu'elle exige diverses hypothèses de la part de l'analyste. De plus, ces hypothèses varient selon le type d'analyse qui est faite. Il est donc préférable que l'analyste prenne les décisions qui s'imposent cas par cas. Les utilisateurs qui désirent combiner les données sur les aliments et les suppléments doivent être conscients notamment des problèmes suivants :

1. **Apport quotidien et apport habituel** : Les données sur les suppléments ont été recueillies dans le cadre du volet général sur la santé et non dans le cadre du rappel alimentaire de 24 heures. Les questions sur la consommation de suppléments portaient sur la consommation au cours des 30 derniers jours. Nous ne disposons donc pas d'une mesure de l'apport quotidien pour les suppléments comme dans le cas des aliments.
2. **Inexistence d'une deuxième interview de rappel** : Comme les données sur les suppléments ont été recueillies le cadre du volet général sur la santé, nous ne disposons pas d'une deuxième interview de rappel comme dans le cas des aliments. Il est donc impossible de connaître dans quelle mesure la variation quotidienne de consommation de suppléments influence la variation quotidienne fondée sur les aliments, ce qui influera sur les rajustements faits à l'aide du logiciel SIDE pour calculer l'apport habituel à partir de l'apport quotidien.
3. **Consommateurs occasionnels** : Ces répondants posent un problème, que les données soient recueillies dans le cadre de la deuxième interview de rappel ou non. Il s'agit des personnes qui prennent des suppléments à l'occasion et peut-être à très fortes doses. Les consommateurs occasionnels ont une variance intra-individuelle élevée.
4. **Les nutriments ne peuvent pas tous être combinés** : La vitamine E est incluse dans le fichier des suppléments, mais pas dans celui des aliments. De même, il existe plus de variables concernant les lipides pour les aliments que pour les suppléments. Enfin, la vitamine A ne peut être combinée puisqu'elle est mesurée différemment pour les aliments et les suppléments.

Méthodes

Il est très important de tenir compte de la distinction entre l'apport quotidien et habituel en combinant les données sur les aliments et les suppléments. L'apport quotidien est calculé à partir d'un rappel alimentaire. Il présente une variation non seulement entre les personnes, mais aussi entre les jours pour une même personne (variance intra-individuelle). L'apport habituel est l'apport moyen à long terme. Pour calculer l'apport habituel, le logiciel SIDE (ou un autre logiciel semblable) est utilisé pour rajuster le premier interview de rappel en fonction du deuxième pour supprimer la variance intra-individuelle.

Le module des suppléments de l'ESCC 2.2 mesure l'apport habituel (consommation pendant une période de 30 jours). Toutefois nous ne disposons pas d'une mesure de l'apport quotidien à partir

d'un rappel alimentaire. Sans une mesure de l'apport quotidien, il est impossible d'évaluer les effets de la variabilité intra-individuelle sur la consommation de suppléments. Comme la British Columbia Nutrition Survey (BCNS) a recueilli des renseignements sur l'apport quotidien et habituel, il a été possible d'évaluer les effets de la variance intra-individuelle. La BCNS a révélé que la variance intra-individuelle était extrêmement faible, avoisinant zéro pour 80 % à 90 % de l'échantillon. Par conséquent, la majorité des répondants ont consommé régulièrement des suppléments. La variabilité intra-individuelle dans l'échantillon de l'ESCC 2.2 est inconnue.

Bien que Statistique Canada ne compte pas combiner les données sur l'apport en aliments et en suppléments, les analystes peuvent recourir à diverses stratégies pour combiner ces données. On trouvera ci-dessous les trois méthodes les plus courantes qui ont été utilisées avec d'autres ensembles de données pour combiner les données sur les aliments et les suppléments.

1. Première méthode : La première méthode est la méthode idéale, puisqu'elle est corrigée pour la variance intra-individuelle à la fois pour les aliments et les suppléments. Comme cette méthode exige au moins deux interviews de rappel pour les données sur les aliments et les suppléments, elle ne peut être utilisée pour les données de l'ESCC 2.2. Pour combiner les données sur les aliments et les suppléments, l'utilisateur devra :
 - Additionner l'apport quotidien en aliments et l'apport quotidien en suppléments (premier interview de rappel).
 - Additionner l'apport quotidien en aliments et l'apport quotidien en suppléments (deuxième interview de rappel).
 - Rajuster le premier interview de rappel en fonction du deuxième.

Bien que cette méthode prenne en compte la variabilité intra-individuelle à la fois pour la consommation d'aliments et pour la consommation de suppléments, les consommateurs occasionnels peuvent demeurer un problème avec cette méthode. Il se peut que ces répondants ne prennent des suppléments qu'à l'occasion. Par conséquent, les consommateurs occasionnels qui ont pris un supplément le jour des interviews de rappel seraient considérés comme des consommateurs habituels, alors que les répondants qui n'en ont pas consommé le jour de ces interviews ne seraient pas saisis. Un problème semblable est celui des consommateurs qui ont pris de très fortes doses le jour des interviews de rappel, sans toutefois en consommer régulièrement. Ces répondants pourraient fausser la distribution avec leur apport très élevé, mais occasionnel.

