

Enquête nationale sur la santé de la population – une enquête longitudinale

Larry Swain, Gary Catlin et Marie P. Beaudet

Résumé

Objectifs

Le présent article expose certains avantages qu'offrent les données recueillies auprès d'un panel longitudinal, ainsi que certains défis que pose la collecte de ces données, en prenant pour exemple l'Enquête nationale sur la santé de la population (ENSP).

Source des données

L'ENSP permet de recueillir des données transversales et longitudinales auprès d'un échantillon de personnes sélectionnées au hasard. Les membres de l'échantillon longitudinal se prêteront à une entrevue tous les deux ans pendant une période allant jusqu'à 20 ans. Deux cycles ont été exécutés jusqu'à présent : le cycle 1 en 1994-1995 et le cycle 2 en 1996-1997.

Sommaire

Quelques résultats de l'ENSP illustrent dans un premier temps les avantages des données longitudinales. L'article donne ensuite un aperçu du contenu du questionnaire, des méthodes de collecte des données et enchaîne en décrivant le plan d'échantillonnage. Un sommaire des taux de réponse est ensuite présenté, suivi d'une discussion des méthodes utilisées pour maintenir ces taux et pour corriger les poids d'échantillonnage afin de réduire le biais dû à la non-réponse. Enfin, certaines questions sont soulevées en ce qui concerne la nature confidentielle des renseignements, la diffusion des données, les incohérences de déclaration, la déclaration par personne interposée et la modification des conventions de codage.

Mots-clés

Enquêtes-santé, études longitudinales.

Auteurs

Larry Swain (613) 951-8569; swailar@statcan.ca, Gary Catlin (613) 951-8571; catlgar@statcan.ca, et Marie P. Beaudet (613) 951-7025; beaumar@statcan.ca travaillent à la Division des statistiques sur la santé, Statistique Canada, Ottawa K1A 0T6.

Certains problèmes de santé ont une relation de cause à effet bien évidente. Ainsi, l'exposition à un virus peut provoquer un accès de grippe.

Un accident de la circulation peut, quant à lui, causer des fractures. Cependant, dans la plupart des cas, le lien entre la cause et l'effet est loin d'être aussi manifeste. Par exemple, les personnes dont le niveau de scolarité et le revenu sont élevés ont tendance à être en meilleure santé que celles dont le niveau de scolarité et le revenu sont plus faibles. Mais est-ce le statut socioéconomique élevé qui permet de vivre dans des conditions propices à un bon état de santé ou est-ce cet état de santé qui permet à une personne d'atteindre un statut socioéconomique de haut niveau?

Habituellement, les efforts en vue de répondre à ce genre de questions s'appuient sur des enquêtes transversales qui fournissent des renseignements sur les situations qui prévalent à un point précis dans le temps. On fait donc des

Le présent article s'inspire de la communication intitulée The National Population Health Survey : Its Longitudinal Nature, présentée par Larry Swain et Gary Catlin à la conférence conjointe de l'AISE/AISO tenue en septembre 1998 sous le thème Statistics for Economic and Social Development.

inférences quant aux effets de certaines caractéristiques sur la santé en comparant les groupes qui présentent et qui ne présentent pas les caractéristiques étudiées. Toutefois, au lieu de recueillir des renseignements à divers intervalles auprès de personnes différentes en réalisant plusieurs enquêtes transversales, il serait préférable d'interroger les mêmes personnes à plusieurs reprises afin de déceler les modifications de leurs caractéristiques au fil du temps et de voir si ces modifications sont associées à des changements de leur état de santé (enquête par panel). Même à l'aide de données longitudinales de ce genre, la relation de causalité demeure difficile à cerner, mais les preuves sont plus convaincantes puisqu'on possède certains renseignements sur l'enchaînement des événements. Les données empiriques et théoriques existantes doivent malgré tout orienter les travaux de recherche et l'interprétation des résultats.

L'Enquête nationale sur la santé de la population (ENSP) a été mise sur pied par Statistique Canada pour étudier la dynamique de la santé. Cette enquête est conçue pour recueillir des données transversales et longitudinales sur la santé physique et psychique des Canadiens et des Canadiennes et sur l'utilisation qu'ils font des services de santé. Les objectifs généraux sont les suivants ¹:

- fournir des données sur l'état de santé de la population et sur la relation entre celui-ci, d'une part, et l'utilisation des services de santé et les déterminants de la santé, d'autre part;
- recueillir des données sur les liens entre la santé et les facteurs économiques, sociaux, démographiques, professionnels et environnementaux;
- suivre un panel de personnes dans le temps pour refléter la dynamique de la santé et de la maladie;
- permettre de faire des liens entre les données de l'enquête et des données administratives courantes, dont les statistiques de l'état civil, les mesures environnementales, les variables communautaires et les données sur l'utilisation des services de santé.

Le présent article donne un aperçu du contenu et des méthodes de collecte des données de l'ENSP. On y décrit certaines stratégies méthodologiques et

opérationnelles particulières qu'il faut appliquer aux composantes longitudinale et transversale de l'enquête réalisée auprès des ménages dans les 10 provinces.

