

ENQUÊTE NATIONALE SUR LA SANTÉ DE LA POPULATION

VOLET MÉNAGES

CYCLE 5 (2002-2003)

DOCUMENTATION LONGITUDINALE

STATISTIQUE CANADA

24 novembre 2004

NOTE AUX UTILISATEURS

La documentation longitudinale du cycle 5 de l'Enquête nationale sur la santé de la population (ENSP) fournit un large éventail d'information sur l'enquête: objectifs, développement de contenu, plan d'échantillonnage, collecte, traitement, pondération, qualité, totalisation et accès aux données. Les chapitres 7 et 8 donnent plus de détails concernant les différents sous-ensembles de répondants et le poids qui leur est associé.

Ce document fait référence aux différents cycles de l'ENSP en utilisant parfois les années du cycle en question. À titre de référence, voici la correspondance entre les cycles et les années.

Cycle 1 = 1994-1995

Cycle 2 = 1996-1997

Cycle 3 = 1998-1999

Cycle 4 = 2000-2001

Cycle 5 = 2002-2003

Ce guide s'adresse également aux utilisateurs du fichier partagé de microdonnées, c'est-à-dire les ministères provinciaux de la santé et Santé Canada. Le fichier partagé inclut les répondants qui ont accepté de partager leurs données au cycle 5 de l'ENSP de même que le poids d'échantillonnage correspondant. Ce groupe de répondants est un sous-ensemble du fichier maître de microdonnées. Les utilisateurs du fichier partagé ne doivent pas tenir compte des références qui sont faites concernant les autres sous-ensembles de répondants.

TABLE DES MATIÈRES

	Page
1. Introduction.....	1
2. Contexte.....	2
3. Objectifs.....	3
4. Contenu de l'enquête.....	4
4.1 Critères de sélection.....	4
4.2 Modifications apportées au contenu du cycle 5 (2002-2003).....	5
4.3 Rétroaction des données du cycle 4 (2000-2001) et questions de suivi.....	6
5. Plan d'échantillonnage.....	7
5.1 Plan d'échantillonnage pour le cycle 1 (1994-1995).....	7
5.1.1 Répartition de l'échantillon.....	7
5.1.2 Méthode d'exclusion.....	8
5.1.3 Prélèvement de l'échantillon.....	8
5.1.4 Plan d'échantillonnage au Québec.....	10
5.2 Échantillon longitudinal.....	11
6. Collecte des données.....	13
6.1 Conception du questionnaire et méthode de collecte des données.....	13
6.2 Essais.....	13
6.3 Interviews.....	13
6.4 Non-réponse et dépistage.....	14
7. Traitement des données.....	16
7.1 Vérification.....	16
7.2 Antécédents résidentiels, un nouveau module au cycle 5.....	16
7.3 Codage.....	16
7.4 Création de variables dérivées et groupées.....	17
7.5 Estimation et pondération.....	17
7.6 Sous-ensembles de répondants.....	18
7.7 Définition d'une réponse complète.....	19
8. Pondération.....	20
8.1 Pondération longitudinale.....	20
8.1.1 Point de départ: poids épurés du cycle 1 (1994-1995).....	21
8.1.2 Ajustements apportés pour créer les différents poids longitudinaux.....	21
8.1.2.1 Poids longitudinal carré (WT64LS).....	21
8.1.2.2 Poids longitudinal complet (WT62LF).....	21
8.1.2.3 Poids longitudinal complet cycles 1 et 5 (WT62LFE).....	28
8.1.2.4 Poids longitudinal complet partagé (WT62SLF).....	29

9.	Qualité des données	31
9.1	Erreurs dues à l'échantillonnage	31
9.1.1	Méthode bootstrap pour l'estimation de la variance	32
9.2	Erreurs non dues à l'échantillonnage	33
9.2.1	Taux de réponse	33
9.2.1.1	Taux de réponse du cycle 1 (1994-1995)	34
9.2.1.2	Taux de réponse du cycle 2 (1996-1997)	35
9.2.1.3	Taux de réponse du cycle 3 (1998-1999)	36
9.2.1.4	Taux de réponse du cycle 4 (2000-2001)	37
9.2.1.5	Taux de réponse du cycle 5 (2002-2003)	38
9.2.2	Taux de refus	39
9.2.3	Taux de conversion des refus	40
9.2.4	Taux d'érosion.....	41
9.2.5	La non-réponse attribuable aux personnes non-dépistées.....	42
9.2.6	Taux de refus et de «Ne sait pas» par item	43
10.	Lignes directrices sur les totalisations, l'analyse et la diffusion	45
10.1	Lignes directrices sur l'arrondissement.....	45
10.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation.....	46
10.2.1	Définitions des genres d'estimations : de type nominal par opposition à quantitatif	46
10.2.2	Totalisation d'estimations de type nominal	47
10.2.3	Totalisation d'estimations quantitatives	47
10.3	Lignes directrices pour l'analyse statistique	48
10.4	Lignes directrices pour la diffusion.....	49
11.	Utilisation du fichier longitudinal.....	50
11.1	Utilisation des poids longitudinaux.....	50
11.2	Assurance de la fiabilité des estimations avec l'utilisation des poids bootstrap.....	50
11.3	Convention utilisée pour nommer les variables	51
11.3.1	Structure élémentaire des noms des variables.....	51
11.3.2	Positions 1 et 2 : Nom de la variable / section du questionnaire.....	52
11.3.3	Position 3 : Type d'enquête.....	53
11.3.4	Position 4 : Variable de l'année de référence du cycle	54
11.3.5	Position 5 : Type de variable.....	54
11.3.6	Position 6 à 8 : Nom de la variable	54
12.	Accès aux données de l'ENSP.....	56
12.1	Fichiers de microdonnées	56
12.2	Rapports analytiques et totalisations	57

Liste des tableaux

Tableau 5.A	Taille de l'échantillon longitudinal par province
Tableau 7.A	Sous-ensembles de répondants
Tableau 8.A	Sous-ensembles de répondants, variables de poids et variables indicatrices correspondantes.
Tableau 8.B	Variables pour l'ajustement de la non-réponse au cycle 2 - WT62LF
Tableau 8.C	Variables pour l'ajustement de la non-réponse au cycle 3 - WT62LF
Tableau 8.D	Variables pour l'ajustement de la non-réponse au cycle 4 - WT62LF
Tableau 8.E	Variables pour l'ajustement de la non-réponse au cycle 5 - WT62LF
Tableau 9.A	Taux de refus par cycle
Tableau 9.B:	Taux de conversion des refus
Tableau 9.C:	Source de l'érosion par cycle
Tableau 9.D	Dépistage, cycles 2, 3, 4 et 5
Tableau 9.E	Taux de refus et de ne sait pas au niveau de la personne
Tableau 10.A	Lignes directrices relatives à la variabilité d'échantillonnage
Tableau 11.A	Sous-ensembles de répondants et fichiers de poids bootstrap correspondants
Tableau 11.B	Variables longitudinales «constantes»

Liste des annexes

Annexe A :	ENSP, volet ménages, contenu du questionnaire au cours des 5 cycles
Annexe B :	ENSP, volet ménages, changements au questionnaire pour le cycle 5 (2002-2003)
Annexe C :	ENSP, volet ménages, exemples de rétroaction des données du cycle 4 (2000-2001) et de questions de suivi

Autres documents de référence

Questionnaire
Cliché d'article
Index alphabétique
Index par sujet
Dictionnaire de données
Documentation des variables dérivées, ENSP, cycles 1 à 5

1. Introduction

L'Enquête nationale sur la santé de la population (ENSP) vise à recueillir des renseignements «longitudinaux» sur la santé de la population canadienne ainsi que des renseignements socio-démographiques connexes. Le premier cycle de l'enquête a eu lieu en 1994-1995 et elle aura lieu tous les deux ans pendant dix cycles. L'ENSP a recueilli des renseignements de nature transversale et longitudinale durant les trois premiers cycles mais depuis le cycle 4 (2000-2001) l'enquête est devenue uniquement longitudinale. Des renseignements de nature transversale sont désormais recueillis grâce à l'Enquête sur la santé dans les collectivités canadiennes (ESCC). L'ENSP compte maintenant deux volets : l'enquête auprès des ménages et celle auprès des établissements de soins de santé. À partir du cycle 4, l'enquête dans le Nord est uniquement transversale et est menée par l'ESCC plutôt que l'ENSP.

Le volet ménages de l'ENSP vise les membres des ménages de toutes les provinces en 1994-1995, à l'exclusion de la population des réserves indiennes et terres de la Couronne, des résidents des établissements de soins de santé, des membres à temps plein des bases des Forces canadiennes et de certaines régions éloignées du Québec et de l'Ontario. Le volet établissements de soins de santé inclut les résidents séjournant à long terme dans des établissements de soins de santé (séjours de plus de six mois) comportant au moins quatre lits, dans toutes les provinces à l'exclusion des territoires, des réserves indiennes et des membres à temps plein des bases des Forces canadiennes. Cinq cycles de collecte ont été complétés pour le volet ménages de l'ENSP : l'ENSP cycle 1 (1994-1995), l'ENSP cycle 2 (1996-1997), l'ENSP cycle 3 (1998-1999), l'ENSP cycle 4 (2000-2001) et l'ENSP cycle 5 (2002-2003).

Le volet ménages du cycle 5 de l'ENSP a recueilli des renseignements détaillés sur la santé des répondants longitudinaux sélectionnés au hasard lors du premier cycle de même que des renseignements démographiques sur tous les membres du ménage des répondants longitudinaux. Le questionnaire comprend, entre autres, des questions sur l'état de santé, l'utilisation des services de santé, les déterminants de la santé, les problèmes de santé chroniques et les limitations d'activité. On recueille également des renseignements socio-démographiques tels que l'âge, le sexe, le niveau de scolarité, le revenu du ménage et l'activité sur le marché du travail.

Le présent document a pour but de faciliter l'utilisation des fichiers maître et partagé longitudinaux du cycle 5, du volet ménages de l'ENSP. Ces fichiers sont décrits de façon plus détaillée dans les chapitres suivants. Pour toute question au sujet des données ou leur utilisation, veuillez vous adresser à :

Renseignements sur le contenu et l'accès aux données de l'enquête :

Section ENSP, Division de la statistique de la santé	Téléphone :	1-613-951-1653
	Télécopieur :	1-613-951-4198
	Courriel :	nphs-ensp@statcan.ca

Totalisations personnalisées ou renseignements généraux sur les données :

Services personnalisés à la clientèle	Téléphone :	1-613-951-1746
Division de la statistique de la santé	Courriel :	hd-ds@statcan.ca

2. Contexte

À l'automne 1991, le Conseil national d'information sur la santé a recommandé la réalisation d'une enquête permanente nationale sur la santé de la population. Cette recommandation découlait des pressions économiques et fiscales exercées sur les régimes de soins de santé et du besoin correspondant d'information pouvant servir à améliorer l'état de santé de la population canadienne. Or, les sources existantes de données sur la santé ne fournissaient pas suffisamment de renseignements pour brosser un tableau complet de l'état de santé de la population et de la multitude des facteurs qui ont un effet sur la santé.

Dès le mois d'avril 1992, Statistique Canada a obtenu les fonds nécessaires pour élaborer l'Enquête nationale sur la santé de la population qui a été conçue en prenant pour critères la souplesse de l'enquête ainsi que la validité, la fiabilité et l'actualité des données. Il fallait aussi que l'enquête puisse répondre à l'évolution des besoins, des politiques et des intérêts.

3. Objectifs

Les objectifs de l'ENSP sont les suivants :

- faciliter l'élaboration de politiques gouvernementales en fournissant des données sur l'état de santé de la population, une ventilation de ces données ainsi qu'un aperçu des tendances observées,
- fournir des données qui permettent de réaliser des études analytiques et qui aident à comprendre les déterminants de la santé,
- recueillir des données sur les facteurs économiques, sociaux, démographiques, professionnels et environnementaux ayant une influence sur la santé,
- permettre de mieux comprendre la relation entre l'état de santé et l'utilisation des services de santé, y compris les services parallèles et les services traditionnels,
- fournir des renseignements sur un panel de personnes qui seront suivies dans le temps pour refléter la dynamique de la santé et de la maladie,
- donner aux provinces, aux territoires et à d'autres clients une capacité d'enquête sur la santé qui leur permettra de compléter le contenu ou de grossir l'échantillon,
- permettre de coupler les données de l'enquête à des données administratives courantes, telles que les statistiques de l'état civil, les mesures environnementales, les variables liées aux collectivités ainsi que l'utilisation des services de santé.

4. Contenu de l'enquête

Les objectifs décrits au chapitre 3 représentent l'orientation générale de l'ENSP, notamment en ce qui concerne le type de renseignements à recueillir. La première section de ce chapitre présente les critères généraux utilisés pour choisir le contenu de l'enquête et résume le contenu des diverses sections du questionnaire. La deuxième section décrit en détail les modifications apportées au contenu de l'enquête pour réaliser le cycle 5 de l'ENSP ainsi que son nouveau contenu. La dernière section présente les variables du cycle 4 qui ont été utilisées lors de la collecte du cycle 5. De l'information détaillée additionnelle sur le contenu de l'enquête de 2002-2003 et sur les changements de contenu durant les quatre premiers cycles est disponible dans les annexes A, B et C.

4.1 Critères de sélection

On a choisi le contenu de l'enquête conformément aux critères suivants :

- 1) les renseignements doivent être en rapport avec les objectifs des provinces et des territoires en matière de santé et faciliter le contrôle de ces objectifs. Dans les cas où ces objectifs n'existent pas, par exemple au niveau national, on peut déterminer le contenu de l'enquête à partir des politiques et des programmes,
- 2) les renseignements ne doivent pas faire double emploi avec ceux provenant d'autres sources,
- 3) pour mieux faire comprendre la santé et ses déterminants, l'enquête doit fournir de nouveaux renseignements dans des domaines qui n'ont pas encore été étudiés de manière approfondie,
- 4) l'enquête doit porter principalement sur les comportements ou sur les problèmes de santé qui se prêtent à la prévention, au traitement ou à l'intervention,
- 5) l'enquête doit permettre de recueillir des renseignements sur les problèmes de santé qui causent le plus de souffrance ou qui coûtent le plus cher aux personnes touchées, à l'ensemble de la population ou aux régimes de soins santé,
- 6) l'enquête doit fournir des renseignements sur les facteurs liés non seulement à la maladie mais aussi à la bonne santé.

Lors des trois premiers cycles un certain nombre de renseignements ont été recueillis auprès de tous les membres de chaque ménage (composante Générale - H05, par exemple les caractéristiques socio-démographiques et les renseignements sur l'activité relative au marché du travail). De plus, au cycle 1 une personne a été choisie au hasard dans chaque ménage (répondant longitudinal) pour une interview plus approfondie (composante Santé - H06, par exemple des renseignements sur l'état de santé). À partir du cycle 4, les questionnaires des composantes Générale et Santé ont été combinés en un seul questionnaire. Le répondant longitudinal répond désormais aux questions portant sur le ménage.

Selon les critères susmentionnés, le questionnaire porte sur l'état de santé, l'utilisation des services de santé, les déterminants de la santé, les problèmes de santé chroniques et les limitations d'activité, et la situation démographique et socio-économique. Par exemple, on évalue l'état de santé à l'aide de questions sur la perception personnelle de la santé, la capacité fonctionnelle, les problèmes de santé chroniques et la limitation des activités. L'utilisation des services de santé est mesurée au moyen de questions sur les visites faites chez des professionnels de la santé (médecine traditionnelle et médecine alternative), les soins hospitaliers et la consommation de médicaments et autres remèdes. Les questions sur les déterminants de la santé ont trait à l'usage du tabac, la consommation d'alcool et l'activité physique. Il y a en outre des questions sur les examens préventifs qui demandent la fréquence et les raisons d'utilisation de ces services. Les questions sur les caractéristiques démographiques et socio-économiques couvrent, entre autres, l'âge, le sexe, le niveau de scolarité, le revenu du ménage et l'activité sur le marché du travail.

Une copie du questionnaire du cycle 5 (2002-2003) de l'ENSP volet ménages est fournie avec la documentation et est disponible sur le site Web de Statistique Canada <http://www.statcan.ca>, en cliquant sur « Définition, sources de données et méthodes », ensuite « Questionnaire », puis « Liste par sujet des questionnaires », puis « Santé » et enfin « Enquête nationale sur la santé de la population, Volet ménages, Cycle 5 (2002-2003) - Questionnaire ». Veuillez consulter le chapitre 12 pour d'autres détails relatifs à l'information disponible sur le site Web de Statistique Canada.

4.2 Modifications apportées au contenu du cycle 5 (2002-2003)

L'ordre du questionnaire n'a pas changé par rapport au cycle 4. Les questions portant sur la santé sont posées en premier et sont suivies de celles sur les données socio-économiques (langue, niveau de scolarité, état de la population active et revenu).

Le contenu thématique et le contenu acheté du cycle 5 ont été incorporés à la partie la plus appropriée du questionnaire. Le contenu thématique du cycle 5 porte sur le sommeil, l'image corporelle, la nutrition et les opinions et la dépendance à l'égard de l'alcool. Le contenu acheté porte sur les antécédents résidentiels depuis 1980. Quelques questions sur le diagnostic et la façon de gérer les problèmes de santé chroniques ont été reformulées ou supprimées.

Pour plus d'information, voir les annexes A et B. L'annexe A présente par thème et par sujet le contenu du questionnaire de l'ENSP au cours des cinq premiers cycles. L'annexe B présente en détail, les changements apportés au questionnaire du cycle 5 de l'ENSP.

4.3 Rétroaction des données du cycle 4 (2000-2001) et questions de suivi

Afin de réduire le fardeau de réponse, les questions posées lors des cycles précédents et dont la réponse ne peut changer dans le temps (le pays de naissance par exemple) n'ont pas été posées à nouveau. Les variables pour lesquelles la réponse peut changer au cours du temps, mais uniquement dans certaines conditions (par exemple, le niveau de scolarité) ont été mises à jour au besoin. Dans certains cas, les réponses recueillies lors des cycles précédents pouvaient être rappelées au répondant lors de l'interview du cycle 5, et ce afin d'améliorer la qualité des données recueillies. Par exemple, les renseignements sur certaines conditions chroniques obtenus lors de l'interview du cycle 4 ont été utilisés pour aider à expliquer un changement au cycle 5. Voir l'annexe C pour plus de détails.

5. Plan d'échantillonnage

La population cible du volet ménages de l'ENSP vise les membres des ménages de toutes les provinces canadiennes en 1994-1995 à l'exclusion des personnes habitant dans une réserve indienne ou sur des terres de la Couronne, des résidents des établissements de soins de santé, des membres à temps plein des bases des Forces canadiennes et des personnes habitant certaines régions éloignées de l'Ontario et du Québec. Ce chapitre décrit le plan d'échantillonnage utilisé au cycle 1 et explique comment l'échantillon longitudinal de 17 276 personnes a été obtenu.

5.1 Plan d'échantillonnage pour le cycle 1 (1994-1995)

L'Enquête sur la population active (EPA), remaniée en 1991, a servi de base pour le plan d'échantillonnage dans toutes les provinces, à l'exception du Québec où l'échantillon a été choisi à partir des ménages déjà interviewés par Santé Québec pour l'Enquête sociale et de santé de 1992-1993.

Le plan d'échantillonnage du volet ménages a été établi selon trois critères :

- la taille visée des échantillons nationaux et provinciaux/territoriaux,
- la décision de choisir un membre par ménage pour constituer le panel longitudinal, et
- le choix de l'EPA comme véhicule pour sélectionner l'échantillon.

Ces trois facteurs ont déterminé, respectivement, la répartition de l'échantillon, l'application d'une technique pour accroître la représentativité (la « méthode d'exclusion » décrite plus loin) et la sélection des échantillons provinciaux à l'extérieur du Québec.

5.1.1 Répartition de l'échantillon

L'ENSP visait initialement un échantillon d'une taille de 19 600 ménages. Par ailleurs, les représentants nationaux et provinciaux ont convenu qu'il fallait un minimum de 1 200 ménages par province. Sous réserve de cette restriction, la taille des échantillons provinciaux a été établie à l'aide d'une méthode de répartition bien connue qui met en équilibre les exigences de fiabilité aux niveaux national et régional (Kish, 1988)¹. Selon cette méthode, l'échantillon a été réparti proportionnellement à $\sqrt{(0.804W_h^2 + 1/12^2)}$, où W_h représente la proportion des ménages dans la province/le territoire h , $h=1, \dots, 12$, d'après le Recensement de 1991. Cette répartition a permis de déterminer la taille de l'échantillon principal de chaque province. Quatre provinces ont choisi d'acheter des unités d'échantillonnage additionnelles pour augmenter l'échantillon du cycle 1 qui leur avait été alloué, mais pour des fins transversales seulement. Ces unités additionnelles ne font pas partie de l'échantillon longitudinal.

¹ Kish, L. (1988). Multipurpose Sample Design, Survey Methodology, 14, 19-32

5.1.2 Méthode d'exclusion

Le contenu de l'enquête vise principalement un membre de chaque ménage échantillonné, choisi au hasard, pour faire partie du panel longitudinal. Sans le recours à la méthode d'exclusion, les personnes appartenant à des ménages de grande taille, notamment les parents et leurs enfants, seraient sous-représentées dans le panel tandis que les personnes appartenant à des ménages de petite taille, en général les personnes célibataires et les personnes âgées, seraient sur-représentées.