2. Deuxième méthode : La deuxième méthode, tout en n'étant pas idéale, est possible avec l'ensemble de données de l'ESCC 2.2. Pour utiliser cette méthode, l'utilisateur doit faire l'hypothèse que la variabilité intra-individuelle de la consommation de suppléments est nulle, ce qui serait le cas pour les répondants qui ont consommé la même quantité du même supplément chaque jour. Avec cette méthode, l'utilisateur devra :
 - Additionner l'apport quotidien en aliments (premier interview de rappel) et l'apport moyen en suppléments.
 - Additionner l'apport quotidien en aliments (deuxième interview de rappel) et l'apport moyen en suppléments.
 - Rajuster le premier interview de rappel en fonction du deuxième.

La principale limite de cette approche est qu'elle suppose que la variance intra-individuelle de la consommation de suppléments est nulle, ce qui n'est pas le cas. La mesure dans laquelle la variance intra-personnelle existe dans l'ESCC 2.2 est inconnue et ne peut être

évaluée sans une mesure de l'apport habituel et quotidien. L'utilisateur doit, au minimum, examiner les habitudes de consommation de suppléments contenant le nutriment d'intérêt afin d'évaluer la validité de l'hypothèse.

Cette méthode suppose en outre que tous les suppléments sont absorbés de la même manière. Comme nous utilisons un apport quotidien moyen pendant une période de 30 jours, l'hypothèse est que la prise d'une dose quotidienne équivaut à la prise de la dose équivalente hebdomadaire, ce qui peut être le cas ou non, selon l'nutriment et le supplément.

3. Troisième méthode : Comme pour la deuxième approche, l'utilisateur doit supposer, s'il désire appliquer cette troisième méthode, que la consommation de suppléments est la même chaque jour (la variance intra-individuelle est nulle). Pour cette méthode, l'utilisateur devra :
 - Rajuster le premier interview de rappel en fonction du deuxième.
 - Additionner l'apport en aliments ajusté et l'apport moyen en suppléments (non ajusté)

La différence entre cette méthode et la précédente est le moment de l'ajustement pour calculer l'apport habituel à partir de l'apport quotidien. Avec la deuxième méthode, les apports en aliments et en suppléments sont additionnés *avant* le rajustement du premier interview de rappel en fonction du deuxième. Dans le cas présent, les apports en aliments et en suppléments sont additionnés *après* le rajustement pour l'apport habituel.

Comme pour la deuxième méthode, la principale limite de cette approche est que la variance intra-individuelle n'est pas nulle. Dans ce cas également, nous supposons que tous les suppléments sont absorbés de la même manière. De plus, l'utilisateur suppose aussi que les apports en aliments et en suppléments sont indépendants. Nous savons que cela n'est pas nécessairement vrai, puisque les recherches antérieures ont démontré que les consommateurs de suppléments sont plus susceptibles d'avoir des apports en nutriments adéquats. Enfin, cette méthode exige également l'utilisation d'un apport habituel individuel. Même si le logiciel SIDE peut être utilisé pour dériver cet apport habituel individuel, le résultat pourra être biaisé, particulièrement avec seulement deux rappels alimentaires. Selon la théorie, avec deux seuls rappels alimentaires, on ne devrait faire que des répartitions de l'apport habituel global.

S'il est possible de combiner les données sur les aliments et les suppléments, les utilisateurs de données doivent être au fait des problèmes entourant ce type d'analyse. L'analyste devra faire des hypothèses qui ont une incidence sur l'analyse et l'interprétation des résultats. Les utilisateurs de données sont priés de consulter les documents suivants, qui décrivent plus en détail les méthodes exposées ci-dessus.

Références

British Columbia Ministry of Health Services: British Columbia Nutrition Survey, Report on Energy and Nutrient Intakes. 2004.

Carriquiry, AL: Estimation of Usual Intake Distributions of Nutrients and Foods. The Journal of Nutrition 2003: 601S-608S.

Dwyer J, Picciano MF, Raiten DJ and Members of the Steering Committee: Estimation of Usual Intakes: What we Eat in America—NHANES. *The Journal of Nutrition* 2003: 609S-623S.

2. État de sécurité alimentaire des ménages

Introduction

Depuis 2004, Statistique Canada recueille des données détaillées sur la sécurité alimentaire liée au revenu dans les ménages au cours de certains cycles de l'Enquête sur la santé dans les collectivités canadiennes (ESCC). Le module de la sécurité alimentaire qui est inclus dans ces cycles de l'ESCC, c'est-à-dire le module d'enquête sur la sécurité alimentaire des ménages (MESAM), a été adapté du module d'enquête sur la sécurité alimentaire élaboré aux États-Unis. Le MESAM contient 18 questions sur la situation de sécurité alimentaire du ménage au cours des 12 mois qui ont précédé l'enquête. Dix questions ont trait à la situation des membres adultes du ménage ou à celle du ménage en général, et les huit autres concernent la situation des enfants présents dans le ménage. Le niveau de sécurité alimentaire est déterminé d'après le nombre de réponses affirmatives à ces questions. La variable dérivée originale de situation de sécurité alimentaire du ménage qui est incluse dans l'ensemble de données de l'ESCC (c.-à-d. FSCnDHFS) est basée sur la méthode standard appliquée pour la publication des données aux États-Unis. Pour plus de renseignements sur cette méthode de mesure de la sécurité alimentaire du ménage et sur la façon dont cette variable est dérivée, veuillez consulter la documentation sur les variables dérivées.