Le premier cycle de collecte de données a eu lieu en 1994-1995 et le deuxième, en 1996-1997. Le troisième a débuté en juin 1998 et se poursuivra jusqu'en juin 1999. Seules des données de deux cycles sont donc jusqu'ici disponibles pour l'analyse. Malgré cela, les résultats témoignent de la valeur unique des renseignements longitudinaux.

En deux ans seulement...

Certains facteurs physiques, sociaux et environnementaux ont sans doute sur la santé des effets qui ne se manifestent qu'après de nombreuses années. Pourtant, les données recueillies, après deux cycles seulement de l'ENSP, donnent à penser que certains de ces facteurs exercent une action importante^{2,3}.

Par exemple, deux ans seulement après le premier cycle, on constate les effets nocifs de l'usage du tabac. Chez les personnes qui n'avaient déclaré aucune limitation dans leurs activités en 1994-1995, la cote exprimant le risque d'être limité dans les activités de la vie quotidienne à cause d'une maladie respiratoire, d'une maladie cardiaque ou d'une autre maladie circulatoire est deux fois plus élevée pour les personnes qui fument régulièrement que pour celles qui n'ont jamais fumé. Cette observation persiste, même après avoir corrigé les données pour tenir compte des effets de l'âge, du sexe, du niveau de scolarité et du niveau de revenu du ménage. En outre, la cote exprimant le risque qu'une personne qui avait entre 45 et 84 ans en 1994-1995 soit décédée en 1996-1997 est deux fois plus élevée pour les fumeurs que pour les non-fumeurs.

Les problèmes de santé chroniques et les incapacités fonctionnelles sont en général plus fréquents chez les personnes âgées que chez les autres. Cependant, l'évolution n'a pas nécessairement lieu à sens unique. Parmi les personnes de 65 ans et plus qui présentaient un handicap physique en 1994-1995 et qui avaient à ce moment-là besoin d'aide pour se livrer à des activités comme préparer les repas ou faire les courses, une

sur cinq n'avait plus besoin de cette aide deux ans plus tard. On note aussi des améliorations de la capacité fonctionnelle chez certaines personnes âgées qui faisaient face à des limitations plus graves de l'activité et qui devaient compter sur d'autres pour des soins plus fondamentaux comme se laver, s'habiller ou manger.

Les données transversales donnent parfois l'impression d'une stabilité que ne confirme pas l'examen des données longitudinales. Ainsi, on estime qu'en 1994-1995, 2,4 % de la population de 18 ans et plus a reçu une certaine forme de soins à domicile financés par l'État. À 2,5 %, l'estimation calculée pour 1996-1997 est comparable. Cependant, 36 % seulement des personnes qui avaient reçu des soins en 1994-1995 en recevaient encore en 1996-1997. Presque la moitié d'entre elles (46 %) vivaient chez elles, mais ne recevaient plus de soins à domicile, 14 % étaient décédées et 4 % avaient été placées en établissement de soins.

Les données transversales indiquent que la prévalence du tabagisme diminue chez les personnes de 15 ans et plus, étant passée de 31 % en 1994-1995 à 29 % en 1996-1997. Toutefois, les données longitudinales renseignent sur les changements qui sous-tendent cette diminution globale. Selon ces données, 6 % de la population qui ne fumait pas en 1994-1995 avait commencé à fumer pour la première fois ou avait recommencé à fumer en 1996-1997. En revanche, 14 % des personnes qui fumaient en 1994-1995 avaient renoncé au tabac en 1996-1997.

Ces premiers résultats témoignent de la valeur des données que l'on peut tirer d'une enquête longitudinale. Ce genre d'enquête oblige toutefois à prendre plusieurs mesures d'ordre opérationnel et méthodologique afin de s'assurer que l'échantillon demeure, dans la mesure du possible, statistiquement représentatif au cours du temps et afin de réduire au minimum l'erreur de réponse.

Contenu et collecte

Dans chaque ménage qui participe à l'ENSP, une personne bien informée fournit des renseignements généraux sur les caractéristiques démographiques et socioéconomiques, ainsi que sur la santé de chaque membre du ménage. Les données sur ces personnes

forment la *composante générale* (voir *Définitions*). En outre, on choisit au hasard dans chaque ménage une personne – pas nécessairement celle qui a donné les renseignements généraux – qui fournit des renseignements détaillés sur sa propre santé pour la *composante santé*. Si la personne sélectionnée a moins de 12 ans, on recueille les réponses *par personne interposée*. Seules les personnes sélectionnées au hasard dans les ménages qui participent à l'enquête sont suivies à intervalle de deux ans pour les besoins de l'étude longitudinale.