Par conséquent, une méthode d'exclusion pour accroître la représentativité des parents et des jeunes dans le panel a été adoptée. Pour ce faire, une partie de l'échantillon a été soumise à une sélection préliminaire. Une fois la liste des membres de ces ménages dressée, les ménages n'ayant aucun membre âgé de moins de 25 ans ont été considérés comme remplissant les conditions de rejet et exclus de l'enquête. Afin de maintenir la taille voulue des échantillons, on a ajouté le nombre prévu de ménages exclus au nombre de ménages visités dans chaque province.

La méthode d'exclusion des ménages qui n'ont aucun membre âgé de moins de 25 ans a été adoptée parce qu'elle était plus efficace que d'autres méthodes qui ont été envisagées. Pour des raisons financières et opérationnelles, le pourcentage de ménages soumis à la sélection préliminaire a généralement été limité à 25-30 % en Ontario. Il était à 37,5-40 % dans les régions urbaines des autres provinces et à 25-30 % dans toutes les régions rurales. Étant donné que les strates d'appartements avaient une forte concentration de petits ménages, on a réduit la taille de ces échantillons plutôt que d'y appliquer la méthode d'exclusion. De plus, la méthode d'exclusion n'a pas été appliquée dans les régions éloignées en raison des coûts élevés pour contacter les ménages à ces endroits.

5.1.3 Prélèvement de l'échantillon

Un plan d'échantillonnage stratifié à plusieurs degrés a été utilisé pour le volet ménages de l'ENSP. Le premier degré consistait à former des strates homogènes et à prélever des échantillons indépendants de grappes dans chaque strate. Le deuxième degré consistait à dresser la liste de logements présents dans chaque grappe choisie et à en sélectionner un certain nombre à partir de cette liste.

Dans chaque province, sauf au Québec, l'ENSP a fait appel à la méthode d'échantillonnage polyvalente mise au point pour le remaniement de l'EPA. Les échantillons de logements répartis en grappes qui sont constitués selon cette méthode peuvent servir à des enquêtes ménages générales, ce qui réduit considérablement les frais de listage et de collecte de données.

Le plan de sondage de l'EPA est constitué d'un échantillon stratifié à plusieurs degrés composé de logements sélectionnés parmi les grappes. Chaque province est divisée en trois types de régions (grands centres urbains, régions urbaines et régions rurales) à partir desquelles sont formées des strates géographiques ou socio-économiques distinctes. Dans la plupart des strates, on choisit six grappes, généralement des secteurs de dénombrement (SD), selon une méthode d'échantillonnage avec probabilité proportionnelle à la taille (PPT). Dans quelques cas où la densité de la population est faible, on ajoute un degré supplémentaire en sélectionnant d'abord deux ou trois unités primaires d'échantillonnage, en les divisant en grappes et en prélevant un échantillon de six grappes à l'intérieur de chacune. Dans tout le plan d'échantillonnage, on utilise six grappes afin de pouvoir renouveler chaque mois le sixième de l'échantillon aux fins de l'EPA.

L'échantillon de logements est obtenu une fois que les opérations de listage des grappes sont terminées. Étant donné que les taux d'échantillonnage sont déterminés au préalable, il y a souvent un écart entre la taille prévue de l'échantillon et le nombre d'unités obtenu. Le rendement excessif de l'échantillon est corrigé en supprimant une partie des unités sélectionnées à l'origine. Cette opération, qui se fait habituellement à un niveau agrégé, est appelée stabilisation de l'échantillon. De plus, on augmente la taille de l'échantillon afin de représenter les logements plutôt que les ménages, étant donné qu'un certain nombre de non-réponse était attendu, et qu'une partie des logements sont habituellement vacants ou hors du champ de l'enquête.

Le plan d'échantillonnage de l'EPA est établi de manière à obtenir environ 60 000 ménages. Pour les enquêtes qui exigent des échantillons de moindre taille, on « réserve » habituellement de un à six renouvellements par province, un renouvellement correspondant au sixième de l'échantillon total. On procède à la stabilisation de l'échantillon pour le maintenir au niveau voulu, notamment lorsqu'on réserve deux renouvellements mais que la taille de l'échantillon requise représente seulement 1,5 renouvellement.

En raison des besoins particuliers de l'ENSP, deux modifications ont été apportées à cette stratégie d'échantillonnage. Le nombre de renouvellements « réservés » a été déterminé à l'échelle des strates plutôt qu'à l'échelle provinciale afin de répondre aux exigences relatives à la répartition infra-provinciale de l'échantillon pour des fins transversales au premier cycle. Il fallait aussi que le nombre de grappes sélectionnées par strate soit un multiple de quatre, aux fins de l'estimation de la variance et de la représentativité saisonnière (les strates pouvaient ainsi être réparties en deux ou plusieurs échantillons indépendants de quatre grappes chacun - soit un par période de collecte). Étant donné que l'ENSP ne nécessitait généralement que de deux à six grappes par strate de l'EPA, des strates similaires de l'EPA ont été regroupées de manière à former de plus grandes strates de l'ENSP comportant le nombre requis de grappes de l'échantillon. Une fois les strates regroupées, leurs grappes ont, elles aussi, été regroupées pour former des échantillons répétés.

À la suite de ces modifications, l'échantillon de grappes de l'ENSP peut être considéré comme un échantillon stratifié répété dont les strates sont constituées de groupes de strates de l'EPA et dont les échantillons répétés sont habituellement des échantillons indépendants, répartis de manière identique et composés de quatre grappes chacun. Les quelques exceptions à cette règle ne devraient pas avoir d'incidence appréciable sur les résultats de l'enquête. Deux variables de plan, « stratum » et « replicat », se trouvent dans le fichier maître. Elles représentent la strate de l'EPA et l'échantillon répété de l'ENSP.

5.1.4 Plan d'échantillonnage au Québec

Au Québec, l'échantillon de l'ENSP a été prélevé parmi les logements qui ont participé à l'Enquête sociale et de santé (ESS) réalisée par Santé Québec en 1992-1993. Cette enquête a porté sur 16 010 logements qui ont été sélectionnés selon un plan d'échantillonnage à deux degrés semblable à celui de l'EPA. La province a été divisée géographiquement en recoupant quinze secteurs sanitaires avec quatre classes de densité urbaine (la région métropolitaine de recensement de Montréal, les capitales régionales, les petites agglomérations urbaines et le secteur rural). Dans chaque secteur, les grappes ont été stratifiées selon des caractéristiques socio-économiques et sélectionnées à l'aide d'une méthode d'échantillonnage PPT. On a dénombré les grappes sélectionnées et on a prélevé des échantillons aléatoires de logements parmi celles-ci, à raison de 10 logements par grappe dans les grandes villes et de 20 ou 30 logements ailleurs.

Santé Québec a fourni des renseignements non confidentiels qui ont permis de classer l'échantillon de l'ESS en quatre types de ménages : les ménages n'ayant qu'un seul membre, les ménages ayant des enfants, les autres ménages comptant des jeunes (personnes âgées de moins de 25 ans) et les autres (plus d'un membre mais sans enfant, ni jeune). Le type de ménages dans le cas des non répondants de l'ESS a été déterminé par le personnel de l'ENSP.

La taille de l'échantillon de l'ENSP a d'abord été répartie parmi les quatre classes de densité urbaine. Afin d'éviter la concentration de l'échantillon à Montréal, la répartition a été proportionnelle à $\sqrt{(2W_h^2 + 1/4^2)}$, où W_h représente la part de la population pour la classe h , $h=1,2,3,4$. Dans chaque classe, on a tenté d'obtenir un sous-échantillon de l'ESS, en ce qui concerne le membre sélectionné pour faire partie du panel, qui serait proportionnel aux populations des quatre types de ménages. Pour ce faire, on a tiré de l'ESS un nombre suffisant de ménages pour donner le rendement voulu de ménages ayant des enfants (le groupe le plus sous-représenté), après quoi on a supprimé la part excédentaire des trois autres groupes de ménages. On a donc prélevé un échantillon initial supérieur de près de 50 % à l'échantillon requis. Après avoir supprimé les deux tiers des ménages n'ayant qu'un seul membre, la moitié des autres ménages sans enfant ni jeune et le sixième des ménages ayant des jeunes mais pas d'enfant, l'objectif a été presque atteint.

Au Québec comme ailleurs, la représentation saisonnière, l'estimation de la variance et l'intégration avec l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ) ont eu une incidence sur le sous-échantillonnage. On a donc groupé des strates de l'ESS pour pouvoir former des échantillons répétés, les grappes de chacun couvrant les quatre trimestres (une grappe couvre deux trimestres dans les régions rurales et les petites agglomérations urbaines car la taille des échantillons y est plus élevée). L'échantillon des ménages ayant des enfants a été réparti, selon un rapport de 3 à 2, entre un échantillon « adultes » et un échantillon « enfants », ces termes ayant la même signification que dans les autres provinces. Les ménages de l'échantillon « enfants » des trimestres 1 et 2 ont été « réattribués » aux trimestres 3 et 4. Pour les trimestres 3 et 4, l'échantillon des ménages sans enfant a lui aussi été réparti selon un rapport de 2 à 3, entre un échantillon «adultes» et un échantillon « enfants », étant donné que l'ENSP vise les occupants actuels des logements choisis pour l'ESS et que des changements sont survenus dans certains de ces logements.

5.2 Échantillon longitudinal

L'échantillon longitudinal, aussi appelé panel longitudinal ou simplement panel, est composé des 17 276 personnes sélectionnées au cycle 1 qui ont au moins rempli la composante générale du questionnaire du cycle 1. Il comprend également les 2 022 enfants provenant du premier cycle de l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ). Ces enfants ont été interviewés par l'ELNEJ pour l'ENSP au cycle 1 et sont interviewés par l'ENSP depuis le deuxième cycle. C'est cet échantillon longitudinal qui a été suivi aux cycles 2, 3, 4 et 5 et qui sera suivi au cours des cycles à venir de l'ENSP. Il est à noter que les échantillons supplémentaires ajoutés aux cycles 1, 2 et 3 à des fins transversales ne font pas partie de l'échantillon longitudinal.

L'échantillon longitudinal ne se renouvelle pas au fil du temps. Aucun membre du panel n'a été et ne sera classé hors du champ de l'enquête. La taille de l'échantillon longitudinal demeurera donc égale à 17 276 pour tous les cycles. Par conséquent, au cycle 5, les membres du panel longitudinal étaient tous âgés d'au moins huit ans et l'échantillon longitudinal ne contenait aucune personne ayant immigré au Canada après 1994-1995.

Le nombre de personnes répondant à l'enquête diminue légèrement d'un cycle à l'autre en raison de l'érosion causée par les non-répondants, les refus et les personnes qui n'ont pu être retracées. Malgré l'érosion, l'échantillon longitudinal est toujours représentatif de la population de 1994-1995. L'érosion étant relativement faible (voir section 9.2.4), elle ne devrait pas entraîner de fortes augmentations de la variance des estimations. Il est à noter que les membres du panel de l'ENSP qui sont décédés ou qui ont déménagé en établissement de soins de santé font toujours partie de l'échantillon longitudinal et sont considérés comme étant répondants. Ces personnes ne contribuent donc pas à l'érosion du panel longitudinal de l'ENSP.

Le tableau 5.A présente la taille de l'échantillon longitudinal par province en 1994-1995 et montre également le nombre de personnes ayant répondu de façon complète aux cinq cycles de l'ENSP (ce compte comprend les personnes décédées).

Tableau 5.A : Taille de l'échantillon longitudinal par province

Province	Échantillon longitudinal Cycle 1 (1994-1995)	Nombre de répondants, réponse complète aux cycles 1, 2, 3, 4 et 5
Terre-Neuve	1 082	822
Île-du-Prince-Édouard	1 037	803
Nouvelle-Écosse	1 085	775
Nouveau-Brunswick	1 125	824
Québec	3 000	2 189
Ontario	4 307	2 990
Manitoba	1 205	921
Saskatchewan	1 168	922
Alberta	1 544	1 111
Colombie-Britannique	1 723	1 189
Total	17 276	12 546

6. Collecte des données

6.1 Conception du questionnaire et méthode de collecte des données

Les questions de l'enquête ont été conçues pour être posées selon la méthode d'interview assistée par ordinateur (IAO), ce qui veut dire que l'ordre logique des questions a été spécifié à mesure que celles-ci ont été élaborées. On a notamment spécifié le genre de réponse voulue, les valeurs minimales et maximales, les vérifications en ligne liées aux questions et la suite à donner en cas de non-réponse.

L'IAO permet de contrôler l'interview d'après les déclarations du répondant. Lorsqu'une réponse invalide est saisie, un message s'affiche à l'écran, ce qui permet au répondant ou à l'intervieweur d'obtenir une rétroaction immédiate pour corriger les incohérences. L'ajout automatique de la période de référence d'après la date du jour est un autre avantage. Il est possible de pré-insérer du texte ou des données selon les renseignements recueillis durant l'interview en cours et aussi durant les interviews des cycles précédents. Ceci permet à l'intervieweur de continuer son travail sans devoir revenir en arrière pour trouver des réponses antérieures. Cette possibilité permet notamment d'insérer le nom du répondant ou de faire les accords en genre à l'intérieur même des questions. Il est également possible de programmer des intervalles ou des réponses acceptables d'après les données recueillies durant l'interview. En d'autres termes, le questionnaire est adapté au répondant selon les données recueillies pendant l'interview en cours et lors des cycles précédents.

6.2 Essais

L'application IAO a été soumise à des essais à grande échelle à l'interne pour repérer toute erreur dans le cheminement du programme ou dans le texte. De plus, à chaque cycle, deux essais sur le terrain ont été effectués. Quatre des bureaux régionaux de Statistique Canada ont participé à ces essais. Ces deux essais visaient principalement à observer les réactions des répondants faces à de nouvelles questions ou nouveaux modules, à estimer la durée des diverses sections du questionnaire, à déterminer les taux de réponse et à évaluer les questions avec rétroaction. Les opérations et les procédures sur le terrain, la formation des intervieweurs et l'application IAO (c.-à-d., questionnaire informatisé) ont aussi été soumis à des essais.

6.3 Interviews

Au cycle 5, la collecte des données a été répartie en quatre trimestres (débutant en juin, août et octobre 2002, et en janvier 2003). Une autre période de collecte s'est tenue en juin 2003 avec des tentatives supplémentaires de suivi des non-répondants des trimestres précédents.

Les intervieweurs étaient des employés formés spécifiquement pour mener des enquêtes selon la méthode d'interview assistée par ordinateur et la majorité étaient des intervieweurs expérimentés de l'Enquête sur la population active (EPA). On a appliqué à la collecte des données de l'ENSP la structure de supervision et de vérification de l'EPA. Les intervieweurs ont suivi une formation ciblée sur le contenu de l'ENSP et ils ont reçu un manuel de l'intervieweur à utiliser comme outil de référence.

Tous les membres vivants du panel longitudinal ont reçu par courrier une lettre d'introduction à l'enquête indiquant la tenue éminente de la collecte du cycle 5 de l'ENSP. Ils ont également reçu une brochure présentant entre autres quelques résultats tirés des données de l'ENSP, un dépliant contenant des articles de journaux relatifs à l'ENSP et un aimant arborant le logo de l'enquête, le tout, pour les remercier.

En général, on communique d'abord par téléphone avec les répondants de l'échantillon des ménages. En fait, plus de 98 % des interviews au cycle 5 ont été effectuées par téléphone. On a procédé à des interviews en personne si le répondant n'avait pas le téléphone, dans le cadre d'un dépistage en personne, ou tout simplement à la demande du répondant. L'interview complète a duré en moyenne moins d'une heure.

Lors de la collecte, on a recueilli des renseignements généraux sur tous les membres du ménage (âge, sexe, liens de parenté entre les membres) du répondant longitudinal. Dans le cas des répondants longitudinaux âgés de 12 ans et plus, la déclaration par personne interposée n'a été permise que pour des raisons de maladie ou d'incapacité. Ces déclarations par personne interposée représentent 4,8 % des données recueillies pour les répondants de 12 et plus. Presque toutes les interviews auprès des répondants âgés de moins de 12 ans ont été effectuées par personne interposée.

6.4 Non-réponse et dépistage

De nombreuses stratégies ont été adoptées pour réduire le nombre de cas de non-réponse. Par exemple, afin d'éviter de surcharger les intervieweurs, on a déterminé à partir des résultats des cycles précédents la taille maximale d'une tâche à attribuer à chaque intervieweur. La formation offerte aux intervieweurs incluait divers moyens pour réduire le nombre de cas de non-contact (par exemple, faire des visites ou des appels à divers moments de la journée) en se servant des renseignements de base recueillis au cours de l'interview précédente.

On a demandé aux intervieweurs de déployer tous les efforts raisonnables pour que les répondants longitudinaux acceptent de répondre au questionnaire de l'ENSP. Dans les cas où il n'y avait personne au domicile, plusieurs tentatives ont été faites pour contacter le répondant. Aux personnes qui ont refusé de participer à l'ENSP, le bureau régional concerné a envoyé une lettre pour souligner l'importance de l'enquête et de la coopération du répondant. Puis, l'intervieweur a fait un deuxième appel (ou une deuxième visite). Si le moment de cet appel (ou de cette visite) était mal choisi, l'intervieweur convenait avec la

personne intéressée d'un moment plus opportun pour rappeler (ou pour revisiter).

Les cas de refus ont fait l'objet d'un suivi par les intervieweurs principaux, les surveillants de projet ou d'autres intervieweurs, afin de tenter de convaincre les répondants de participer à l'enquête. Afin de maximiser le taux de réponse, un grand nombre de cas de non-réponse ont fait l'objet d'un suivi lors des périodes de collecte subséquentes.

Le fait de ne pas dépister un répondant longitudinal est considéré comme un autre type de non-réponse. Les intervieweurs disposaient de plusieurs moyens pour dépister un répondant. Les données relatives au cas comprenaient la dernière adresse et le dernier numéro de téléphone ainsi que le nom et l'adresse d'un ou deux autres contacts. En outre, on a formé les intervieweurs pour qu'ils se servent des outils publics disponibles, comme les annuaires téléphoniques locaux et l'assistance annuaire. Les cas pour lesquels ces démarches n'ont donné aucun résultat ont été transmis à un intervieweur spécialement formé pour dépister les répondants. Les intervieweurs de dépistage avaient accès aux répertoires téléphoniques de l'ensemble du Canada et à des annuaires par ordre de numéros. Le taux de non-réponse cumulatif lié à l'impossibilité de dépister le répondant longitudinal se chiffre à 4,1 % du panel total, ce qui est relativement faible pour le cinquième cycle de l'enquête. La section 9.2.5 traite plus en détail des taux de non-réponse due au non-dépistage.

On a tenté d'interviewer tous les membres du panel longitudinal qui ont déménagé au Canada ou aux États-Unis. On a également tenté de confirmer le lieu de résidence des membres du panel longitudinal qui ont déménagé à l'extérieur du Canada et des États-Unis. Aucune interview n'a été menée lorsque le membre demeure à l'extérieur du Canada et des États-Unis et l'information a été mise à jour pour le prochain cycle.

7. Traitement des données

7.1 Vérification

Une vérification des données a été d'abord effectuée lors de l'interview assistée par ordinateur. L'intervieweur ne pouvait pas saisir des valeurs situées hors de la fourchette permise. Les erreurs d'enchaînement étaient contrôlées à l'aide du système programmé de sauts de questions (commandes « passer à ») de l'IAO. Par exemple, le système d'IAO veillait à ce que toutes les questions ne s'appliquant pas au répondant ne soient pas posées. En cas de réponses incohérentes, des messages d'avertissement étaient donnés. Dans certains cas, l'incohérence a dû être résolue avant de poursuivre l'interview et dans d'autres cas, l'incohérence a été acceptée et aucune mesure corrective n'a été prise. Après la collecte des données, un programme de vérification a été exécuté au bureau central. En général, les incohérences ont été corrigées en donnant la valeur « non déclaré » pour l'une des variables en cause ou les deux. Aucune imputation n'a été effectuée.

7.2 Antécédents résidentiels, un nouveau module au cycle 5

Les données recueillies dans le cadre du module des antécédents résidentiels ont été placées dans un fichier distinct. Il existe également un cliché d'article et un dictionnaire de données distincts pour le fichier des antécédents résidentiels. Lors de la collecte, les répondants avaient le choix de rapporter leur(s) lieu(x) de résidence soit en commençant en 1980 pour remonter le temps ou en reculant dans le temps pour finir en 1980. Lors du traitement, les données sur les lieux de résidence ont été converties en reculant dans le temps pour finir en 1980.

7.3 Codage

Pour plusieurs questions auxquelles il était possible de saisir une réponse écrite, on a codé les renseignements inscrits de façon à les classer dans de nouvelles catégories distinctes ou dans une catégorie énumérée existante si les renseignements inscrits étaient le calque d'une catégorie déjà énumérée. Dans la mesure du possible (par exemple pour la profession, l'industrie, les maladies), on a codé les renseignements conformément aux systèmes de classification type utilisés pour le Recensement de la population ou pour d'autres enquêtes de Statistique Canada, comme l'Enquête sur la santé et les limitations d'activités et le cycle 6 de l'Enquête sociale générale.