Comparativement à la variable de sécurité alimentaire dérivée qui est publiée dans le fichier de données du cycle 2.2 de l'ESCC (basée sur la méthode standard américaine), Santé Canada a utilisé une autre méthode pour produire les estimations de la situation de sécurité alimentaire publiées dans le rapport de 2007 intitulé *Enquête sur la santé dans les collectivités canadiennes, cycle 2.2, Nutrition (2004) – Sécurité alimentaire liée au revenu dans les ménages canadiens*. Une brève description des différences entre la méthode standard américaine de calcul du niveau de sécurité alimentaire et de la méthode suivie par Santé Canada figure plus bas. Les utilisateurs de ces données sont invités à consulter le rapport *Sécurité alimentaire liée au revenu dans les ménages canadiens* pour obtenir plus de renseignements sur la méthodologie utilisée par Statistique Canada.

Principaux changements méthodologiques dans la détermination de la situation de sécurité alimentaire : Santé Canada

La méthode appliquée par Santé Canada mène à trois catégories de situation de sécurité alimentaire comparativement aux quatre qui sont incluses dans l'ensemble de données de l'ESCC en se fondant sur la méthode standard utilisée originalement aux États-Unis. Les trois catégories de situation de sécurité alimentaire établies conformément à la méthode de Santé Canada sont les suivantes :

1. **sécurité alimentaire** : aucun ou un seul signe de difficulté d'accès à des aliments en raison du revenu.
2. **insécurité alimentaire modérée** : signes que la qualité et/ou la quantité des aliments consommés est compromise.
3. **insécurité alimentaire grave** : signes de réduction de l'apport alimentaire et de perturbation des habitudes alimentaires.

La méthode de Santé Canada diffère de la méthode standard américaine à deux égards :

- Pour déterminer la situation de sécurité alimentaire des ménages comptant des enfants, dans la méthode standard utilisée aux États-Unis, les 18 questions du MESAM

sont évaluées sur une échelle unique. Inversement, dans la méthode de Santé Canada, on détermine d'abord la situation de sécurité alimentaire des membres adultes du ménage et celle des enfants faisant partie du ménage. Si les adultes ou les enfants faisant partie du ménage sont en situation d'insécurité alimentaire, le ménage est considéré comme en état d'insécurité alimentaire. Cette modification devrait, en soi, produire des estimations un peu plus faibles de l'insécurité alimentaire dans les ménages comparativement à la méthode américaine.

- Selon la méthode de Santé Canada, le seuil d'insécurité alimentaire est fixé à deux réponses affirmatives, soit un seuil plus faible que celui utilisé habituellement aux États-Unis (c.-à-d. trois). Ce changement devrait, en soi, produire des estimations plus élevées de l'insécurité alimentaire que la méthode américaine standard.

Analyse des données : Utilisation de pondérations statistiques

L'analyse des variables de sécurité alimentaire en utilisant la variable de poids de la personne produit des estimations du nombre de personnes qui vivent dans les ménages ayant une situation de sécurité alimentaire donnée, ainsi que les taux de prévalence connexes. **Il importe de souligner que cela ne revient pas à estimer le nombre de personnes ayant une situation de sécurité alimentaire donnée, un concept qu'il est impossible de mesurer à l'aide du MESAM.**

Dans le cas du cycle 2.2 de l'ESCC, il est également possible d'analyser les données en utilisant la variable de poids du ménage. L'analyse avec ce poids produit des estimations du nombre de ménages ayant une situation de sécurité alimentaire donnée, ce qui est approprié car les données recueillies à l'aide du MESAM reflètent le ménage et non les répondants individuels.

À l'heure actuelle, les poids du ménage pour le MESAM ne sont disponibles que pour le cycle 2.2 de l'ESCC. Toutefois, ils le seront bientôt pour le cycle 3.1, pour l'analyse au niveau de la province ou du territoire uniquement. Pour des raisons méthodologiques, les poids des ménages ne peuvent être produits pour les régions sociosanitaires. À partir du cycle 4.1, les poids des ménages au niveau provincial ou territorial deviendront un produit ordinaire.

Pour plus de renseignements concernant l'utilisation des poids des ménages, veuillez consulter la documentation sur le cycle 2.2 de l'ESCC publiée par Statistique Canada.

Pour plus de renseignements

Si vous avez des questions ou avez besoin d'autres renseignements, veuillez envoyer un courriel à cnha_acns@hc-sc.gc.ca ou cchs-escc@statcan.ca