Dans le fichier maître du cycle 5 (contenant les 17 276 membres du panel longitudinal), les données concernant l'industrie et l'occupation pour tous les cycles sont codées en utilisant le « Système de classification des industries de l'Amérique du Nord (SCIAN) » et la « Classification type des professions de 1991 (SOC-91) ».

La classification des médicaments se fonde sur la Classification anatomique thérapeutique chimique (ATC) développée par l'Organisation mondiale de la santé (OMS) telle que

contenue dans la Base de données sur les produits pharmaceutique (BDPP) de Santé Canada en septembre 2003. Au cycle 5, une révision complète des codes de médicaments a été faite pour tous les cycles de l'ENSP. La liste complète des codes est disponible sur demande.

Les problèmes de santé causant une limitation des activités ont été codés conformément à la Classification internationale des maladies, 9^e édition (CIM-9) ou au Système de codage supplémentaire pour les incapacités ostéo-articulaires et musculaires établi au Canada aux fins de l'Enquête sur la santé et les limitations d'activités (ESLA).

7.4 Création de variables dérivées et groupées

Pour simplifier l'analyse des données, plusieurs variables ont été dérivées d'après certains éléments du questionnaire de l'ENSP. Le cinquième caractère du nom des variables dérivées est habituellement un D (voir section 11.3 sur la convention utilisée pour nommer les variables). Dans d'autres cas, plusieurs variables sont combinées pour en former une nouvelle. Pour plus de détails sur le calcul des variables dérivées, consulter le document « Enquête nationale sur la santé de la population, Documentation des variables dérivées, Cycles 1 à 5 ».

À partir de certaines variables on a créé des variables dites groupées, c'est à dire que les valeurs ont été regroupées afin de créer une autre variable. Dans certains cas, les variables dites groupées sont des variables dérivées simplement obtenues par la réduction du nombre de catégories de réponse. Le cinquième caractère du nom des variables groupées est habituellement un G (voir section 11.3 sur la convention utilisée pour nommer les variables).

7.5 Estimation et pondération

Le principe sur lequel repose l'estimation pour un échantillon probabiliste, comme celui de l'ENSP, veut que chacune des personnes faisant partie de l'échantillon « représente », à part elle-même, plusieurs autres personnes qui n'en font pas partie. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne échantillonnée en représente 50 dans la population. Selon cette terminologie, on peut dire que chaque personne a un facteur de pondération (ou poids) de 50.

La phase de pondération est l'étape où l'on calcule ce poids pour chaque personne. Il faut utiliser ce poids pour produire des estimations valables à partir des résultats de l'enquête. Par exemple, si l'on veut estimer le nombre de personnes au Canada dont l'état général de santé s'est détérioré entre deux cycles, on choisit les enregistrements se rapportant aux personnes qui possèdent cette caractéristique et on fait la somme des poids inscrits sur ces enregistrements.

La méthode de pondération utilisée par l'ENSP est décrite au chapitre 8.

7.6 Sous-ensembles de répondants

Afin d'offrir davantage de flexibilité aux utilisateurs, un seul fichier maître de microdonnées a été créé pour le cycle 5 de l'ENSP. Ce fichier contient les 17 276 membres du panel longitudinal de l'ENSP peu importe le type de réponse obtenu aux cycles 1 à 5. Dans ce fichier maître quatre sous-ensembles de répondants ont été créés avec les poids de sondage correspondants et les variables indicatrices rendant leur identification plus facile. Pour plus d'information relative au calcul des poids longitudinaux de chacun des sous-ensembles de répondants, veuillez vous référer au chapitre 8. Pour l'utilisation des poids longitudinaux veuillez consulter la section 11.1. Le tableau 7.A donne une description de ces quatre sous-ensembles de répondants en fonction du type de réponse.

Tableau 7.A : Sous-ensembles de répondants

Sous-ensembles de répondants	Type de réponse	Variables indicatrices	Nombre de répondants
Longitudinal carré	Panel entier : tous les membres du panel peu importe le type réponse aux cycles 1, 2, 3, 4 et 5.	Aucune, tous les enregistrements	17 276
Longitudinal complet	Tous les membres du panel qui ont fourni une réponse complète aux cycles 1, 2, 3, 4 et 5.	WF62LF	12 546
Longitudinal complet C1 et C5	Tous les membres du panel qui ont fourni une réponse complète aux cycles 1 et 5, peu importe le type de réponse aux cycles 2, 3 et 4.	WF62LFE	13 629
Longitudinal complet partagé	Tous les membres du panel qui ont fourni une réponse complète aux cycles 1, 2, 3, 4 et 5 et qui ont autorisé le partage des données au cycle 5.	WF62SLF	12 226

Les utilisateurs du fichier partagé, c'est-à-dire les ministères provinciaux de la santé et Santé Canada, doivent prendre note que le sous-ensemble «complet partagé» est fourni séparément sur un CD-ROM avec les poids d'échantillonnage correspondants. Les poids d'échantillonnage et les variables indicatrices des autres sous-ensembles ne sont pas disponibles sur le CD-ROM du fichier partagé.

7.7 Définition d'une réponse complète

Les trois derniers sous-ensembles de données ont été créés en utilisant la définition de « réponse complète ». Depuis le cycle 4, l'ENSP est strictement longitudinale. La définition d'une réponse n'est pas la même du point de vue longitudinal et transversal. Pour le panel longitudinal de l'ENSP une « réponse complète » comprend :

- **Les membres du panel** qui ont fourni une réponse complète à l'interview, c'est à dire qui ont répondu à toutes les questions jusqu'à un point déterminé dans le questionnaire de l'ENSP.
- **Les membres du panel décédés.** Le décès d'un membre du panel a été, lorsque possible, confirmé avec la Base canadienne de données de l'état civil - Décès. Lorsque le décès est confirmé, on saisit la cause reliée au décès et la date du décès (si non recueillie lors de l'enquête). Les variables des personnes décédées sont représentées par un « 9 », c'est-à-dire non déclaré dans le fichier.
- **Les membres du panel qui ont déménagé en établissement de soins de santé.** La collecte de données pour ces personnes se fait via le volet établissement de soins de santé. Les données recueillies sont ramenées dans le fichier du volet ménages de l'ENSP. Toutefois, le volet établissements de soins de santé recueille moins d'information que le volet ménages. Les variables non recueillies pour ces répondants sont codées avec un « 6 », c'est-à-dire «sans objet».

8. Pondération

Le présent chapitre décrit le processus de pondération longitudinale pour chacun des sous-ensembles de répondants décrits à la section 7.6. Le processus de pondération longitudinale est nécessairement différent de celui de la pondération transversale pour plusieurs raisons. Premièrement, les poids longitudinaux doivent représenter la probabilité de sélection de l'unité d'analyse au moment de la sélection de l'échantillon. Puisque l'échantillon longitudinal a été sélectionné en 1994-1995, les poids doivent représenter la probabilité de sélectionner l'individu au cycle 1, et non dans les cycles qui suivent. De plus, la définition d'une réponse longitudinale est différente d'une réponse transversale, ce qui nécessite des ajustements différents particuliers à chaque type de non-réponse. Les analystes doivent toujours utiliser les poids longitudinaux pour les études faites à partir des sous-ensembles de répondants. Les poids longitudinaux ont été conçus spécifiquement pour représenter la population cible de 1994-1995. Aux cycles 1, 2 et 3 des fichiers transversaux et longitudinaux ont été produits. Malgré le fait que les membres du panel faisaient partie à la fois des fichiers transversaux et longitudinaux, leurs poids n'étaient pas identiques dans ces deux types de fichier, mais plutôt ajustés afin de bien représenter la population d'intérêt.

Pour le cycle 5, quatre ensembles de poids ont été créés : WT64LS, WT62LF, WT62LFE et WT62SLF. Le tableau 8.A présente les variables de poids et les variables indicatrices associées avec chaque sous-ensemble. Un membre du panel fait partie d'un sous-ensemble donné lorsque la variable indicatrice de celui-ci est égale à 1.

Tableau 8.A : Sous-ensembles de répondants, variables de poids et variables indicatrices correspondantes

Sous-ensemble de répondants	Poids	Variables indicatrices
Longitudinal carré	WT64LS	Aucune, tous les enregistrements
Longitudinal complet	WT62LF	WF62LF
Longitudinal complet C1 et C5	WT62LFE	WF62LFE
Longitudinal complet partagé	WT62SLF	WF62SLF

Les quatre poids ont été post-stratifiés selon des groupes d'âge-sexe pour chaque province, conformément aux estimations de la population de 1994-1995 basées sur le Recensement de 1996. Les poids WT62LF, WT62LFE et WT62SLF ont été corrigés pour la non-réponse. La post-stratification est utilisée afin de s'assurer que les quatre sous-ensembles de répondants représentent correctement la population cible de 1994-1995. Les sections suivantes décrivent en détail la méthode de pondération longitudinale de l'ENSP.

8.1 Pondération longitudinale

La procédure utilisée pour la pondération longitudinale est basée essentiellement sur la pondération effectuée pour l'échantillon transversal du cycle 1 de l'ENSP. Quelques ajustements avaient été apportés à ce poids transversal afin d'incorporer les échantillons

supplémentaires utilisés uniquement à des fins transversales. Pour la pondération du panel longitudinal, ces ajustements ont été enlevés pour créer un poids «épuré». Ce poids épuré est le point de départ pour obtenir les poids longitudinaux.

8.1.1 Point de départ: poids épurés du cycle 1 (1994-1995)

Les poids épurés du cycle 1 sont obtenus à partir du poids de base de l'EPA pour toutes les provinces à l'exception du Québec, et à partir du poids de base de l'Enquête Sociale et de Santé pour le Québec. Plusieurs ajustements sont apportés à ces poids, afin de tenir compte des particularités de l'ENSP et de tenir compte de la probabilité réelle de sélection de chaque membre du panel. Tous ces ajustements effectués au cycle 1 sont conservés pour les cycles suivants puisque l'échantillon longitudinal réfère toujours à la même population de référence, c'est-à-dire celle de 1994-1995.

Une description complète des procédures de pondération utilisées au cycle 1 et toujours valides pour les cycles subséquents, est incluse aux sections 11.3 et 11.4 de la documentation des fichiers de microdonnées à grande diffusion des cycles 2 et 3.

À partir de ce point, des ajustements ont été appliqués au poids épuré afin d'obtenir les différents ensembles de poids longitudinaux.

8.1.2 Ajustements apportés pour créer les différents poids longitudinaux

8.1.2.1 Poids longitudinal carré (WT64LS)

Le poids longitudinal carré **WT64LS** doit être utilisé avec le sous-ensemble longitudinal carré. Il est calculé en post-stratifiant le poids épuré du cycle 1 conformément aux estimations de la population de 1994-1995 basées sur le Recensement de 1996, par groupe d'âge (0-11, 12-24, 25-44, 45-64, 65 et plus) et par sexe à l'intérieur de chaque province. L'ajustement du à la post-stratification est donné par :

$$\frac{\text{Estimation démographique dans une classe province-âge-sexe}}{\text{Somme des poids épurés des membres longitudinaux dans une classe province-âge-sexe}}$$

8.1.2.2 Poids longitudinal complet (WT62LF)

Le sous-ensemble longitudinal complet inclut seulement les membres du panel qui ont fourni une réponse complète à chaque cycle ou qui sont décédés. Les membres du panel exclus de ce sous-ensemble ont donc été non-répondants à un certain point durant les cinq premiers cycles de l'enquête, et leur poids doit donc être redistribué pour compenser pour la non-réponse.

Le point de départ est le poids épuré du cycle 1 et des ajustements pour la non-réponse sont faits. Un ajustement différent est effectué pour la non-réponse à chacun des cycles et ces ajustements sont cumulatifs d'un cycle à l'autre. Par exemple, pour obtenir le poids du cycle 5, les ajustements pour la non-réponse aux cycles 2, 3, 4 et 5 sont appliqués successivement aux poids épurés du cycle 1.

Les ajustements nécessaires en vue d'obtenir le poids longitudinal complet du cycle 5 sont décrits ci-dessous.

Ajustement 1 : Ajustement pour la non-réponse au cycle 2 (1996-1997)

L'ajustement pour la non-réponse a été faite en utilisant l'approche des classes de pondération. Les classes de pondération représentent des groupes de répondants qui affichent la même propension à répondre à l'enquête. On se fonde sur les caractéristiques des répondants et des non-répondants au cycle 2 de l'enquête obtenues d'après les données du cycle 1 pour déterminer à quelle classe de pondération ils appartiennent. On définit les classes à l'aide d'un algorithme de mise en grappe qui permet de disposer les unités d'échantillonnage en une structure arborescente en divisant successivement les ensembles de données en « branches » d'après les caractéristiques des unités. Chaque division a pour objet de répartir les unités existantes en au moins deux groupes les plus dissemblables possible en ce qui a trait aux taux observés de non-réponse (et à l'intérieur desquels on s'attend à observer des taux de non-réponse fort similaires). On peut se servir d'une autre caractéristique pour définir chaque division. Par exemple, on peut d'abord répartir les unités en un groupe de logements occupés par le propriétaire et un groupe de logements loués. Puis, on peut subdiviser le premier groupe en cinq sous-groupes, d'après le niveau de revenu du ménage, tandis qu'on peut subdiviser le deuxième groupe d'après l'âge du répondant. Chaque groupe nouvellement formé peut ensuite être divisé d'après d'autres caractéristiques et ainsi de suite. Les résultats des subdivisions finales correspondent aux classes de pondération.

L'algorithme CHAID (« chi-square automatic interaction detection ») a été utilisé pour définir les classes de pondération. Un minimum de 30 unités par classe de pondération a été utilisé afin de produire des ajustements plus stables.

Des classes de pondération ont été créées séparément pour chaque province. Notez que dans ce cas, le terme province réfère à la province de résidence utilisée au moment de la sélection de l'échantillon en 1994-1995. Les caractéristiques du ménage ainsi que les caractéristiques personnelles du membre longitudinal au cycle 1 ont été prises en considération. Quelques caractéristiques reliées au plan de sondage de l'enquête ou au poids

d'échantillonnage ont aussi été prises en considération dans le but d'incorporer le plan de sondage de l'enquête dans l'analyse. Les caractéristiques personnelles provenant de la composante Santé n'ont pas été utilisées car elles n'étaient pas disponibles pour plusieurs membres longitudinaux en 1994-1995.

Les variables choisies par l'algorithme CHAID afin de produire les classes de pondération pour corriger la non-réponse au cycle 2 sont énumérées dans le tableau 8.B. Deux variables du plan d'échantillonnage du cycle 1 indiquant la présence dans le ménage de membres de moins de 25 ans et la présence de membres de moins de 12 ans ont été incluses. Une variable indiquant la non-réponse aux questions sur le revenu au cycle 1 ainsi qu'une variable indiquant si l'individu était âgé de 16 ans ou moins ont aussi été incluses. Veuillez consulter le dictionnaire des données pour une description complète des variables du tableau 8.B.

Tableau 8.B : Variables pour l'ajustement de la non-réponse au cycle 2

DHC4_AGE	DHC4_MAR	GE34DURB	LFC4_1	SDC4DRAC
DHC4DECF	DHC4_OWN	HCC4DMDC	RAC4F1	SDC4GCB7
DHC4_DWE	GE34DCMA	INC4DIA5	SDC4DAIM	SEX

Afin de corriger pour les membres longitudinaux qui n'ont pas répondu au cycle 2, l'ajustement qui suit est apporté aux poids des répondants :

Somme des poids de tous les membres longitudinaux

Somme des poids des membres longitudinaux répondants aux cycles 1 et 2

Cet ajustement est apporté à l'intérieur de chacune des classes de pondération.

Ajustement 2 : Ajustement pour la non-réponse au cycle 3 (1998-1999)

Les 15 670 enregistrements avec un statut de réponse complet pour les deux premiers cycles servent de point de départ. Le profil de réponse longitudinal (la variable LONGPAT) a été créé de telle sorte qu'à chaque cycle, on rajoute simplement un chiffre à la fin de cette variable pour indiquer le statut de ce cycle. Pour chaque cycle, le code utilisé dans la construction de cette variable reflète le statut du membre du panel. Les codes sont : 1=complet, 2=décédé, 3=en établissement de soins de santé, 4=partiel et 5=non-réponse.

Une « réponse longitudinale complète » après trois cycles est représentée par un des profils de réponse suivants : 111, 112, 113, 122, 131, 132 ou 133.

Tous les autres profils de réponse forment la non-réponse (c.-à-d. 114, 115, 134 et 135). Les enregistrements pour lesquels le membre du panel était décédé au cycle 2 (c.-à-d. profil 122) ou en établissement de soins de santé depuis le cycle 2 (c.-à-d. profil 133) sont traités différemment des autres

enregistrements. Pour ces enregistrements, aucun ajustement pour la non-réponse n'est fait puisque leur poids au cycle 2 a déjà été ajusté pour tenir compte du fait que certains non-répondants au cycle 2 étaient probablement décédés ou en établissement de soins de santé.

L'approche des classes de pondération a été utilisée pour effectuer l'ajustement pour la non-réponse. Des classes distinctes ont été créées pour chaque province, c.-à-d. la province de résidence en 1994-1995. Pour l'ajustement de la non-réponse au cycle 3, les caractéristiques du ménage ainsi que les caractéristiques personnelles du membre du panel longitudinal au cycle 2 seulement ont été considérées. Comme pour le cycle 2, quelques caractéristiques reliées au plan de sondage de l'enquête ou au poids d'échantillonnage ont aussi été prises en considération dans le but d'incorporer le plan de sondage de l'enquête dans l'analyse. Cependant, contrairement à l'ajustement pour la non-réponse au cycle 2, les caractéristiques personnelles provenant de la composante Santé ont été utilisées puisqu'elles étaient disponibles pour tous les enregistrements impliqués dans l'ajustement pour la non-réponse au cycle 3.

Les variables choisies par l'algorithme CHAID afin de produire les classes de pondération pour corriger la non-réponse au cycle 3 sont énumérées dans le tableau 8.C. Une variable du plan au cycle 1 indiquant si le répondant provient de l'échantillon adulte ou enfant a été incluse. Une variable indiquant la non-réponse aux questions portant sur le revenu au cycle 2 a aussi été incluse. Veuillez consulter le dictionnaire des données pour une description complète des variables du tableau 8.C.

Tableau 8.C : Variables pour l'ajustement de la non-réponse au cycle 3

AD_6_1	DHC6_AGE	INS6_4	SDC6DAIM	SMS6_9A
AD_6_7	DHC6_MAR	INS6_6	SDC6_4P	SMS6_13A
ALC6WKY	DV_6_65J	LFC6_41	SDC6_5A	SMS6_13C
ALC6_3	EDC6_3	MHC6DWK	SDC6_5F	SMS6_13E
AM56_SHA	ES_6_80	MHC6_1A	SDC6_6B	SMS6_16D
AM66_PXY	GE36LMOV	MHC6_1B	SDC6_7A	SMS6_18A
AM66_SHA	HCC6F1	MHC6_1F	SDC6_7B	SMS6_18D
BPC6_10	HSC6DPAD	MHC6_1L	SDC6_7D	SP36_CPA
CCC6DNUM	HWS_5	MHC6_13	SEX	SSC6D2
CCC6_1L	INC6DIA5	PC_6_40	SHS6_4	SSC6_3
CCC6_1N	INC6_1A	RPC6_3	SMC6_2	SSS6_2
DGC6_1D	INC6_3B	RSS6_1	SMC6_5	SSS6_4

Afin de corriger pour les membres du panel qui n'ont pas répondu au cycle 3, l'ajustement suivant est apporté aux poids des membres répondants :

Somme des poids des membres longitudinaux répondants aux cycles 1 et 2
Somme des poids des membres longitudinaux répondants aux cycles 1, 2 et 3

Cet ajustement est apporté à l'intérieur de chacune des classes de pondération. Seuls les enregistrements avec les profils de réponse longitudinale suivants ont été impliqués dans l'ajustement : 111 à 115, 131, 132, 134 et 135. Encore une fois, les enregistrements pour lesquels le membre du panel était décédé au cycle 2 ou en établissement de soins de santé depuis le cycle 2 ne font pas partie de cet ajustement.

Ajustement 3 : Ajustement pour la non-réponse au cycle 4 (2000-2001)

Les 14 619 enregistrements avec une réponse complète après trois cycles servent de point de départ. Encore une fois, les enregistrements pour lesquels le membre du panel était décédé au cycle 2 ou 3 ou en établissement de soins de santé depuis le cycle 2 ou 3 sont traités différemment des autres enregistrements. Pour ces enregistrements, aucun ajustement pour la non-réponse n'est fait puisque leur poids au cycle 2 ou 3 a déjà été ajusté pour tenir compte du fait que certains non-répondants au cycle 2 ou 3 étaient probablement décédés ou en établissement de soins de santé.

Comme pour les cycles précédents, l'approche des classes de pondération a été utilisée pour effectuer l'ajustement pour la non-réponse. Des classes distinctes ont été créées pour chaque province, c.-à-d. la province de résidence en 1994-1995. Pour l'ajustement de la non-réponse au cycle 4, les caractéristiques du ménage ainsi que les caractéristiques personnelles du membre du panel longitudinal au cycle 3 seulement ont été considérées. Comme pour le cycle 3, quelques caractéristiques reliées au plan de sondage de l'enquête ou au poids d'échantillonnage ont aussi été prises en considération dans le but d'incorporer le plan de sondage de l'enquête dans l'analyse. Les caractéristiques personnelles provenant de la composante Santé ont été utilisées puisqu'elles étaient disponibles pour tous les enregistrements impliqués dans l'ajustement pour la non-réponse au cycle 4.

Les variables choisies par l'algorithme CHAID afin de produire les classes de pondération pour corriger la non-réponse au cycle 4 sont énumérées dans le tableau 8.D. Une variable indiquant la non-réponse aux questions portant sur le revenu au cycle 3 a aussi été incluse. Veuillez consulter le dictionnaire des données pour une description complète des variables du tableau 8.D.

Tableau 8.D : Variables pour l'ajustement de la non-réponse au cycle 4

CCC8DANY	DGC8_1A	HCC8_1	PAC8_1A	SDC8_6A
CCC8_1C	DHC8_AGE	INC8DIA5	PAC8_1J	SDC8_7A
CCC8_1L	DHC8DECF	ISC8_1	PY_8DH1	SEX
CCC8_1N	DHC8_OWN	NU_8_1B	RAC8F1	SSC8DEMO
CCC8_1V	FIC8F1	PAC8DFD	RPC8_2	SSC8DSOC
DGC8F1	GE38DURB	PAC8DLEI	SDC8_4A	TWC8_5

Afin de corriger pour les membres du panel qui n'ont pas répondu au cycle 4, l'ajustement suivant est apporté aux poids des membres répondants :

Somme des poids des membres longitudinaux répondants aux cycles 1, 2 et 3
Somme des poids des membres longitudinaux répondants aux cycles 1 à 4

Cet ajustement est apporté à l'intérieur de chacune des classes de pondération. Les enregistrements pour lesquels le membre du panel était décédé au cycle 2 ou 3 ou en établissement de soins de santé depuis le cycle 2 ou 3 ne font pas partie de cet ajustement.

Ajustement 4 : Ajustement pour la non-réponse au cycle 5 (2002-2003)

Les 13 582 enregistrements avec une réponse complète après quatre cycles servent de point de départ. Encore une fois, les enregistrements pour lesquels le membre du panel était décédé au cycle 2, 3 ou 4 ou en établissement de soins de santé depuis le cycle 2, 3 ou 4 sont traités différemment des autres enregistrements. Pour ces enregistrements, aucun ajustement pour la non-réponse n'est fait puisque leur poids au cycle 2, 3 ou 4 a déjà été ajusté pour tenir compte du fait que certains non-répondants au cycle 2, 3 ou 4 étaient probablement décédés ou en établissement de soins de santé.

Comme pour les cycles précédents, l'approche des classes de pondération a été utilisée pour effectuer l'ajustement pour la non-réponse. Des classes distinctes ont été créées pour chaque province, c.-à-d. la province de résidence en 1994-1995. Pour l'ajustement de la non-réponse au cycle 5, les caractéristiques du ménage ainsi que les caractéristiques personnelles du membre du panel longitudinal au cycle 4 seulement ont été considérées. Comme pour le cycle 4, quelques caractéristiques reliées au plan de sondage de l'enquête ou au poids d'échantillonnage ont aussi été prises en considération dans le but d'incorporer le plan de sondage de l'enquête dans l'analyse. Les caractéristiques personnelles provenant de la composante Santé ont été utilisées puisqu'elles étaient disponibles pour tous les enregistrements impliqués dans l'ajustement pour la non-réponse au cycle 5.

Les variables choisies par l’algorithme CHAID afin de produire les classes de pondération pour corriger la non-réponse au cycle 5 sont énumérées dans le tableau 8.E. Trois variables de plan du cycle 1 ont été utilisées, indiquant la présence de membres du ménage qui ont moins de 12 ans, moins de 25 ans, et si le répondant provient de l’échantillon adulte ou enfant. Une variable indiquant la non-réponse aux questions portant sur le revenu au cycle 4 a aussi été incluse. Veuillez consulter le dictionnaire des données pour une description complète des variables du tableau 8.E.

Tableau 8.E : Variables pour l'ajustement de la non-réponse au cycle 5

ALC0_3	DHC0_OWN	IMM	MHC0_1J	SMC0_2
ALC0DTYP	DHC0DL12	INC0DIA5	MHC0DCH	ST_0DC4
ALC0DWKY	DHC0DLE5	ISC0_1	MHC0DDS	ST_0DC5
AM60_SHA	GE30DURB	LSC0_1	PAC0DFD	ST_0DC6
BPC0_10	GHC0_21	LSC0DPFT	PAC0DLEI	ST_0DC8
CCC0DANY	HCC0DHPC	MHC0_16	SDC0_4A	ST_0DR2
DGC0F1	HSC0DHSI	MHC0_1A	SDC0_6A	ST_0DW3
DHC0_AGE	HWC0DSW	MHC0_1F	SEX	ST_0DW6

Afin de corriger pour les membres du panel qui n’ont pas répondu au cycle 5, l’ajustement suivant est apporté aux poids des membres répondants :

$$\frac{\text{Somme des poids des membres longitudinaux répondants aux cycles 1, 2, 3 et 4}}{\text{Somme des poids des membres longitudinaux répondants aux cycles 1 à 5}}$$

Cet ajustement est apporté selon les classes de pondération. Les enregistrements pour lesquels le membre du panel était décédé au cycle 2,3 ou 4 ou en établissement de soins de santé depuis le cycle 2, 3 ou 4 ne font pas partie de cet ajustement.

Ajustement 5 : Post-stratification

Le poids des unités faisant partie du sous-ensemble a finalement été post-stratifié conformément aux estimations de la population de 1994-1995 basées sur le Recensement de 1996, par groupe d’âge (0-11, 12-24, 25-44, 45-64, 65 et plus) et par sexe à l’intérieur de chaque province. Ceci est fait pour s’assurer que la population de 1994-1995 est représentée correctement dans toute estimation provenant du fichier longitudinal. Cet ajustement de post-stratification est donné par :

$$\frac{\text{Estimation démographique dans une classe province-âge-sexe}}{\text{Somme des poids des membres longitudinaux répondants aux cycles 1 à 5 dans une classe province-âge-sexe}}$$

Le poids final pour une réponse longitudinale complète, **WT62LF**, est calculé en prenant d'abord le poids épuré du cycle 1, et en le multipliant ensuite par les ajustements 1 à 5.

8.1.2.3 Poids longitudinal complet cycles 1 et 5 (WT62LFE)

L'ENSP inclut maintenant, à partir du cycle 4, un poids additionnel à être utilisé avec le sous-ensemble de membres du panel qui ont fourni une réponse complète au cycle 1 et au cycle le plus récent (dans ce cas, le cycle 5). Ce sous-ensemble inclut les répondants faisant partie du sous-ensemble complet, auquel on rajoute les répondants ayant une réponse complète seulement aux cycles 1 et 5 (c'est-à-dire les non-répondants et les réponses partielles au cycle 2, 3 ou 4). Puisqu'un bon nombre de répondants faisant partie de ce sous-ensemble font aussi partie du sous-ensemble complet, la méthode de pondération utilisée pour ce dernier est à nouveau utilisée, avec quelques modifications.

Le point de départ est le poids longitudinal complet du cycle 5 juste avant l'étape de post-stratification. Tel que décrit dans les sections précédentes, ce poids a été corrigé pour la non-réponse à chacun des cycles. Cependant, certains non-répondants au cycle 2, 3 ou 4, dont les poids avaient été répartis aux répondants du sous-ensemble complet, ont répondu de nouveau au cycle 5 et leurs poids doivent maintenant être récupérés. Le poids de départ qui est attribué à chaque enregistrement récupéré au cycle 5 est le poids que cet enregistrement avait au dernier cycle où il faisait partie du sous-ensemble complet. Ceci permet d'utiliser le poids le plus récent, qui tient compte des ajustements pour la non-réponse aux cycles précédents. L'équivalent des poids ainsi attribués doit être enlevé des poids du sous-ensemble complet du cycle 5. Cette diminution se fait séparément pour chacune des classes de pondération déterminées au cycle où les enregistrements récupérés étaient non-répondants pour la première fois. Il s'agit en fait de redonner le poids aux enregistrements de la même façon qu'ils l'avaient perdu. Ce poids doit cependant être ajusté afin que les enregistrements récupérés au cycle 5 représentent la même proportion dans le sous-ensemble complet C1 et C5, qu'ils représenteraient s'ils avaient fait partie du sous-ensemble complet à tous les cycles, c'est-à-dire s'ils avaient été impliqués dans tous les ajustements pour la non-réponse aux cycles subséquents. À nouveau cet ajustement se fait séparément à l'intérieur de chaque classe de pondération.

Finalement, comme pour les autres poids longitudinaux, le poids des unités faisant partie du sous-ensemble complet C1 et C5 a été corrigé en le post-stratifiant conformément aux estimations de la population de 1994-1995 basées sur le recensement de 1996, par groupe d'âge (0-11, 12-24, 25-44, 45-64, 65 et plus) et par sexe à l'intérieur de chaque province. Cet ajustement est donné par :

Estimation démographique dans une classe province-âge-sexe
Somme des poids des membres du sous-ensemble longitudinal complet
C1 et C5, dans une classe province-âge-sexe

Le poids final pour une réponse complète aux cycles 1 et 5 est le poids WT62LFE. Pour une description plus détaillée et technique des ajustements utilisés dans la création de ce poids, veuillez vous référer à Brisebois et Mathieu (2003)².

8.1.2.4 Poids longitudinal complet partagé (WT62SLF)

Dans certains cas, les répondants ont indiqué qu'ils n'acceptaient pas de partager l'information recueillie lors de toutes les interviews menées dans le cadre de l'ENSP avec les ministères provinciaux de la santé et Santé Canada. Étant donné que ces partenaires reçoivent seulement les enregistrements de ceux qui ont accepté de partager leur information, un poids spécial doit être créé de telle sorte que les estimations calculées à partir de ce sous-ensemble représentent correctement la population totale.

Un seul ajustement simple est apporté au poids longitudinal complet pour créer le poids du sous-ensemble complet partagé (fichier partagé). Cet ajustement est donné par :

Somme des poids des membres longitudinaux répondants aux cycles 1 à 5,
dans une classe province / profil de réponse longitudinale / âge-sexe
Somme des poids des membres longitudinaux répondants aux cycles 1 à 5
et ayant accepté de partager, dans une classe province / profil de réponse
longitudinale / âge-sexe

Il est à noter qu'aux cycles 3, 4 et 5, quelques profils de réponse longitudinale ont été regroupés afin de produire des ajustements plus stables. Le regroupement a été fait pour quelques classes province/âge-sexe qui n'avaient que quelques observations dans certains profils de réponse représentant les répondants décédés ou en établissement de soins de santé.

² Brisebois, F. et Mathieu, P. (2003) Creation of a new longitudinal weight for the Canadian National Population Health Survey: Providing data users with greater analytical flexibility. Proceedings of the Survey Methods Section, Statistical Society of Canada

Dans chaque cas, le profil de réponse problématique a été regroupé avec un autre profil de la même classe province/âge-sexe de telle sorte que la somme des poids corresponde toujours aux totaux de la population. Le regroupement des profils de réponse a préservé la distribution pondérée des membres du panel demeurant dans un ménage, dans un établissement de soins de santé ou décédés pour chacun des cycles. Le poids final pour le sous-ensemble complet partagé, (WT62SLF) est obtenu en multipliant le poids longitudinal complet (WT62LF) par cet ajustement. Puisque cet ajustement est fait en respectant les classes utilisées lors de l'ajustement pour la post-stratification, aucune post-stratification supplémentaire n'est nécessaire.

9. Qualité des données

L'évaluation de la qualité des données est un aspect important à toute enquête. L'étude de la qualité des données permet de vérifier la fiabilité et l'exactitude de l'information recueillie, ainsi que de s'en servir pour améliorer la qualité de la prochaine réalisation de l'enquête.

L'enquête permet de produire des estimations fondées sur l'information recueillie à partir d'un échantillon de personnes. On aurait pu obtenir des estimations quelque peu différentes si on avait effectué un recensement complet en utilisant le même questionnaire, les mêmes intervieweurs, les mêmes superviseurs, les mêmes méthodes de traitement, etc. que ceux utilisés pour l'enquête. La différence entre les estimations tirées 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 interprètent incorrectement les questions posées, que les réponses soient mal saisies ou que des erreurs se produisent au moment du traitement et de la totalisation des données. Tous ces exemples représentent des erreurs non dues à l'échantillonnage.

9.1 Erreurs dues à l'échantillonnage

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

Par exemple, supposons qu'une personne estime que 25 % des canadiens âgés de 12 ans et plus ont amélioré leur état de santé général entre les cycles 1 et 2 de l'enquête 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 vérifier la qualité des estimations statistiques produites, et recommande vivement aux utilisateurs produisant des estimations à partir des fichiers de données de l'ENSP d'en faire de même. Pour consulter les

lignes directrices sur la façon d'interpréter les résultats du CV, se référer au tableau à la fin de la section 10.4.

9.1.1 Méthode bootstrap pour l'estimation de la variance

Afin de déterminer la qualité de l'estimation et de calculer le CV, il est nécessaire de calculer l'écart-type. Les intervalles de confiance de même qu'un grand nombre de tests statistiques nécessitent également le calcul de l'écart-type de l'estimation.

Pour l'ENSP, on utilise un plan de sondage à plusieurs degrés, ce qui signifie qu'il n'y a pas de formule simple pour calculer les estimations de la variance. Par conséquent, on a recours à une méthode approximative. On applique la méthode bootstrap parce que l'information sur le plan d'échantillonnage doit être prise en compte lors du calcul des estimations de la variance. La méthode bootstrap permet d'effectuer ce calcul et, conjointement avec le programme Bootvar, elle demeure une méthode relativement facile à utiliser.

La méthode de ré-échantillonnage bootstrap utilisée pour l'ENSP sous-tend la sélection d'échantillons aléatoires simples, appelés les échantillons répétés, et le calcul de l'écart des estimations d'un échantillon répété à l'autre. Dans chaque strate, on sélectionne un échantillon aléatoire simple de $(n-1)$ des n grappes avec remise pour former un échantillon répété. Il est à noter que puisque la sélection se fait avec remplacement, une grappe peut être choisie plus d'une fois. Pour chaque échantillon répété, on recalcule le poids de sondage de chaque enregistrement dans les grappes $(n-1)$ sélectionnées. Ces poids sont par la suite post-stratifiés en fonction de l'information démographique de la même façon que les pondérations du plan de sondage, ce qui permet d'obtenir les poids bootstrap finaux.

Le processus complet (sélectionner les échantillons aléatoires simples, recalculer et post-stratifier les poids de chaque strate) est répété B fois, B prenant une grande valeur. En général, on utilise $B=500$ pour l'ENSP afin de produire 500 poids bootstrap. Pour obtenir l'estimateur bootstrap de la variance, on doit calculer l'estimation ponctuelle de chacun des B échantillons. L'écart-type de ces estimations représente l'estimateur bootstrap de la variance. Statistique Canada a élaboré un programme qui peut effectuer tous ces calculs pour l'utilisateur : le programme Bootvar. Pour plus d'information concernant les poids bootstrap, veuillez vous référer au chapitre 11.2.

Le programme Bootvar est offert en formats SAS et SPSS. Il est constitué de macros qui calculent les variances de totaux, ratios, différences entre ratios, et pour des régressions linéaires et logistiques.

Le programme Bootvar est fourni avec les poids bootstrap et un document qui explique comment modifier et utiliser le programme selon les besoins de l'utilisateur.

9.2 Erreurs non dues à l'échantillonnage

Beaucoup de temps et d'efforts sont investis afin de 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 de l'ENSP, il n'y a presque pas de non-réponse partielle car une fois le questionnaire débuté les répondants ont tendance à le terminer. Dans la plupart des cas, la non-réponse partielle à l'enquête était attribuable au fait que le répondant ne comprenait pas une question ou l'interprétait mal, refusait d'y répondre, ne se souvenait pas de l'information demandée ou n'était pas en mesure de fournir de l'information sur lui-même ou sur la personne pour laquelle il fournissait une réponse par procuration. Il y a eu non-réponse totale lorsqu'il était impossible de dépister ou de joindre le répondant, qu'aucun membre du ménage ne pouvait fournir les renseignements voulus ou que le répondant refusait de participer à l'enquête. On a traité les cas de non-réponse 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 8.1.2 pour avoir de plus amples détails sur l'ajustement de la pondération pour la non-réponse.

Cette section présente de l'information concernant différentes composantes de la non-réponse. On aborde d'abord la non-réponse de façon globale en présentant les taux de non-réponse obtenus à chaque cycle. Suivent ensuite les sections relatives aux refus, la conversion des refus, l'érosion, puis le non-dépistage des répondants longitudinaux. Finalement, la non-réponse d'item est examinée brièvement.

9.2.1 Taux de réponse

Ce chapitre présente les taux de réponse et décrit la façon dont ils sont calculés. Le calcul des taux de réponse du cycle 1 diffère de celui des cycles subséquents. Les taux de réponse du cycle 1 sont basés sur les 20 095 personnes faisant partie du champ d'enquête et choisies en vue de former le panel, alors que les cycles subséquents sont basés sur les 17 276 personnes formant le panel. Autre différence importante, pour les trois premiers cycles, les taux de réponse des personnes choisies sont calculés pour la composante générale (H05) et la composante Santé (H06). Comme l'enquête est devenue purement longitudinale au cycle 4 et qu'il n'y a plus de distinction entre ces deux composantes, le taux de réponse du panel longitudinal n'est calculé qu'une seule fois aux cycles 4 et 5.

9.2.1.1 Taux de réponse du cycle 1 (1994-1995)

Les taux de réponse du cycle 1 sont basés sur les 20 095 personnes faisant partie du champ de l'enquête et choisies en vue de former le panel. Par conséquent, les personnes qui faisaient partie d'un ménage hors du champ de l'enquête (code de statut = 017, 018, 023, 024)³ ont été exclues des calculs des taux de réponse du cycle 1.

Taux de réponse H05 des personnes choisies

nbre de personnes choisies ayant répondu à la composante H05
toutes les personnes choisies à l'intérieur du champ de l'enquête

Pour la composante H05 de l'ENSP, le taux de réponse des personnes choisies à l'échelle nationale était de **86,0 %**. À l'échelle provinciale, ce taux variait de 80,7 % en Ontario à 91,0 % en Alberta.

Taux de réponse H06 des personnes choisies

nbre de personnes choisies ayant répondu à la composante H06
toutes les personnes choisies à l'intérieur du champ de l'enquête

Pour la composante H06, le taux de réponse des personnes choisies était de **83,6 %** à l'échelle du Canada, mais à l'échelle provinciale il variait de 77,8 % en Ontario à 89,1 % en Alberta.

Information pertinente pour le calcul des taux de réponse :

Nombre de personnes choisies à l'intérieur du champ de l'enquête :	20 095
Nombre de répondants à la composante H05	
parmi les personnes choisies :	17 276
Nombre de répondants à la composante H06	
parmi les personnes choisies :	16 794
Nombre de non-répondants à la composante H05	
parmi les personnes choisies :	2 819
Nombre de non-répondants à la composante H06	
parmi les personnes choisies :	3 301

³ 017 = Autre logement non admissible (p. ex. ambassade)
018 = Ménage rejeté
023 = En construction ou démolé
024 = Logement vacant

Calcul des taux de réponse des personnes choisies :

$$\begin{aligned} \text{Taux H05 parmi les personnes choisies} &= \\ \frac{17\,276}{17\,276 + 2\,819} &= \frac{17\,276}{20\,095} = 86,0\% \end{aligned}$$

$$\begin{aligned} \text{Taux H06 parmi les personnes choisies} &= \\ \frac{16\,794}{16\,794 + 3\,301} &= \frac{16\,794}{20\,095} = 83,6\% \end{aligned}$$

9.2.1.2 Taux de réponse du cycle 2 (1996-1997)

Les taux de réponse du cycle 2 sont fondés sur les 17 276 personnes qui forment le panel longitudinal. Les personnes décédées de même que celles qui ont déménagé et qui ont été interviewées dans un établissement de soins de santé sont considérées comme une réponse aux fins longitudinales. Aucun membre du panel n'est classé comme étant hors du champ d'enquête.

Taux de réponse du panel à la composante H05

$$\frac{\text{nbre de membres du panel ayant répondu à la composante H05 ou qui sont décédés ou qui ont déménagé en établissement de soins de santé}}{\text{nbre de membres du panel}}$$

À l'échelle canadienne, le taux de réponse du panel était de **93,6 %**. À l'échelle provinciale, ce taux variait de 90,4 % en Colombie-Britannique à 96,2 % à Terre-Neuve.

Taux de réponse du panel à la composante H06

$$\frac{\text{nbre de membres du panel ayant répondu à la composante H06 ou qui sont décédés ou qui ont déménagé en établissement de soins de santé}}{\text{nbre de membres du panel}}$$

À l'échelle canadienne, le taux de réponse du panel était de **92,8 %**. À l'échelle provinciale, ce taux variait de 89,6 % en Colombie-Britannique à 95,1 % à Terre-Neuve.

Information pertinente pour le calcul des taux de réponse :

Nombre de membres du panel longitudinal :	17 276
Nombre de membres du panel décédés ⁴ :	287
Nombre de membres du panel qui ont déménagé en établissement de soins de santé :	62
Nombre de membres du panel répondants au H05 :	15 819
Nombre de membres du panel répondants au H06 :	15 687
Nombre de membres du panel non-répondants au H05 :	1 108
Nombre de membres du panel non-répondants au H06 :	1 240

Calcul des taux de réponse du panel :

$$\text{Taux de réponse du panel à la composante H05} = \frac{15\,819 + 287 + 62}{15\,819 + 287 + 62 + 1\,108} = \frac{16\,168}{17\,276} = 93,6 \%$$

$$\text{Taux de réponse du panel à la composante H06} = \frac{15\,687 + 287 + 62}{15\,687 + 287 + 62 + 1\,240} = \frac{16\,036}{17\,276} = 92,8 \%$$

9.2.1.3 Taux de réponse du cycle 3 (1998-1999)

Tout comme au cycle 2, les taux de réponse longitudinaux du cycle 3 sont fondés sur le panel longitudinal complet. Les personnes décédées de même que celles qui ont déménagé et ont été interviewées dans un établissement de soins de santé sont considérées comme une réponse aux fins longitudinales. Aucun membre du panel n'est classé comme étant hors du champ d'enquête.

Taux de réponse du panel à la composante H05

$$\frac{\text{nbre de membres du panel ayant répondu à la composante H05}}{\text{nbre de membres du panel}} \text{ ou qui sont décédés ou qui ont déménagé en établissement de soins de santé}$$

À l'échelle canadienne, le taux de réponse du panel était de **88,9 %**. À l'échelle provinciale, ce taux variait de 84,2 % en Colombie-Britannique à 92,5 % à Terre-Neuve.

⁴ À ce moment, les décès n'avaient pas été confirmés à l'aide de la Base canadienne de données sur l'état civil – décès. Il a été découvert que quatre de ces personnes étaient vivantes au cycle 3.

Taux de réponse du panel à la composante H06

nbre de membres du panel ayant répondu à la composante H06
ou qui sont décédés ou qui ont déménagé en établissement de soins de santé
nbre de membres du panel

À l'échelle canadienne, le taux de réponse du panel était de **88,2 %**. À l'échelle provinciale, ce taux variait de 83,9 % en Colombie-Britannique à 92,0 % à Terre-Neuve.

Information pertinente pour le calcul des taux de réponse :

Nombre de membres du panel longitudinal :	17 276
Nombre de membres du panel décédés ⁵ :	599
Nombre de membres du panel qui ont déménagé en établissement de soins de santé :	114
Nombre de membres du panel répondants au H05 :	14 647
Nombre de membres du panel répondants au H06 :	14 532
Nombre de membres du panel non-répondants au H05 :	1 916
Nombre de membres du panel non-répondants au H06 :	2 031

Calcul des taux de réponse du panel :

$$\text{Taux de réponse du panel à la composante H05} = \frac{14\,647 + 599 + 114}{14\,647 + 599 + 114 + 1\,916} = \frac{15\,360}{17\,276} = 88,9 \%$$

$$\text{Taux de réponse du panel à la composante H06} = \frac{14\,532 + 599 + 114}{14\,532 + 599 + 114 + 2\,031} = \frac{15\,245}{17\,276} = 88,2 \%$$

9.2.1.4 Taux de réponse du cycle 4 (2000-2001)

Comme dans les cycles 2 et 3, les taux de réponse longitudinaux du cycle 4 sont obtenus à l'aide des 17 276 membres du panel longitudinal. Les personnes décédées de même que celles qui ont déménagé et ont été interviewées dans un établissement de soins de santé sont considérées comme une réponse aux fins longitudinales. Aucun membre du panel n'est classé comme étant hors du champ d'enquête. Comme l'enquête est maintenant purement longitudinale et ne fait plus de distinction entre les questionnaires H05 et H06, un seul taux de réponse est donc calculé.

⁵ À ce moment, les décès n'avaient pas été confirmés à l'aide de la Base canadienne de données sur l'état civil - décès. Il a été découvert qu'une de ces personnes était vivante au cycle 4.

Taux de réponse

$$\frac{\text{nbre de membres du panel répondants ou qui sont décédés
ou qui ont déménagé en établissement de soins de santé}}{\text{nbre de membres du panel}}$$

À l'échelle du Canada, le taux de réponse du panel était de **84,8 %**. À l'échelle provinciale, ce taux variait de 80,5 % en Colombie-Britannique à 90,9 % en Saskatchewan.

Information pertinente pour le calcul des taux de réponse :

Nombre de membres du panel longitudinal :	17 276
Nombre de membres du panel décédés ⁶ :	957
Nombre de membres du panel qui ont déménagé en établissement de soins de santé :	135
Nombre de membres du panel répondants :	13 559
Nombre de membres du panel non-répondants :	2 625

Calcul du taux de réponse du panel :

$$\text{Taux de réponse du panel} = \frac{13\,559 + 957 + 135}{13\,559 + 957 + 135 + 2\,625} = \frac{14\,651}{17\,276} = 84,8 \%$$

9.2.1.5 Taux de réponse du cycle 5 (2002-2003)

Comme dans les cycles 2, 3 et 4, les taux de réponse longitudinaux du cycle 5 sont obtenus à l'aide des 17 276 membres du panel longitudinal. Les personnes décédées de même que celles qui ont déménagé et qui ont été interviewées dans un établissement de soins de santé sont considérées comme une réponse aux fins longitudinales. Aucun membre du panel n'est classé comme étant hors du champ d'enquête. Comme l'enquête est maintenant purement longitudinale et ne fait plus de distinction entre les questionnaires H05 et H06, un seul taux de réponse est donc calculé.

Taux de réponse

$$\frac{\text{nbre de membres du panel répondants ou qui sont décédés
ou qui ont déménagé en établissement de soins de santé}}{\text{nbre de membres du panel}}$$

⁶ À ce moment, les décès n'avaient pas été confirmés à l'aide de la Base canadienne de données sur l'état civil - décès. Il a été découvert qu'une de ces personnes était vivante au cycle 5.

À l'échelle du Canada, le taux de réponse du panel était de **80,6 %**. À l'échelle provinciale, ce taux variait de 77,7 % en (Colombie-Britannique) à 84,7 % en (Saskatchewan).

Information pertinente pour le calcul des taux de réponse :

Nombre de membres du panel longitudinal :	17 276
Nombre de membres du panel décédés :	1 279
Nombre de membres du panel qui ont déménagé en établissement de soins de santé :	161
Nombre de membres du panel répondants :	12 484
Nombre de membres du panel non-répondants :	3 352

Calcul du taux de réponse du panel :

$$\text{Taux de réponse du panel} = \frac{12\,484 + 1\,279 + 161}{12\,484 + 1\,279 + 161 + 3\,352} = \frac{13\,924}{17\,276} = 80,6 \%$$

9.2.2 Taux de refus

Les refus sont la plus grande source de non-réponse pour l'ENSP. Les refus représentent 49 % de la non-réponse du cycle 2, 56 % de la non-réponse du cycle 3 et 61 % de la non-réponse des cycles 4 et 5. Même si l'intention est de suivre tous les 17 276 membres du panel au fil du temps, il y a des enregistrements qui ne sont pas envoyés à la collecte à chaque cycle, tels que les refus fermes. Notez que lorsque les décès de membres du panel ont été confirmés par appariement avec les fichiers des mortalités, ces cas sont considérés comme des réponses complètes pour le reste de l'enquête et ne sont plus envoyés à la collecte.

Deux taux de refus peuvent être calculés pour chaque cycle, un basé sur les enregistrements qui sont envoyés à la collecte et l'autre basé sur tous les 17 276 enregistrements. Le tableau 9.A, qui présente les deux taux pour les cycles 2, 3, 4 et 5, démontre que les taux de refus augmentent à chaque cycle.

Tableau 9.A : Taux de refus par cycle

	Nombre de cas envoyés à la collecte	Nombre de refus lors de la collecte	Taux de refus basé sur les enregistrements envoyés à la collecte	Nombre de refus qui n'ont pas été envoyés à la collecte	Nombre total de refus	Taux de refus basé sur le panel complet 17,276
Cycle 2	17 266	539	3,1 %	1	540	3,1 %
Cycle 3	16 582	601	3,6 %	469	1070	6,2 %
Cycle 4	16 186	1017	6,3 %	526	1543	8,9 %
Cycle 5	15 616	1265	8,1 %	687	1952	11,3 %

9.2.3 Taux de conversion des refus

Le tableau 9.B montre le taux de conversion des refus entre chaque cycle. Le taux de conversion des refus est le pourcentage de répondants qui ont refusé dans un cycle particulier et qui donnent une réponse complète dans un cycle suivant. Ces taux sont basés seulement sur les enregistrements qui ont été envoyés à la collecte car seuls les enregistrements qui ont été envoyés pouvaient être convertis. Ainsi, le nombre de refus du tableau 9.B est différent du nombre de refus dans le tableau 9.A. Par exemple, le taux de conversion entre les cycles 2 et 3 est basé sur seulement 73 enregistrements car 467 des 540 refus du cycle 2 n'ont pas été envoyés à la collecte au cycle 3.

Tableau 9.B: Taux de conversion des refus

	Nombre de refus	Convertis au cycle suivant	Convertis 2 cycles plus tard	Convertis 3 cycles plus tard	Jamais convertis
Cycle 2	73	22 (30 %)	16 (22 %)	2 (3 %)	33 (45 %)
Cycle 3	539	164 (30 %)	34 (6 %)		341 (63 %)
Refus aux cycles 2 et 3	26	8 (31 %)	0 (0 %)		18 (69 %)
Cycle 4	849	164 (19 %)			685 (81 %)
Refus aux cycles 3 et 4	259	26 (10 %)			233 (90 %)
Refus aux cycles 2, 3 et 4	15	0 (0 %)			15 (100 %)

Comme on peut observer dans le tableau 9.B, les taux de conversion des refus sont beaucoup plus élevés à court terme qu'à long terme.

9.2.4 Taux d'érosion

L'érosion dans le cadre d'une enquête longitudinale est une réduction de la taille d'un échantillon attribuable aux non-répondants, aux passages hors du champ d'enquête et aux personnes qu'on ne peut dépister. L'érosion est dans cette section définie relativement au fait qu'un membre du panel fasse ou non partie du sous-ensemble complet. Ainsi, aussitôt qu'une non-réponse est observée pour un membre du panel, celui-ci fait partie de l'érosion. Deux taux d'érosion différents sont calculés : un taux qui présente l'érosion entre deux cycles consécutifs et un taux qui présente l'érosion cumulative en fonction de l'échantillon initial. Ces deux taux sont calculés en utilisant le nombre de répondants dans le sous-ensemble complet. Pour de plus amples renseignements sur le sous-ensemble complet, voir la section 7.7.

Information pertinente pour le calcul des taux d'érosion :

Nombre de membres du panel de l'échantillon longitudinal :	17 276
Nombre de personnes dans le sous-ensemble complet du cycle 2 :	15 666
Nombre de personnes dans le sous-ensemble complet du cycle 3 :	14 618
Nombre de personnes dans le sous-ensemble complet du cycle 4 :	13 582
Nombre de personnes dans le sous-ensemble complet du cycle 5 :	12 546

Taux d'érosion d'un cycle à l'autre:

Cycle 2 (1996-1997) :	$\frac{17\,276-15\,666}{17\,276}$	=	$\frac{1\,610}{17\,276}$	=	9,3 %
Cycle 3 (1998-1999) :	$\frac{15\,666-14\,618}{15\,666}$	=	$\frac{1\,048}{15\,666}$	=	6,7 %
Cycle 4 (2000-2001) :	$\frac{14\,618-13\,582}{14\,618}$	=	$\frac{1\,036}{14\,618}$	=	7,1 %
Cycle 5 (2002-2003) :	$\frac{13\,582-12\,546}{13\,582}$	=	$\frac{1\,036}{13\,582}$	=	7,6 %

Taux d'érosion cumulatifs

Cycle 2 (1996-1997) :	$\frac{17\ 276-15\ 666}{17\ 276}$	=	$\frac{1\ 610}{17\ 276}$	=	9,3 %
Cycle3 (1998-1999) :	$\frac{17\ 276-14\ 618}{17\ 276}$	=	$\frac{2\ 658}{17\ 276}$	=	15,4 %
Cycle 4 (2000-2001) :	$\frac{17\ 276-13\ 582}{17\ 276}$	=	$\frac{3\ 694}{17\ 276}$	=	21,4 %
Cycle 5 (2002-2003) :	$\frac{17\ 276-12\ 546}{17\ 276}$	=	$\frac{4\ 730}{17\ 276}$	=	27,4 %

Tel qu'observé typiquement dans les enquêtes longitudinales, le taux d'érosion entre les cycles 1 et 2 est considérablement plus élevé (9,3 %) que ceux observés par la suite. Ces derniers sont demeurés constants au cours des cycles. Cumulativement, on observe qu'après cinq cycles, plus d'un quart du panel fait partie de l'érosion.

Tableau 9.C: Source de l'érosion par cycle - Sous-ensemble longitudinal complet

Source de l'érosion	Cycle 2		Cycle 3		Cycle 4		Cycle 4		Total	
	Nombre	%								
Données partielles au cycle 1	482	29,9	s/o	s/o	s/o	s/o	s/o	s/o	482	10,2
Données partielles	110	6,8	93	8,8	61	5,9	86	8,3	350	7,4
Déménagés	75	4,7	55	5,2	53	5,1	20	1,9	203	4,3
Refus	492	30,6	474	45,2	541	52,2	523	50,5	2030	42,9
Non-dépistés	275	17,1	131	12,5	175	16,9	162	15,6	743	15,7
Autre non-réponse	176	10,9	295	28,1	206	19,9	245	23,6	922	19,5
Total	1610		1048		1036		1036		4730	100

9.2.5 La non-réponse attribuable aux personnes non-dépistées

L'incapacité de dépister un membre du panel de l'échantillon longitudinal représente un type de non-réponse. Malgré les nombreux efforts des intervieweurs (décrits à la section 6.4), le taux cumulatif attribuable aux personnes non-dépistées augmente cycle après cycle. Cependant, de nombreuses tentatives ont été faites pour maintenir ce taux le plus faible possible. Le tableau 9.D présente le profil des membres du panel selon qu'ils ont pu ou non être dépistés au cours des cycles 2, 3, 4 et 5.

Tableau 9.D : Dépistage, cycles 2, 3, 4 et 5

Cycle 2	Cycle 3	Cycle 4	Cycle 5	Nombre de membres du panel
D	D	D	D	16,230
D	D	D	ND	325
D	D	ND	D	120
D	D	ND	ND	136
D	ND	D	D	56
D	ND	D	ND	17
D	ND	ND	D	24
D	ND	ND	ND	72
ND	D	D	D	66
ND	D	D	ND	18
ND	D	ND	D	6
ND	D	ND	ND	13
ND	ND	D	D	43
ND	ND	D	ND	17
ND	ND	ND	D	24
ND	ND	ND	ND	109
Total				17 276

D: dépisté
 ND: non-dépisté

On observe que plus un membre du panel reste non-dépisté, moins il est probable qu'il soit dépisté au prochain cycle.

À l'aide des données de l'enquête, on a également observé que les non-répondants d'un cycle ont une probabilité environ quatre fois supérieure à celle des répondants d'être non-dépistés. Environ 3,5 % des non-répondants du cycle 2 n'ont pas été dépistés au cycle 3, 4,6 % des non-répondants du cycle 3 n'ont pas été dépistés au cycle 4 et 7,2 % des non-répondants du cycle 4 n'ont pas été dépistés au cycle 5.

Le tableau 9.D montre toutefois que les efforts de dépistage ont beaucoup de valeur car même des membres du panel qui étaient non-dépistés pendant trois cycles consécutifs ont été dépistés au cycle 5.

9.2.6 Taux de refus et de «Ne sait pas» par item

Les taux de refus par variable varient de 0 % à 5 %. Ceux des modules varient de 0 % à 0,9 %. D'un cycle à l'autre, on observe que ce sont les mêmes variables et modules qui présentent un taux de refus supérieurs aux autres. Par exemple, les variables du revenu ont un taux de refus de près de 4% et certains modules de la population active ainsi que le sous module de contrôle (module stress) ont le plus

haut taux de refus soit 0,11 % et 0,12 % respectivement.

Les taux de «Ne sait pas» par variable varient de 0 % à 72,7 %. Ceux des modules varient de 0,01 % à 2,74 %. D'un cycle à l'autre, on observe que ce sont les mêmes variables et modules qui présentent un taux de «Ne sait pas» supérieurs aux autres. Par exemple, les variables du module assurance ont un taux de «Ne sait pas» de près de 2,0 % et celles du module du revenu ont un taux de 1,8 %.

Des taux de refus et de «Ne sait pas» ont aussi été calculés au niveau du répondant pour déterminer le pourcentage de questions auxquelles un individu refuse de répondre ou répond «Ne sait pas». Le tableau 9.E montre la proportion des membres du panel qui n'ont refusé à aucune question, qui ont refusé à moins de 1 % des questions posées, et qui ont refusé à moins de 3 % des questions posées.

Tableau 9.E: Taux de refus et de «Ne sait pas» au niveau de la personne

	Refus à 0 % des questions	Refus à moins de 1 % des questions	Refus à moins de 3 % des questions	«Ne sait pas» à 0 % des questions	«Ne sait pas» à moins de 1% des questions	«Ne sait pas» à moins de 5% des questions
Total	92,3 %	98,1 %	98,8 %	62,6 %	93,7 %	99,4 %
Hommes	92,2 %	97,7 %	98,6 %	63,7 %	93,9 %	99,3 %
Femmes	92,3 %	98,5 %	98,9 %	61,6 %	93,6 %	99,5 %
< 12 ans	94,5 %	95,7 %	98,0 %	82,9 %	95,5 %	99,5 %
12-24 ans	93,2 %	98,5 %	98,6 %	46,5 %	88,6 %	98,3 %
25-44 ans	94,4 %	98,9 %	99,2 %	72,4 %	97,6 %	99,8 %
45-64 ans	92,4 %	98,5 %	99,0 %	65,9 %	96,3 %	99,7 %
> = 65 ans	86,8 %	96,6 %	97,9 %	48,0 %	86,6 %	99,0 %
Par procuration	91,6 %	95,9 %	97,9 %	69,5 %	88,3 %	98,6 %
Pas par procuration	92,3 %	98,3 %	98,8 %	61,9 %	94,2 %	99,5 %

On verra aussi que les taux de «Ne sait pas» ont beaucoup plus de variation que les taux de refus quand ils sont séparés par sexes, groupes d'âges et types d'interview. De tous les répondants qui ont complété l'interview en entier au cycle 5, 92,3 % ont accepté de répondre à toutes les questions. Presque 99 % ont des taux de refus inférieurs à 3 %. Pour les taux de «Ne sait pas», 62,6 % ont des taux de 0 %. Plus de 99 % ont des taux de «Ne sait pas» inférieurs à 5 % et presque 94 % ont des taux inférieurs à 1 %. Ceci démontre que presque tout le monde qui refuse de répondre ou répond « Ne sait pas » le fait pour très peu de questions.

10. Lignes directrices sur les totalisations, l'analyse et la diffusion

Cette section de la documentation décrit les lignes directrices que les utilisateurs devraient suivre pour créer des tableaux, analyser, publier ou diffuser de quelque façon que ce soit les données de l'ENSP. Ces lignes directrices devraient permettre aux utilisateurs de produire des chiffres qui se rapprochent de ceux produits par Statistique Canada, tout en étant capables d'obtenir, conformément à ces lignes directrices, des chiffres qui n'ont pas été publiés.

10.1 Lignes directrices sur l'arrondissement

Pour que les estimations de l'ENSP qui seront publiées correspondent à celles de Statistique Canada, nous conseillons aux utilisateurs de se conformer aux lignes directrices suivantes pour l'arrondissement de ces estimations. Des estimations qui ne sont pas arrondies laissent supposer qu'elles sont beaucoup plus précises qu'elles le sont en réalité.

- a) Les estimations dans le corps d'un tableau statistique devraient être arrondies à la centaine près au moyen de la technique d'arrondissement classique. Selon cette technique, si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente de 1 la valeur du dernier chiffre à conserver. Par exemple, lorsqu'on veut arrondir à la centaine près de la façon classique, si les deux derniers chiffres se situent entre 00 et 49, il faut les remplacer par 00 et laisser le chiffre précédent (le chiffre des centaines) tel quel. Si les deux derniers chiffres se situent entre 50 et 99, on les remplace par 00 et on augmente de 1 le chiffre précédent.
- b) Les totaux partiels marginaux et les totaux marginaux des tableaux statistiques devraient être calculés à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la centaine près à l'aide de la technique d'arrondissement classique.
- c) Les moyennes, les proportions, les taux et les pourcentages devraient être calculés à partir d'éléments non arrondis (c'est-à-dire les numérateurs ou les dénominateurs), puis arrondis à une décimale au moyen de la technique d'arrondissement classique. Lorsqu'on veut arrondir à un seul chiffre par cette technique, si le dernier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente de 1 le dernier chiffre à conserver.
- d) Les sommes et les différences d'agrégats (ou de rapports) devraient être calculées à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la centaine près (ou à la décimale près) en employant la technique d'arrondissement classique.
- e) Si, en raison de limitations d'ordre technique ou de toute autre nature, on utilise une autre technique que l'arrondissement classique, de sorte que les estimations à publier ou à diffuser sous une forme quelconque diffèrent des estimations correspondantes publiées par Statistique Canada, nous conseillons aux utilisateurs d'indiquer la raison de ces divergences dans le ou les documents à publier ou à diffuser.

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

Le plan d'échantillonnage utilisé pour l'ENSP n'était pas auto-pondéré, c'est-à-dire que les personnes faisant partie l'échantillon n'ont pas toutes le même poids d'échantillonnage. Pour produire des estimations simples, y compris des tableaux statistiques, les utilisateurs doivent employer le poids d'échantillonnage approprié. Sinon, les estimations calculées à partir des différents sous-ensembles ne pourront être considérées comme représentatives de la population cible de 1994-1995 et ne correspondront pas à celles de Statistique Canada.

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

10.2.1 Définitions des genres d'estimations : de type nominal par opposition à quantitatif

Avant de discuter de la façon dont on peut totaliser et analyser les données de l'ENSP, il est utile de décrire les deux principaux genres d'estimations ponctuelles des caractéristiques de la population qui peuvent être produites.

Estimations de type nominal :

Les estimations de type nominal sont des estimations du nombre ou du pourcentage de personnes dans la population visée par l'enquête qui possèdent certaines caractéristiques ou qui font partie d'une classe définie. Le nombre de personnes ayant cessé de fumer entre les cycles est un exemple de ce type d'estimation. Une estimation du nombre de personnes possédant une caractéristique donnée peut également être appelée une estimation d'agrégat.

Exemple de question de type nominal :

Actuellement, est-ce que ... fume(z) des cigarettes tous les jours, à l'occasion ou jamais? (SMC2_2)

- Tous les jours
- À l'occasion
- Jamais

Estimations quantitatives :

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

Un exemple d'estimation quantitative est l'augmentation moyenne du nombre de cigarettes fumées par jour par des fumeurs quotidiens qui ont augmenté leur

consommation entre deux cycles.

Exemple de question quantitative :

*Actuellement, combien de cigarettes est-ce que ... fume(z) chaque jour?
(SMC2_4)*

||| Nombre de cigarettes

10.2.2 Totalisation d'estimations de type nominal

On peut obtenir des estimations du nombre de personnes qui possèdent une caractéristique donnée à partir du fichier de microdonnées en additionnant les poids finaux de tous les enregistrements contenant la ou les caractéristiques d'intérêt. On obtient les proportions et les rapports de la forme \hat{X} / \hat{Y} comme suit :

- a) en additionnant les poids finaux des enregistrements contenant la caractéristique voulue pour le numérateur (\hat{X}),
- b) en additionnant les poids finaux des enregistrements contenant la caractéristique voulue pour le dénominateur (\hat{Y}), puis
- c) en divisant l'estimation du numérateur par celle du dénominateur.

10.2.3 Totalisation d'estimations quantitatives

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

- a) en multipliant la valeur de la variable étudiée par le poids final de chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements étudiés afin d'obtenir le numérateur (\hat{X}),
- b) en additionnant les poids finaux des enregistrements présentant la caractéristique étudiée pour obtenir le dénominateur (\hat{Y}), puis
- c) en divisant l'estimation du numérateur par l'estimation du dénominateur.

Par exemple, pour estimer le nombre moyen de cigarettes fumées chaque jour par les personnes qui fument tous les jours, on calcule d'abord le numérateur (\hat{X}) en faisant la somme des produits de la variable **SMC2_4** par le poids final pour tous les enregistrements pour lesquels la valeur de la variable **SMC2_2** est « tous les jours ». On obtient le dénominateur (\hat{Y}) en additionnant le poids final de tous les enregistrements pour lesquels la valeur de la variable **SMC2_2** est « tous les

jours ». Le nombre moyen de cigarettes fumées chaque jour par les personnes qui fument tous les jours est finalement obtenu en divisant (\hat{X}) par (\hat{Y}) .

10.3 Lignes directrices pour l'analyse statistique

L'ENSP est basée sur un plan d'échantillonnage complexe qui comporte une stratification et de multiples degrés de sélection ainsi que des probabilités inégales de sélection des répondants. L'utilisation de données provenant d'enquêtes aussi complexes pose des difficultés aux analystes car le plan de sondage de l'enquête et les probabilités de sélection influent sur les méthodes d'estimation et de calcul de la variance à utiliser.

Bien que de nombreuses méthodes d'analyse intégrées aux progiciels statistiques permettent d'utiliser des poids, le sens ou la définition du poids diffère de celui qui convient à une enquête par sondage. Par conséquent, si les estimations faites au moyen de ces progiciels sont exactes dans bien des cas, les variances calculées n'ont pratiquement aucune signification.

Dans le cas de nombreuses techniques d'analyse (par exemple la régression linéaire, la régression logistique et l'analyse de variance), il existe un moyen de rendre l'application des progiciels courants plus raisonnable. Si l'on transforme les poids des enregistrements contenus dans le fichier de manière à ce que le poids moyen soit un (1), les résultats produits par les progiciels courants seront plus raisonnables et tiendront compte des probabilités inégales de sélection, bien qu'ils continueront à ne faire aucun cas de la stratification et de la répartition en grappes du plan d'échantillonnage. On peut effectuer cette transformation en employant dans l'analyse un poids égal au poids final divisé par la moyenne des poids finaux des unités d'échantillonnage (personnes) qui contribuent à l'estimation en question.

Auparavant, on produisait des tableaux de CV pour les données transversales. Des tableaux de CV n'ont pas été créés pour les fichiers longitudinaux car il existe un très grand nombre de combinaisons de variables possibles pour l'analyse. Pour estimer correctement la variance, les responsables de l'ENSP recommandent d'utiliser la méthode bootstrap. Grâce à cette méthode, la complexité de la pondération et le plan d'échantillonnage de l'enquête sont intégrés dans le calcul de la variance. Un programme d'estimation de la variance bootstrap de format SAS a été élaboré, offrant une documentation d'accompagnement et des exemples d'utilisation, pour faciliter le calcul de la variance à l'aide de la méthode bootstrap. Le programme permet également de calculer le coefficient de variation correspondant. Une version semblable du programme est aussi offerte en SPSS. Il importe que les utilisateurs apprennent à l'utiliser car le programme générera les estimations de variances exactes pour évaluer la qualité des estimations mises en tableaux. On recommande fortement d'utiliser ce programme plutôt que l'approche des poids transformés. Certains progiciels statistiques, comme STATA, peuvent lire l'information des strates et des grappes servant à l'estimation de la variance, ce qui améliore la qualité de l'estimation mais ne tient pas compte des différents ajustements appliqués aux poids.

10.4 Lignes directrices pour la diffusion

Avant de diffuser ou de publier des estimations de totaux et de proportions tirées du fichier maître, les utilisateurs doivent 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 10, l'estimation pondérée ne doit pas être diffusée, quelque que soit la valeur de son coefficient de variation. Cela tient au fait que la possibilité d'obtenir une variance artificiellement faible est plus grande si l'échantillon compte moins de 10 personnes. Pour les estimations pondérées basées sur des échantillons composés de 10 répondants ou plus, les utilisateurs doivent calculer le coefficient de variation de l'estimation et suivre les lignes décrites dans le tableau 10.A.

Tableau 10.A : Lignes directrices relatives à la variabilité d'échantillonnage

Type d'estimation	CV (en %)	Lignes directrices
Acceptable	0,0 – 16,5	On peut envisager une diffusion générale non restreinte des estimations. Aucune annotation particulière n'est nécessaire.
Marginale	16,6 – 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	Plus de 33,3	Statistique Canada recommande de ne pas publier des estimations dont la qualité est inacceptable. Toutefois, si l'utilisateur choisit de le faire, il doit alors adjoindre la lettre F (ou un autre identificateur semblable) et les diffuser avec l'avertissement suivant : «Nous avisons l'utilisateur que ...(précisez les données)... ne répondent pas aux normes de qualité de Statistique Canada pour ce programme statistique. Les conclusions tirées de ces données ne sauraient être fiables et seront fort probablement erronées. Ces données et toute conclusion qu'on pourrait en tirer ne doivent pas être publiées. Si l'utilisateur choisit de les publier, il est alors tenu de publier également le présent avertissement».

11. Utilisation du fichier longitudinal

11.1 Utilisation des poids longitudinaux

Le fichier maître du cycle 5 contient les 17 276 membres du panel et quatre poids (section 8.1) correspondant à quatre sous-ensembles de répondants (section 7.6). Cet ensemble de données inclut les répondants qui sont devenus non-répondants. Des variables indicatrices ont été créées afin d'identifier les enregistrements appartenant à chacun des sous-ensembles de répondants (tableau 8.A). Les enregistrements qui ne font pas partie d'un sous-ensemble de répondants particulier ont une variable indicatrice et un poids égal à 0 pour ce sous-ensemble de répondants particulier. Pour constituer le sous-ensemble de répondants d'intérêt, sélectionnez les enregistrements où la variable indicatrice appropriée est égale à 1.

Le poids WT64LS est appelé « poids carré » et s'applique à l'ensemble des 17 276 membres qui constituent l'échantillon longitudinal. Il est à noter que tous les cas de non-réponse devront être pris en considération lors des calculs.

Le poids WT62LF est appelé « poids longitudinal complet » et s'applique aux 12 546 enregistrements qui forment le sous-ensemble « complet ».

Le poids WT62LFE est appelé « poids longitudinal complet C1 et C5 » et s'applique aux 13 629 enregistrements qui forment le sous-ensemble « complet C1 et C5 ».

Le poids WT62SLF est appelé « poids longitudinal complet partagé » et s'applique aux 12 226 enregistrements du sous-ensemble « complet partagé ».

11.2 Assurance de la fiabilité des estimations avec l'utilisation des poids bootstrap

Les poids bootstrap sont nécessaires au calcul de l'estimation de la variance. La section 9.1.1 contient de l'information sur la méthode bootstrap à appliquer pour le calcul de l'estimation de la variance. À chacun des sous-ensembles de répondants correspond un ensemble de poids bootstrap. Pour les données du cycle 5, quatre ensembles différents de poids bootstrap ont été créés, soit le carré, le complet, le complet C1 et C5 et le complet partagé. Le tableau 11.A présente les sous-ensembles de répondants et le nom des fichiers de poids bootstrap correspondants.

Tableau 11.A : Sous-ensembles de répondants et fichiers de poids Bootstrap correspondants

Sous-ensemble de répondants	Nom du fichier contenant les poids Bootstrap
Longitudinal carré	B5long
Longitudinal complet	B5lngf
Longitudinal complet C1 et C5	B5lngfe
Longitudinal complet partagé	B5lngf (partagé)

En raison de la complexité du plan d'échantillonnage, les utilisateurs devraient utiliser le programme Bootvar fourni pour calculer la variance. Les données de sortie standards pour l'écart-type provenant d'autres logiciels statistiques, comme SAS et SPSS, peuvent nettement sous-estimer la variance d'une estimation dans le cadre de cette enquête. **Il incombe à l'utilisateur d'assurer la qualité/fiabilité des estimations qu'il produit en respectant les lignes directrices énoncées au chapitre 10 et de calculer correctement la variance pour toutes les estimations.** Le non-respect de cette consigne pourrait conduire à une mauvaise interprétation des résultats et compromettre la qualité des travaux de recherche.

Certains logiciels statistiques peuvent intégrer l'information sur les strates et les grappes comme données d'entrée dans le traitement analytique, ce qui génère une estimation de variance qui se rapproche beaucoup plus de l'estimation de la variance réelle. Cependant, ces logiciels ne tiennent pas compte des divers ajustements de pondération qui peuvent, dans certains cas, influencer considérablement sur les estimations de la variance.

11.3 Convention utilisée pour nommer les variables

On a adopté, pour nommer les variables de l'ENSP, une convention qui permet aux utilisateurs des données de faire facilement référence à des données similaires provenant de différentes périodes de collecte et de divers volets de l'ENSP. Les exigences qui suivent devaient être satisfaites : d'une part limiter les noms des variables à au plus huit caractères pour qu'il soit facile de les utiliser avec les logiciels d'analyse et, d'autre part, préciser la période de référence de l'enquête (1994-1995, 1996-1997, 1998-1999, 2000-2001, 2002-2003) dans le nom pour permettre de repérer facilement les variables conceptuellement identiques d'une période de référence à l'autre. Par exemple, des données conceptuellement identiques sur l'usage du tabac ont été recueillies en 1994-1995, en 1996-1997, en 1998-1999, en 2000-2001 et en 2002-2003. Les noms des variables pour ces questions ne devaient différer qu'à la position réservée à l'année qui identifie le cycle durant lequel les données ont été recueillies. Cette convention est appliquée pendant toute la durée de l'enquête longitudinale et est adoptée pour tous les volets de l'ENSP, soit le volet ménages, le volet établissements de soins de santé et, auparavant, le volet des résidents du Nord et les suppléments.

11.3.1 Structure élémentaire des noms des variables

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

Positions 1 et 2 : Nom de la variable / section du questionnaire

Position 3 : Type d'enquête

Position 4 : Année / cycle durant lequel apparaît la variable

Position 5 : Type de variable

Positions 6 à 8 : Numéro / nom de la variable tiré du questionnaire

Par exemple : la signification des noms de variable DHC4_AGE, DHC6_AGE, DHC8_AGE, DHC0_AGE et DHC2_AGE est la suivante :

- DH** : dans la section du questionnaire sur la démographie et la composition du ménage ;
- C** : questions qui font partie du contenu principal de l'enquête auprès des ménages ;
- 4/6/8/0/2** : figure dans le cycle 1 (1994-1995), figure dans le cycle 2 (1996-1997), figure dans le cycle 3 (1998-1999), figure dans le cycle 4 (2000-2001), figure dans le cycle 5 (2002-2003);
- _** : peut être trouvé sur le questionnaire ;
- AGE** : nom de la variable.

11.3.2 Positions 1 et 2 : Nom de la variable / section du questionnaire

AD	Dépendance à l'égard de l'alcool	MH	Santé mentale
AL	Alcool	NU	Nutrition
AM	Administration de l'enquête	PA	Activités physiques
CC	Problèmes de santé chroniques	PH	Santé préventive
DG	Consommation de médicaments	PY	Ressources psychologiques (estime de soi, contrôle, sentiment de cohésion)
DH	Démographie et ménage	RA	Limitation des activités
ED	Niveau de scolarité	RH/MB	Antécédents résidentiels
FV	Consommation de fruits et légumes	RP	Mouvements répétitifs
GE	Identificateurs géographiques	SD	Renseignements socio-démographiques
GH	État de santé général	SL	Sommeil
HC	Utilisation des soins de santé	SM	Usage du tabac
HS	État de santé	SP	Identificateurs d'échantillon
HW	Taille et poids, Image corporelle	SS	Soutien social
IJ	Blessures	ST	Stress
IN	Revenu	TU	Bronzage et exposition aux rayons UV
IS	Assurance	WF	Variable indicatrice des sous-ensembles
LF/LS	Population active	WT	Poids d'échantillonnage des sous-ensembles

Quelques variables d'identification importantes ne suivent pas la convention : p. ex. : REALUKEY, PERSONID, CYCLE, SUBCYCLE, DESIGPRV, STRATUM et REPLICAT.

Il existe aussi quelques variables qui sont « constantes ». Le tableau 11.B présente les variables qui n'apparaissent qu'une seule fois sur le fichier de données. Le nom de ces variables ne suit pas la convention.

Tableau 11.B : Variables longitudinales « constantes »

Variables longitudinales	Concept
SEX	Sexe
DOB	Jour de naissance
MOB	Mois de naissance
YOB	Année de naissance
COB	Pays de naissance
COBC	Code pour le pays de naissance
COBGC	Pays de naissance (7 groupes) - Groupé
IMM	Indicateur d'immigration pour le répondant
AOI	Âge au moment de l'immigration - Dérivé
HWB	Poids à la naissance
DOD	Jour de décès
MOD	Mois de décès
YOD	Année de décès
COD	Code pour la cause de décès

11.3.3 Position 3 : Type d'enquête

- A Enquête supplémentaire sur l'asthme
- B Contenu de l'enquête auprès des échantillons supplémentaires des provinces - questions posées aux enfants
- C Questions du contenu principal répétées à chaque cycle
- F Supplément sur la précarité alimentaire
- I Établissements de soins de santé
- K Questions posées aux enfants membres du panel
- N Nord (Yukon / Territoires du Nord-Ouest)
- P Contenu de l'enquête auprès des échantillons supplémentaires des provinces - questions posées aux adultes
- S Suppléments nationaux (Enquête promotion santé)
- Questions particulières à un cycle qui ne sont pas répétées chaque fois (par ex. stress aux cycles 1, 4 et 5, accès aux services de santé au cycle 2)
- 3 Variables d'administration de l'enquête pour le volet des ménages et la composante démographique (H03)
- 5 Variables d'administration de l'enquête pour la composante Générale (H05)
- 6 Variables d'administration de l'enquête pour la composante Santé (H06) (par exemple les variables des poids, de l'accord de partage ou des dates d'interview).

11.3.4 Position 4 : Variable de l'année de référence du cycle

4	Cycle 1 (1994-1995)
6	Cycle 2 (1996-1997)
8	Cycle 3 (1998-1999)
0	Cycle 4 (2000-2001)
2	Cycle 5 (2002-2003)
A	Cycle 6 (2004-2005)
B	Cycle 7 (2006-2007)
C	Cycle 8 (2008-2009)
D	Cycle 9 (2010-2011)
E	Cycle 10 (2012-2013)

11.3.5 Position 5 : Type de variable

_	Variable recueillie	Variable qui figure directement sur le questionnaire
C	Variable codée	Variable codée à partir d'une ou de plusieurs variables recueillies (p. ex. Système de classification des industries de l'Amérique du Nord (SCIAN))
D	Variable dérivée	Variable calculée d'après une ou plusieurs variables recueillies ou codées, ordinairement pendant le traitement au Bureau central (p. ex. Système de classification type de l'état de la santé (CHSMS-HUI3))
F	Variable indicatrice	Variable calculée par l'application informatique à partir d'une ou de plusieurs variables recueillies (comme une variable dérivée) aux fins de son utilisation ultérieure durant l'interview (p. ex. indicateur de travail). Cela peut indiquer aussi qu'une réponse élaborée a été recueillie pour cette variable (p. ex. indicateur de limitation d'activité)
G	Variable groupée	Variables recueillies, codées, supprimées ou dérivées agrégées en groupes (p. ex. groupes d'âge)
L	Variable longitudinale dérivée	Variable calculée d'après des variables provenant d'au moins deux cycles d'enquête

11.3.6 Position 6 à 8 : Nom de la variable

En général, les trois dernières positions correspondent au nom qui figure sur le questionnaire. On utilise des nombres dans la mesure du possible : ainsi Q1 devient 1. Dans le cas des questions de type « Choisissez toutes les réponses appropriées », on se sert des lettres correspondant à chaque catégorie possible de réponse : ainsi Q1 (Choisissez toutes les réponses appropriées) devient 1A, 1B, 1C, etc. On représente les variables démographiques utilisées fréquemment par les

analystes par trois lettres indicatrices plutôt que par un numéro de question. Par exemple, « âge » est représenté par DHC4_AGE au cycle 1 (1994-1995), DHC6_AGE au cycle 2 (1996-1997), etc. Dans les cas où les réponses à des groupes de questions portant sur le même sujet ont été recueillies dans des sections du questionnaire portant des noms différents, on réserve la position 6 pour identifier la sous-section. Par exemple, la première question sur le stress chronique a été nommée ST_2_C1, la première question sur le stress au travail a été nommée ST_2_W1. Un autre exemple survient dans le cas des questions sur l'état de santé général de l'Enquête sur la promotion de la santé. Ces questions ont été réparties en trois sections aux fins de leur intégration au questionnaire et les noms des variables correspondantes reflètent cette situation. Le caractère de la position 6 indique dans quelle section la question figure.

12. Accès aux données de l'ENSP

12.1 Fichiers de microdonnées

Le souci de protéger la confidentialité des données longitudinales écarte la possibilité d'émettre un fichier de microdonnées à grande diffusion (FMGD). Cependant, l'accès aux fichiers maîtres de microdonnées longitudinales, incluant les données des cycles 4 et 5 (et l'accès aux fichiers maîtres de microdonnées transversales des trois premiers cycles de l'ENSP) est possible via le programme de Centres de données de recherche (CDR) de Statistique Canada. Le programme des CDR s'inscrit dans une initiative de Statistique Canada, du Conseil de recherches en sciences humaines du Canada (CRSH) et de consortiums d'universités visant à renforcer la capacité de recherche sociale du Canada et à soutenir le milieu de la recherche sur les politiques.

Les centres de données de recherche (CDR) permettent aux chercheurs d'accéder aux microdonnées d'enquêtes sur les ménages et sur la population. Les centres comptent des employés de Statistique Canada. Ils sont exploités en vertu des dispositions de la Loi sur la statistique et sont administrés conformément à toutes les règles de confidentialité. Ils ne sont accessibles qu'aux chercheurs dont les propositions ont été approuvées et qui ont prêté serment en qualité de personnes réputées être employées de Statistique Canada. On trouve des CDR à l'étendue du pays. Ainsi, les chercheurs n'ont pas à se déplacer vers Ottawa pour avoir accès aux microdonnées de Statistique Canada. Vous pouvez trouver de l'information supplémentaire sur le site du Programme des centres de données de recherche à http://www.statcan.ca/francais/rdc/index_f.htm.

Les chercheurs qui n'ont pas accès à un CDR ont une autre option qui est le service de téléaccès de la Division de la statistique de la santé. Ce service permet aux chercheurs de développer et de tester des programmes informatiques sur leur ordinateur en utilisant les fichiers fictifs qui sont des répliques des fichiers maîtres. Les chercheurs envoient ensuite leurs programmes à une adresse électronique. Les programmes sont tournés en utilisant les fichiers maîtres de microdonnées dans un environnement de serveur interne. Par la suite, les fichiers de sortie sont vérifiés pour la confidentialité et retournés aux chercheurs par courriel. Pour plus d'information relative au service, veuillez contacter l'Unité d'accès aux données à nphs-ensp@statcan.ca.

Des FMGD sont disponibles pour les trois premiers cycles de l'ENSP et procurent un accès élargi aux fichiers transversaux de l'Enquête. Les FMGD de l'ENSP sont aussi accessibles via l'Initiative de démocratisation des données (IDD) des universités et collègues participants. Pour de l'information à ce sujet, veuillez consulter le site de Statistique Canada sous <http://www.statcan.ca/francais/edu/index.htm>. Les FMGD de l'ENSP des cycles 1,2 et 3 peuvent également être achetés. Veuillez communiquer par courriel avec le personnel du soutien technique de la Division de la statistique de la santé à hd-ds@statcan.ca ou avec un des bureaux régionaux de Statistique Canada.

12.2 Rapports analytiques et totalisations

Il est fréquent que des articles basés sur les données de l'ENSP apparaissent dans les *Rapports sur la santé*, une revue trimestrielle produite par la Division de la statistique de la santé. Ce produit est disponible en imprimé sous le numéro de catalogue 82-003-XPF ou en format électronique dans le site de Statistique Canada sous le numéro de catalogue 82-003-XIF. Pour plus d'informations, visiter notre site Web www.statcan.ca et sélectionner « Produits et services ».

Des tableaux gratuits longitudinaux Cansim sont aussi disponibles sur le site web de Statistique Canada. Ils présentent des changements, d'un cycle à l'autre de l'ENSP, de l'usage du tabac, de l'auto-évaluation de la santé, de l'indice de masse corporelle et de l'activité physique. On peut accéder aux tableaux Cansim, en cliquant sur «En santé aujourd'hui, en santé demain? Résultats de l'Enquête nationale sur la santé de la population» ([82-618-MWF](http://www23.statcan.ca/82-618-MWF)), ensuite sous «Tableaux de données».

Des tableaux personnalisés de l'ENSP peuvent aussi être obtenus à frais recouvrables. Pour des estimations de faisabilité et des coûts, communiquer avec le soutien technique de la Division de la statistique de la santé à hd-ds@statcan.ca.

Finalement, le Guide de la statistique de la santé (82-573-GIF) sur le site Web de Statistique Canada est un bon point de départ pour toute information reliée aux indicateurs de la santé provenant de différentes sources incluant l'ENSP. Le guide fait aussi le lien avec plusieurs documents relatifs à l'ENSP incluant les questionnaires. Pour consulter le Guide allez à http://www.statcan.ca/francais/freepub/82-573-GIF/guide_f.htm.

Annexe A : ENSP volet ménages, contenu du questionnaire au cours des 5 cycles

Thème	Sujet	Variable	Univers	1994-1995 (cycle 1)	1996-1997 (cycle 2)	1998-1999 (cycle 3)	2000-2001 (cycle 4)	2002-2003 (cycle 5)
Activités physiques		PA	>=12	P	P	P	P	P
Administration		AM	Tous les répondants	P	P	P	P	P
Alcool	Consommation	AL	>=12	P	P	P	P	P
	Consommation durant la grossesse	AL	Cycle 1: Femmes qui ont donné naissance à un enfant dans les 5 dernières années. Cycle 2: Femmes 15 à 49 ans, de l'Alberta seulement	EPS	CAP			
	Dépendance	AD	>=12		T			T
	Opinions	AL	Cycle 2: >=12 Aucun CA de l'Alberta. Cycle 5: >=12		EPS			T
Allaitement		BF GH	Cycle 1: Femmes qui ont donné naissance à un enfant dans les 5 dernières années. Cycle 2: EPS: Femmes >=15 et <=49 et qui ont donné naissance à un enfant depuis la dernière interview, CAP: Alberta	EPS	EPS CAP			
Antécédents	Médicaux de la famille	FH	>=18			T		
Antécédents résidentiels		RH MB	>=12					ACA
Assurance		DV EX DG IS	Cycle 2: >=12. Cycles 3 à 5: tous les répondants			T	P	P
Attitudes	Face aux parents	AP	12 à 17 ans en Alberta		CAP			

P: principal, **T:** thématique, **EPS:** Enquête promotion santé, **SA:** supplément sur l'asthme, **SPA:** Supplément sur la précarité alimentaire, **CAP:** Contenu acheté par une ou des provinces, **E:** Questions pour les enfants, **ACA:** Autre contenu acheté.

Thème	Sujet	Variable	Univers	1994-1995 (cycle 1)	1996-1997 (cycle 2)	1998-1999 (cycle 3)	2000-2001 (cycle 4)	2002-2003 (cycle 5)
Blessures	Blessures	IJ	Tous les répondants	P	P	P	P	P
	Prévention des blessures	INJ	>=12	EPS				
État de santé	Dextérité	HS	>=4	P	P	P	P	P
	Douleurs et malaises	HS	>=4	P	P	P	P	P
	Élocution	HS	>=4	P	P	P	P	P
	Mémoire	HS	>=4	P	P	P	P	P
	Mobilité	HS	>=4	P	P	P	P	P
	Ouïe	HS	>=4	P	P	P	P	P
	Pensée	HS	>=4	P	P	P	P	P
	Sentiments	HS	>=4	P	P	P	P	P
Vision	HS	>=4	P	P	P	P	P	
État de santé général	Auto-évaluation	GH	Tous les répondants	P	P	P	P	P
Exposition aux rayons ultraviolets et bronzage		TU	>=12. Cycle 2 - CAP: Alberta		CAP		P	P
Grossesse		HW PH	Femmes de 15 à 49 ans	P	P, EPS	P	P	P
Image corporelle		HW	>=12	EPS	EPS			T
Incapacité au cours des 2 dernières semaines		TW	Tous les répondants	P	P	P		
Limitation des activités		RA	Cycles 1 et 2: >=12. Cycles 3 à 5: tous les répondants	P	P	P	P	P
Médicaments - utilisation		DG	>=12	P	P	P	P	P
Mouvements répétitifs		RP	>=12		P	P	P	P
Nutrition	Auto-évaluation des habitudes alimentaires	GH	Tous les répondants					P
	Choix d'aliments	NU	Cycles 1 et 2: >=12. Cycles 3 et 5: >=15	EPS	EPS	T		T
	Fruits et légumes	FV	>=15					T
	Habitudes alimentaires	HW	>=12	EPS	EPS			
	Suppléments alimentaires	NU	>=15			T		T

P: principal, **T**: thématique, **EPS**: Enquête promotion santé, **SA**: supplément sur l'asthme, **SPA**: Supplément sur la précarité alimentaire, **CAP**: Contenu acheté par une ou des provinces, **E**: Questions pour les enfants, **ACA**: Autre contenu acheté.

Thème	Sujet	Variable	Univers	1994-1995 (cycle 1)	1996-1997 (cycle 2)	1998-1999 (cycle 3)	2000-2001 (cycle 4)	2002-2003 (cycle 5)
Poids corporel		HW	Tous les répondants	P	P	P	P	P
Précarité alimentaire		IN FI	Cycle 2: tous les répondants. Cycle 3: Tous les répondants identifiés dans un état de précarité alimentaire à la suite de leurs réponses aux questions filtres supplémentaires de l'ENSP Ménages		EPS	SPA		
Problèmes de santé chroniques	Acné nécessitant un traitement sur ordonnance	CC	<30	P				
	Allergies alimentaires	CC	Cycle 1: >=12. Cycle 2 à 5: tous les répondants	P	P	P	P	P
	Arthrite ou rhumatisme	CC	>=12	P	P	P	P T	P
	Asthme	CC	Cycle 1: >=12. Cycle 2 à 5: tous les répondants. Cycle 2 - SA: Tous les répondants >=12 qui ont déclaré souffrir d'asthme diagnostiqué par un professionnel de la santé dans l'enquête principale.	P	P SA	P	P	P
	Autres allergies	CC	Cycle 1: >=12. Cycle 2 à 5: tous les répondants	P	P	P	P	P
	Bronchite chronique ou emphysème	CC	Cycles 1 et 5: >=12. Cycle 2 à 4: tous les répondants	P	P	P	P	P

P: principal, **T:** thématique, **EPS:** Enquête promotion santé, **SA:** supplément sur l'asthme, **SPA:** Supplément sur la précarité alimentaire, **CAP:** Contenu acheté par une ou des provinces, **E:** Questions pour les enfants, **ACA:** Autre contenu acheté.

Thème	Sujet	Variable	Univers	1994-1995 (cycle 1)	1996-1997 (cycle 2)	1998-1999 (cycle 3)	2000-2001 (cycle 4)	2002-2003 (cycle 5)
Problèmes de santé chroniques	Cancer	CC	>=12	P	P	P	P	P
	Cataracte	CC	>=18	P	P	P	P	P
	Diabète	CC	Cycle 1 à 4: >=12. Cycle 5: tous les répondants	P	P	P	P T	P
	Enfants	CC	<=3	E	E	E		
	Épilepsie	CC	Cycle 1: >=12. Cycle 2 à 5: tous les répondants	P	P	P	P	P
	Fibromyalgie	CC	>=12				P	P
	Glaucome	CC	>=18	P	P	P	P	P
	Hypertension	CC	>=12	P	P	P	P	P
	Incontinence urinaire	CC	>=12	P	P	P	P	P
	Maladie cardiaque	CC	Cycle 1: >=12. Cycle 2 à 5: tous les répondants	P	P	P	P T	P
	Maladie d'Alzheimer ou autre forme de démence cérébrale	CC	>=18	P	P	P	P	P
	Maux de dos	CC	>=12	P	P	P	P	P
	Migraines	CC	Cycle 1 à 4: >=12. Cycle 5: tous les répondants	P	P	P	P	P
	Problème de la thyroïde	CC	>=12		P	P	P	P

P: principal, **T:** thématique, **EPS:** Enquête promotion santé, **SA:** supplément sur l'asthme, **SPA:** Supplément sur la précarité alimentaire, **CAP:** Contenu acheté par une ou des provinces, **E:** Questions pour les enfants, **ACA:** Autre contenu acheté.

Thème	Sujet	Variable	Univers	1994-1995 (cycle 1)	1996-1997 (cycle 2)	1998-1999 (cycle 3)	2000-2001 (cycle 4)	2002-2003 (cycle 5)
Problèmes de santé chroniques	Sinusite	CC	>=12	P	P	P		
	Troubles dus à un accident vasculaire cérébral	CC	>=12	P	P	P	P	P
	Troubles intestinaux tels que la maladie de Crohn ou la colite	CC	>=12	P	P	P	P	P
	Ulcères à l'estomac ou à l'intestin	CC	>=12	P	P	P	P	P
Renseignements socio-démographiques	Année d'immigration	SD	Tous les répondants	P	P	P		
	Appartenance ethnique	SD	Tous les répondants	P	P	P	P	
	Langues	SD	Tous les répondants	P	P	P	P	P
	Niveau de scolarité	ED	>=12	P	P	P	P	P
	Pays de naissance	SD	Tous les répondants	P	P	P		
	Population active	LF/LS	Cycle 1: >=15. Cycles 2 à 5: 15 à 75 ans	P	P	P	P	P
	Race	SD	Tous les répondants	P	P	P	P	
	Revenu	IN	Tous les répondants	P	P	P	P	P
Renseignements sur la santé	Informations sur la santé	A HI	EPS: >=12, CAP: Alberta 12 à 17 ans	EPS	CAP			
Santé des enfants		IJ DG	<12		P			
Santé des femmes		WH PH	Femmes >=18	P	P	P	P	P
Santé mentale		MH	>=12	P	P	P	P	P

P: principal, **T:** thématique, **EPS:** Enquête promotion santé, **SA:** supplément sur l'asthme, **SPA:** Supplément sur la précarité alimentaire, **CAP:** Contenu acheté par une ou des provinces, **E:** Questions pour les enfants, **ACA:** Autre contenu acheté.

Thème	Sujet	Variable	Univers	1994-1995 (cycle 1)	1996-1997 (cycle 2)	1998-1999 (cycle 3)	2000-2001 (cycle 4)	2002-2003 (cycle 5)
Santé préventive	Auto-examen des seins	WH	Femmes >=18		EPS			
	Dentiste - visites	DV	>=12		T			
	Examen de la vue	EX	>=12		T			
	Examen des seins	WH	Femmes >=18		P			
	Examen général	PC	>=12		T			
	Mammographie	WH PH	Femmes >=35	P	P	P	P	P
	Mesures pour améliorer la santé	GH A	>=12	EPS	EPS			
	Tension artérielle	BP, PH	>=12	P	P	P	P	P
	Test de Papanicolaou	WH PH	Femmes >=18	P	P	P	P	P
	Vaccination antigrippale	FS	>=12		T			
	VIH	SH HV	Cycle 1: 15 à 45 ans. Cycle 2: >=18	EPS	EPS			
Santé sexuelle		SH	Cycle 1: 15 à 45 ans. Cycle 2: 15 à 59. CAP: Alberta	EPS	EPS, CAP			
Sécurité personnelle et violence		VS	>=12 Alberta		CAP			
Sécurité routière - conduite sous influence		RS	>=12		EPS			
Soins de santé	Services d'urgence	ES	>=12		T			
	Utilisation	SV HC	Principal: Tous les répondants. EPS: >=20. CAP: >=12 Alberta	P, EPS	P, CAP	P	P	P
Soins de santé à domicile		HC	>=18	P	P	P	P	P
Soins personnels		SC	>=18			T		
Sommeil		SL	>=12					T
Soutien social		SS	>=12. Cycle 2- CAP: Alberta	P	P, EPS, CAP	P	P	P

P: principal, **T:** thématique, **EPS:** Enquête promotion santé, **SA:** supplément sur l'asthme, **SPA:** Supplément sur la précarité alimentaire, **CAP:** Contenu acheté par une ou des provinces, **E:** Questions pour les enfants, **ACA:** Autre contenu acheté.

Thème	Sujet	Variable	Univers	1994-1995 (cycle 1)	1996-1997 (cycle 2)	1998-1999 (cycle 3)	2000-2001 (cycle 4)	2002-2003 (cycle 5)
Stress	Auto perception	ST	Cycle 4: >=18. Cycle 5: >=12				P	P
	Capacité de faire face au stress	CO	Cycle 1: >=18 sans procuration Manitoba et Alberta seulement. Cycle 2: Alberta >=18	CAP	CAP			
	Contrôle	PY	Cycle 1 à 5: >=12	T			T	T
	Estime de soi	PY	>=12	T			T	
	Événements récents	ST	>=18	T			T	
	Facteurs de stress chez les enfants et les adultes (traumatismes)	ST	Cycle 1: >=18. Cycle 4: Répondants âgés de moins de 18 ans au cycle 1 et âgés de 18 ans et plus au cycle 4	T			T	
	Problèmes actuels	ST	>=18	T			T	T
	Sentiment de cohésion	PY	>=18	T		T		
Tabac	Stress au travail	ST	Cycle 1: >=15. Cycles 4 à 5: 15 à 75 ans	T			T	T
	Attitudes	SM	>=12		EPS			
	Connaissances, attitudes, cessation	SM	>=12	EPS				
	Pendant la grossesse	SM	Cycle 2: Femmes de 15 à 49 ans		EPS			
	Usage du tabac	SM GH	>=12	P	P, EPS	P	P	P
Taille	Variants du tabagisme	TA	>=12			EPS		
		HW	Tous les répondants	P	P	P	P	P

P: principal, **T:** thématique, **EPS:** Enquête promotion santé, **SA:** supplément sur l'asthme, **SPA:** Supplément sur la précarité alimentaire, **CAP:** Contenu acheté par une ou des provinces, **E:** Questions pour les enfants, **ACA:** Autre contenu acheté.

Annexe B : ENSP, volet ménages, changements au questionnaire pour le cycle 5 (2002-2003)

1. Introduction

Cette annexe décrit les changements apportés au questionnaire entre 2000-2001 et 2002-2003. Des questions de suivi ont été modifiées et quelques autres questions ayant les mêmes réponses dans le temps, comme l'origine ethnique et la race, ont été supprimées au cycle 5. Le contenu principal et la formulation de certaines questions ont été améliorés. À l'instar des autres cycles, le questionnaire contient des questions thématiques. Au cycle 5, certaines questions relatives au stress des cycles 1 et 4 ont été répétées alors que d'autres ont été supprimées. Le sommeil et l'image corporelle sont nouveaux au cycle 5. La nutrition, la dépendance à l'alcool et le supplément relatif à l'alcool des cycles 2 et 3 ont aussi été intégrés. Certaines questions sur le diagnostic et la gestion de problèmes de santé chroniques ont été reformulées ou supprimées. Le contenu ajouté au cycle 5 portait sur les antécédents résidentiels. Pour plus de détails, veuillez consulter les questionnaires spécifiques à chacun des cycles.

2. Changements à la structure du questionnaire

Comme les questions ne sont posées dorénavant qu'au répondant longitudinal, l'ordre du questionnaire est le même qu'au cycle 4. Le contenu thématique et le contenu ajouté ont été incorporés à la partie la plus appropriée du questionnaire ou le module a été inséré séparément au contenu principal, pour le cycle 5.

3. Changements au contenu principal de l'enquête

Pour ce qui suit, les noms de variables externes du questionnaire sont utilisés. Quelques-uns des noms internes ont été renommés pour assurer la cohérence dans le questionnaire, ceci est causé par la suppression ou l'addition de questions. Les noms des variables du fichier maître et du fichier partagé suivent la convention utilisée pour nommer les variables.

Variable du dossier ménage

- Addition / révision : question sur le type de logement (DHC2_DWE) - changement pour la catégorie de réponse #2 « Maison double (semi-détachée) » pour « Maison double ».
- Addition / révision : question sur la source de l'information (AM32_SRC) - changement de terme.

État de santé général (GH)

- Révision de l'univers (GH_C2) – les questions sont dorénavant posées aux personnes âgées de moins de 12 ans et non aux personnes âgées de moins de 18 ans. De « Si l'âge < 18 » à « Si l'âge < 12 ».

- Addition : nouvelle question sur l'évaluation personnelle des habitudes alimentaires (GHC2_4).

Santé préventive (PH)

- Addition : question sur l'hystérectomie (WHC2_5) - changement de terme dans la question.

Utilisation des soins de santé (HC)

- Addition / révision : changement de terme pour l'univers (HC_C02).
- Addition / révision : question sur la consultation auprès d'un intervenant en médecine douce (HCC2_4) - changement de terme.
- Addition / révision : question sur les services à domicile (HCC2_9) - changement de terme.
- Addition / révision : question sur les soins infirmiers reçus (HCC2_10A) - changement de terme.
- Addition : question sur les services à domicile non couverts par le gouvernement (HCC2_11) - changement de terme.
- Addition : question sur l'offre des services à domicile (HCC2_12A à HCC2_12F) - choisissez toutes les réponses appropriées.
- Addition : question sur le genre de services reçus (HCC2_3AA à HCC2_3FH) - choisissez toutes les réponses appropriées.

Limitation des activités (RA)

- Addition / révision : question sur l'apparition ou l'aggravation de limitation des activités (RAC2_2A) - changement de terme.
- Addition / révision : question sur la disparition ou l'amélioration de limitation d'activité (RAC2_2B) - changement de terme.
- Addition / révision : clarification de l'univers (RA_C5).
- Addition : question sur le besoin d'aide pour se déplacer à l'extérieur peu importe le climat - (RAC2_6G).

Problèmes de santé chroniques (CC)

Le module des problèmes de santé chroniques est totalement intégré au contenu principal du cycle 5, contrairement au cycle 4 où il y avait des questions qui faisaient partie du contenu thématique.

Allergies Alimentaires

- Suppression : question sur l'âge au premier diagnostic d'allergies alimentaire (CCC0_A3).

Asthme

- Révision : question sur le mois et l'année du diagnostic d'asthme (CCC2_C3M et CCC2_C3Y) - remplace la question sur l'âge du premier diagnostic d'asthme (CCC0_C3).
- Addition : question sur l'asthme avant la dernière interview (CCC2_C4).

Fibromyalgie

- Révision: question sur le mois et l'année du diagnostic de fibromyalgie (CCC2_X3M et CCC2_X3Y) - remplace la question sur l'âge au premier diagnostic de fibromyalgie (CCC0_X3).
- Addition : question sur la fibromyalgie avant la dernière interview (CCC2_X4).
- Addition : question la disparition de la fibromyalgie depuis la dernière interview (CCC2_X1).
- Addition : question sur le mois et l'année de la disparition de la fibromyalgie (CCC2_X2M et CCC2_X2Y)
- Addition : question sur le traitement ou les médicaments contre la fibromyalgie (CCC2_X5).
- Addition : question sur le type de traitement ou de médicaments contre la fibromyalgie (CCC2_X6A à CCC2_X6D) - choisissez toutes les réponses appropriées.

Arthrite ou rhumatisme, excepté la fibromyalgie

- Révision : question sur le mois et l'année du diagnostic d'arthrite ou de rhumatisme, excepté la fibromyalgie (CCC2_D3M et CCC2_D3Y) - remplace la question sur l'âge au premier diagnostic d'arthrite ou de rhumatisme, excepté la fibromyalgie (CCC0_D3).
- Addition : question sur l'arthrite ou de rhumatisme, excepté la fibromyalgie avant la dernière interview (CCC2_D4).
- Addition : question sur la disparition de l'arthrite ou du rhumatisme, excepté la fibromyalgie depuis la dernière interview (CCC2_D1).
- Addition : question sur le mois et l'année de la disparition de l'arthrite ou le rhumatisme, excepté la fibromyalgie (CCC2_D2M et CCC2_D2Y).
- Addition : question sur le genre d'arthrite (CCC2_D11) - fait maintenant partie du contenu principal, et non du contenu thématique (CC_0_D1).
- Addition / révision : question sur le traitement ou les médicaments contre l'arthrite ou le rhumatisme (CCC2_D5) - fait maintenant partie du contenu principal, et non du contenu thématique (CC_0_D3), aussi changement de terme dans la question.
- Suppression : question sur une intervention chirurgicale pour l'arthrite ou le rhumatisme (CC_0_D2).
- Suppression : questions sur le type d'intervention chirurgicale pour l'arthrite ou le rhumatisme (CC_0_D2A à CC_0_D2D) - choisissez toutes les réponses appropriées.
- Suppression : questions sur le genre de traitement pour soulager l'arthrite ou le rhumatisme (CC_0_D3A à CC_0_D3H) - choisissez toutes les réponses appropriées.
- Suppression : questions sur le type de médicament pour soulager l'arthrite ou le

- rhumatisme (CC_0_D4A à CC_0_D4C) - choisissez toutes les réponses appropriées.
- Addition : question sur le type de traitement ou médicament pour l'arthrite ou le rhumatisme (CCC2_D6A à CCC2_D6D) - choisissez toutes les réponses appropriées.

Maux de dos

- Suppression : question sur l'âge au premier diagnostic de maux de dos (CCC0_E3).

Hypertension

- Révision : question sur le mois et l'année du diagnostic d'hypertension (CCC2_F3M et CCC2_F3Y) - remplace la question sur l'âge au premier diagnostic d'hypertension (CCC0_F3).
- Addition : question sur l'hypertension avant la dernière interview (CCC2_F4).
- Révision : changement de la condition et changement d'expression (CC_C073).

Migraines

- Suppression : révision de l'univers (CC_C081). Les questions sont dorénavant posées à tous les répondants - changement de « Si âge < 12 » à aucune condition.
- Révision : question sur le mois et l'année du diagnostic de migraines (CCC2_G3M et CCC2_G3Y) - remplace la question sur l'âge au premier diagnostic de migraines (CCC0_G3).
- Addition : nouvelle question les migraines avant la dernière interview (CCC2_G4).

Bronchite Chronique ou Emphysème

- Addition : l'univers (CC_C091). Les questions sont posées aux personnes âgées de 12 ans et plus - condition « Si âge < 12 ».
- Suppression : question sur l'âge au premier diagnostic de bronchite chronique ou emphysème (CCC0_H3).

Diabète

- Suppression : révision de l'univers (CC_C101). Les questions sont dorénavant posées à tous les répondants - changement de « Si âge < 12 » à aucune condition.
- Révision : question sur le mois et l'année du diagnostic de diabète (CCC2_J3M et CCC2_J3Y) - remplace la question sur l'âge au premier diagnostic de diabète (CCC0_J3).
- Addition : nouvelle question sur le diabète avant la dernière interview (CCC2_J4).
- Suppression : question sur le diagnostic de diabète durant la grossesse (CC_0_J3A).
- Suppression : question sur le diagnostic de diabète autre que durant la grossesse (CC_0_J3B).
- Suppression : question sur le laps de temps entre le diagnostic et la prise d'insuline (CC_0_J3C).

- Suppression : question sur l'insuline prise chaque jour (CC_0_J4).
- Suppression : question sur la fréquence quotidienne de prise d'insuline (CC_0_J4A).
- Suppression : question sur le nombre d'unités d'insuline par jour (CC_0_J4B).
- Suppression : question sur les cours pour contrôler le diabète (CC_0_J5A).
- Suppression : question sur les membres du ménage qui ont suivi un cours pour contrôler le diabète (CC_0_J5B).
- Suppression : question sur les professionnels de la santé qui donnent des renseignements pour contrôler le diabète (CC_0_J5C).
- Suppression : question sur le type de professionnel de la santé qui donne des renseignements pour contrôler le diabète (CC_0_J6A à CC_0_J6E) - choisissez toutes les réponses appropriées.
- Suppression : question sur l'apprentissage à vérifier la concentration de sucre (CC_0_J7A).
- Suppression : question sur un membre du ménage qui a appris à vérifier la concentration de sucre (CC_0_J7B).
- Suppression : question sur la fréquence de vérification de la concentration de sucre dans le sang (CC_0_J8A).
- Suppression : question sur la fréquence d'examen des pieds (CC_0_J8B).

Épilepsie

- Révision : question sur le mois et l'année du diagnostic d'épilepsie (CCC2_K3M et CCC2_K3Y) - remplace la question sur l'âge au premier diagnostic d'épilepsie (CCC0_K3)
- Addition : question sur l'épilepsie avant la dernière interview (CCC2_K4)

Maladie cardiaque

- Suppression : question sur l'âge du premier diagnostic de maladie cardiaque (CCC0_L3)
- Suppression : question sur le nombre de crises cardiaques (CC_0_L1B).
- Suppression : question sur l'âge à la première crise cardiaque (CC_0_L1C).
- Suppression : question sur l'âge à la crise la plus récente (CC_0_L1D).
- Suppression : question sur l'admission à l'hôpital pour une crise cardiaque (CC_0_L1E).
- Suppression : questions sur les traitements médicaux reçus pour une crise cardiaque (CC_0_L4A à CC_0_L4E) - choisissez toutes les réponses appropriées.
- Suppression : question sur un programme de réadaptation cardiologique (CC_0_L5A).
- Suppression : question sur la participation à un programme de réadaptation cardiologique (CC_0_L5B).
- Suppression : question sur l'achèvement d'un programme de réadaptation cardiologique (CC_0_L5C).
- Addition : question sur l'apparition d'une crise cardiaque (CCC2_L1A) - fait maintenant partie du contenu principal et non du contenu thématique (CC_0_L1A).
- Addition: question sur l'angine (CCC2_L6) - fait maintenant partie du contenu principal et non du contenu thématique (CC_0_L6).

- Addition : question sur l'insuffisance cardiaque globale (CCC2_L7) - fait maintenant partie du contenu principal et non du contenu thématique (CC_0_L7).

Cancer

- Suppression : révision de l'univers (CC_C131). Les questions sont dorénavant posées à tous les répondants - changement de « Si âge < 12 » à aucune condition.
- Suppression : question sur l'âge au premier diagnostic de cancer (CCC0_M3)

Ulcères à l'intestin ou à l'estomac

- Révision : termes renversés pour intestin et estomac (CCC2_N1)
- Révision : question sur le mois et l'année du diagnostic d'ulcères (CCC2_N3M et CCC2_N3Y) - remplace la question sur l'âge au premier diagnostic d'ulcères (CCC0_N3)
- Addition : nouvelle question les ulcères avant la dernière interview (CCC2_N4)

Troubles dus à un accident vasculaire cérébral

- Révision : question sur le mois et l'année du diagnostic de troubles dus à un accident vasculaire cérébral (CCC2_O3M et CCC2_O3Y) - remplace la question sur l'âge au premier diagnostic de troubles dus à un accident vasculaire cérébral (CCC0_O3).
- Addition : nouvelle question sur les troubles dus à un accident vasculaire cérébral avant la dernière interview (CCC2_O4).

Incontinence urinaire

- Suppression : question sur l'âge au premier diagnostic d'incontinence urinaire (CCC0_P3).

Troubles intestinaux

- Suppression : question sur l'âge au premier diagnostic de troubles intestinaux (CCC0_Q3).

La maladie d'Alzheimer ou autre forme de démence cérébrale

- Addition / révision : question sur la maladie d'Alzheimer ou autre forme de démence cérébrale (CCC2_1R) - clarification du mot « sénilité ».
- Suppression : question sur l'âge au premier diagnostic de la maladie d'Alzheimer ou autre forme de démence cérébrale (CCC0_R3).

Cataractes

- Suppression : question sur l'âge au premier diagnostic de cataractes (CCC0_S3).

Glaucome

- Suppression : question sur l'âge au premier diagnostic de glaucome (CCC0_T3).

Problème de la thyroïde

- Suppression : question sur l'âge au premier diagnostic de problème de la thyroïde (CCC0_U3).

Assurance (IS)

- Aucun changement.

État de santé (HS)

- Aucun changement.

Activités physiques (PA)

- Addition / révision : question sur l'activité au cours des 3 derniers mois (PAC2_1I), changement dans la catégorie de réponse #13 à « ski alpin ou planche à neige ».

Exposition aux rayons ultraviolets (UV) (TU)

- Aucun changement.

Mouvements répétitifs (RP)

- Aucun changement.

Blessures (IJ)

- Addition / révision : question sur les soins 48 heures après la blessure (IJC2_11) - changement dans la phrase.

Consommation de médicaments (DG)

- Addition / révision : question sur la consommation de médicaments (DGC2_1B) - changement pour les exemples de tranquillisants « tels que le Valium ou l'Ativan ».
- Addition / révision : question sur la consommation de médicaments (DGC2_1C) - changement pour les exemples de pilules pour maigrir « tels que le Ponderal, le Dexatrim ou le Fastin ».

- Addition / révision : question sur la consommation de médicaments (DGC2_1F) - changement pour les exemples de médicaments contre les allergies « tels que le Reactine ou l'Allegra ».
- Addition / révision : question sur la consommation de médicaments (DGC2_1L) - changement pour les exemples de diurétiques « ou pilules contre la rétention d'eau ».
- Addition / révision : question sur la consommation de médicaments (DGC2_1P) - changement pour les exemples de somnifères « tels que l'Imovane, le Nytol ou le Starnoc ».
- Addition / révision : question sur le type de médicament: ordonnance d'un médecin ou d'un dentiste jusqu'à concurrence de 12 médicaments. (DGC2_3AA à DGC2_3LA), changement de docteur en médecine à médecin.
- Addition / révision : indicateur de réponse élaborée pour le nom exact du premier produit utilisé pour la santé (DGC2F5A) - changement dans la phrase à « quel est le nom exact d'un des produits pour la santé utilisés ».
- Addition / révision : question sur l'utilisation d'un autre produit pour la santé (DGC2_5AA à DGC2_5LA) - changement dans la phrase à « Avez-vous utilisé un autre produit pour la santé » - jusqu'à concurrence de 12 produits.
- Addition / révision : indicateur de réponse élaborée pour le nom exact du produit suivant utilisé pour la santé (DGC2F5B à DGC2F5L) - changement dans la phrase à « quel est le nom exact de ce produit » - jusqu'à concurrence de 11 produits.

Usage du tabac (SM)

- Addition / révision: changement dans l'usage du tabac depuis la dernière interview (SMC2_4B).
- Addition : question sur la marque de cigarette généralement fumée (SMC2C8B).
- Addition / révision : question sur la raison d'avoir commencé à fumer (SMC2_10) - changement dans la catégories de réponse # 3 « Tout le monde autour de moi fume ».
- Addition / révision : question sur la raison d'avoir augmenté de fumer (SMC2_12) - changement dans la catégories de réponse # 3 « Tout le monde autour de moi fume ».
- Addition / révision : question sur les restrictions de l'usage du tabac au travail (SMC2_2_6) - changement dans la catégories de réponse # 3 « Défendu seulement à certains endroits ».

Consommation d'alcool (AL)

- Addition / révision : question sur la consommation d'alcool (ALC2_5A1 - ALC2_5A7) - changement dans la phrase « Combien de verres avez-vous bus hier, c'est-à-dire » - jusqu'à concurrence de 7 jours.

Santé mentale (MH)

- Addition / révision : question sur la consultation auprès d'un professionnel pour un problème de santé mentale (MHC2_1K) - Changement de terme dans la question.

Soutien social (SS)

- Aucun changement.

Renseignements socio-démographique (SD)

Appartenance ethnique

- Suppression : question sur l'appartenance ethnique (question statique déjà posée) (SDC0_4A à SDC0_4S) - choisissez toutes les réponses appropriées.

Race

- Suppression : question sur la race (question statique déjà posée) (SDC0_7A à SDC0_7L) - choisissez toutes les réponses appropriées.

Langue

- Aucun changement.

Niveau de scolarité (ED)

- Aucun changement.

Population active (LF / LS)

- Aucun changement.

Revenu (IN)

- Aucun changement.

Numéro provincial d'assurance-maladie et Administration (AM)

- Addition / révision : question sur la permission de jumeler les données (AM62_LNK) - Changement de terme dans la question.
- Addition / révision : question sur le partage des données (AM62_SHA) - Changement de terme dans la question.
- Addition / révision : question sur la correction au nom (AM62_15) - Changement de terme dans la question.

4. Changements pour la composante thématique

De nouvelles sections sur le Sommeil et les Antécédents résidentiels sont introduites. Les questions sur la nutrition du cycle 3 sont incluses. Aussi, quelques-unes des questions du module sur le stress (Problèmes actuels, Stress au travail et Contrôle), qui ont été posées au cycle 1 et au cycle 4, sont répétées. Le supplément sur l'attitude face à l'alcool et le module sur la dépendance à l'égard de l'alcool qui ont été demandés au cycle 2 sont aussi répétés au cycle 5.

La composante thématique suivante est incluse pour le cycle 5 :

Sommeil (SL)

- Addition : la sélection de 4 nouvelles questions prises du module sur le sommeil de l'ESCC (SL_2_1 à SL_2_4).

Taille et poids (HW)

Image corporelle

- Addition : 5 nouvelles questions (HW_2_4 à HW_2_8) et une condition (BI_C1).

Nutrition (NU)

Les questions sur le choix alimentaire et les suppléments alimentaires du cycle 3 sont répétées avec les changements suivants :

Choix alimentaire

- Suppression : question sur le choix ou évitement d'aliments pour maintenir ou améliorer la santé (NU_8_1B).
- Suppression : question sur le choix ou évitement d'aliments en raison de préoccupations pour l'hypertension (NU_8_1F).
- Suppression : question sur le choix ou évitement d'aliments en raison de préoccupations pour le diabète (NU_8_1G).
- Suppression : question sur le choix d'aliments pour leur teneur en fer (NU_8_2D).
- Suppression : question sur le choix d'aliments pour leur teneur en autres vitamines ou minéraux (NU_8_2E).
- Suppression : question sur l'évitement d'aliments pour leur teneur en sucre (NU_8_3E).
- Suppression : question sur l'évitement d'aliments pour leur teneur en fer (NU_8_3F).

Suppléments alimentaires

- Aucun changement.

Consommation de fruits et de légumes - FV

- Addition - 12 nouvelles questions prises du module sur la consommation de fruits et de légumes de l'ESCC (FV_2_Q1 à FV_2_Q6) - Chacune des questions contiennent le nombre de fois et la période déclarante.

Stress (ST)

- Ces questions sont les mêmes que celles posées aux cycle 1 et 4 avec les changements suivants:

Problèmes actuels

- Addition : nouveau terme (ST_QINT1B).

Événements récents

- Suppression : questions sur les événements récents (ST_0_R1 à ST_0_R10).

Facteur de stress chez les enfants et les adultes « traumatismes »

- Suppression : questions sur les traumatismes (ST_0_T1 à ST_0_T7).

Stress au travail

- Aucun changement.

Estime de soi

- Suppression : questions sur l'estime de soi (PY_0_E1A à PY_0_E1F).

Contrôle

- Aucun changement.

Alcool (AL)

Supplément sur l'attitude face à l'alcool

- Ces questions sont les mêmes que celles du cycle 2, du supplément national sur l'alcool (AL_2_1 à AL_2_7).

Dépendance à l'égard de l'alcool (AD)

- Les questions sont les mêmes que celles du cycle 2 (AD_2_1 à AD_2_9).

5. Changements à l'achat de contenu

Antécédents résidentiels (RH / MB)

- Addition : le module des antécédents résidentiels au cours des vingt dernières années est nouveau pour le cycle 5. Il y a deux façons d'y répondre: soit à partir du lieu de résidence au moment de l'interview en remontant les 20 dernières années, soit à partir de 1980 (ou année de naissance) jusqu'à la date de l'interview.

Annexe C : ENSP volet ménages, exemples de rétroaction des données du cycle 4 (2000-2001) et de questions de suivi

Hypertension, mammographie, test PAP	Lors du premier cycle et du deuxième cycle, on a demandé à la personne interrogée si on avait « déjà » pris sa tension artérielle (ou si elle avait « déjà » passé une mammographie, etc.). Lors du cycle 3 les questions ont été répétées. Dans les cas où la personne a indiqué qu'elle n'avait pas subi le test en question mais qu'elle avait dit l'avoir subi lors d'un cycle antérieur, on a posé des questions supplémentaires. Lors des cycles 4 et 5, dans le cas où le répondant avait déclaré avoir eu un test lors d'une interview antérieure, seule la question « À quand remontait la dernière fois ? » était posée.
Limitation des activités	On s'est servi des données indiquant si le répondant souffrait ou non d'une incapacité au cycle 3. En cas de changement de situation, on a posé des questions pour obtenir une explication.
Problèmes de santé chroniques	Pour chaque répondant, on a utilisé les réponses du cycle précédent concernant certains problèmes de santé chroniques (asthme, fibromyalgie, arthrite, hypertension, migraine, diabète, épilepsie, ulcères de l'estomac ou de l'intestin et les séquelles d'un accident cérébro-vasculaire) pour essayer d'obtenir des explications concernant le changement. Si le problème était nouveau, on a enregistré la date de sa manifestation.
Usage du tabac	Si un fumeur quotidien avait déclaré lors de la dernière interview l'âge auquel il ou elle avait commencé à fumer tous les jours, cette réponse était utilisée au cycle 5. En ce qui concerne le fumeur occasionnel ou le non-fumeur au cycle 5 qui avait déclaré fumer (ou avoir déjà fumé) tous les jours lors d'une interview antérieure, on a utilisé un indicateur d'usage quotidien. En cas de changement de situation, on a posé des questions pour obtenir une explication.
Renseignements socio-démographiques	Pour chaque répondant, on a utilisé un indicateur montrant que les renseignements concernant le pays de naissance ont été recueillis. Enfin, on a de nouveau posé la question sur la première langue apprise et encore parlée, parce que la réponse à cette question peut changer avec le temps.

Niveau de scolarité	Pour chaque répondant, on a utilisé un indicateur montrant le plus haut niveau de scolarité atteint. On s'est servi de questions filtres pour déterminer si le répondant avait fréquenté un établissement d'enseignement entre les deux cycles. Le cas échéant, on a de nouveau recueilli les renseignements sur le niveau de scolarité.
Population active	À chaque répondant, on a rappelé le nom de l'employeur, le secteur d'activité et le genre de tâches correspondant à l'emploi principal mentionné à l'interview du cycle 4. Aux répondants qui ont dit avoir travaillé l'année précédente, on a demandé de confirmer le nom de l'employeur. Dans le cas de changement, les renseignements ont été saisis.
Numéro d'assurance-maladie	On a utilisé un indicateur montrant si le numéro d'assurance-maladie recueilli lors d'un cycle antérieur était « valide ». Si le numéro d'assurance-maladie était invalide et que le répondant avait donné le même numéro que celui enregistré lors du dernier cycle, on a demandé au répondant de donner son numéro d'assurance-maladie à nouveau.