Le questionnaire comprend des sections réservées au *contenu de base* qui contiennent des questions sur l'état de santé, l'utilisation des services de santé, les facteurs de risque et les caractéristiques démographiques et socioéconomiques (tableau A en annexe). On évalue l'état de santé en posant des questions sur la perception de la santé, la capacité fonctionnelle, les problèmes de santé chroniques et la limitation des activités. On mesure l'utilisation de services de santé en posant des questions sur les visites chez les prestataires de soins de santé, sur les soins hospitaliers et sur la consommation de médicaments. Les facteurs de risque comportementaux incluent l'usage du tabac, la consommation d'alcool et l'activité physique. Les renseignements démographiques et socioéconomiques portent sur l'âge, le sexe, le niveau de scolarité, le groupe ethnique, le revenu et la situation professionnelle.

Lors de chaque cycle, on ajoute au questionnaire un *contenu thématique*, c'est-à-dire des questions sur un sujet particulier. Au premier cycle, ces questions avaient pour thème les facteurs psychosociaux susceptibles d'avoir des répercussions sur la santé, comme le stress, l'estime de soi et le contrôle. Au deuxième cycle, on a mis l'accent sur l'accès aux services de santé, tandis que le troisième cycle a pour thème l'initiative personnelle en matière de soins et les antécédents médicaux de la famille. En général, le contenu thématique ne fait l'objet que d'un seul cycle, mais peut éventuellement être répété lors de cycles ultérieurs.

La collecte des données, qui s'étale sur quatre trimestres, est effectuée par la méthode d'interview assistée par ordinateur (IAO). Toutefois, les données

Définitions

La *composante générale* du questionnaire de l'Enquête nationale sur la santé de la population (ENSP) vise à recueillir des renseignements sur les caractéristiques démographiques et socioéconomiques ainsi que des renseignements généraux sur la santé de chaque membre des ménages échantillonnés. Ces renseignements sont fournis par un membre bien informé du ménage.

La *composante santé* du questionnaire contient des questions détaillées sur la santé qui sont posées à une personne sélectionnée au hasard dans chaque ménage échantillonné. Les renseignements sont ordinairement fournis personnellement par la personne sélectionnée.

Les *réponses par personne interposée* sont celles obtenues au sujet d'un membre particulier d'un ménage auprès d'un autre membre de ce ménage; par exemple, un parent peut répondre au nom d'un jeune enfant. Pour la composante générale du questionnaire, une seule personne répond au nom de tous les membres du ménage (ce qui produit une combinaison de réponses directes et par personne interposée dans le cas des ménages comptant au moins deux personnes). Pour la composante santé, la personne sélectionnée au hasard répond en général personnellement aux questions qui la concernent (réponses directes). Cependant, pour les enfants ou dans des circonstances particulières, on accepte les réponses par personne interposée pour certaines sections de la composante santé du questionnaire. Par contre, elles ne sont pas permises pour d'autres, comme la composante sur la santé mentale.

Le contenu *de base* comprend les questions qui sont posées à chaque cycle de l'enquête. Le contenu *thématique* comprend les questions particulières qui sont posées à l'occasion d'un cycle particulier et que l'on intègre au questionnaire de l'ENSP. Le contenu *supplémentaire* correspond au contenu additionnel parrainé par un client externe à Statistique Canada. Les renseignements sur ce contenu supplémentaire sont recueillis au moyen d'une enquête distincte ou de questions intégrées au questionnaire de l'ENSP.

L'*échantillon longitudinal (principal)* comprend les 17 276 personnes sélectionnées au hasard à l'occasion du premier cycle (1994-1995) qui seront suivies (ou recontactées) tous les deux ans pendant une période allant jusqu'à 20 ans. Les personnes qui font partie de cet échantillon doivent avoir répondu complètement au moins aux questions de la composante générale du questionnaire en 1994-1995. Les personnes qui formaient l'échantillon supplémentaire de 1994-1995 sont exclues.

Les données longitudinales sont contenues dans trois fichiers de l'ENSP (tableau 2). Les données de chaque fichier sont pondérées

de sorte qu'elles soient représentatives de la population du Canada de 1994.

- Le *fichier longitudinal carré* contient les données sur les 17 276 personnes sélectionnées au hasard pour faire partie de l'échantillon longitudinal, indépendamment de leur situation au moment du cycle de 1996-1997.
- Le *fichier longitudinal complet* comprend les données sur les personnes pour lesquelles on possède des renseignements complets pour les cycles de 1994-1995 et 1996-1997. Ce fichier contient uniquement les données sur les 15 670 membres de l'échantillon longitudinal qui ont répondu aux composantes générale et santé du questionnaire en 1994-1995 et qui, en 1996-1997, ont répondu aux composantes générale et santé du questionnaire, avaient été placées en établissement et ont répondu au questionnaire de la composante sur les établissements de santé, ou étaient décédés.
- Le *fichier longitudinal partiel* comprend les 16 168 personnes pour lesquelles on possède au moins des renseignements partiels pour les cycles de 1994-1995 et 1996-1997, c'est-à-dire les 15 670 personnes dont les données sont contenues dans le fichier longitudinal complet, ainsi que les 498 personnes qui ont répondu aux composantes générale et santé du questionnaire en 1994-1995, mais uniquement à la composante générale en 1996-1997, ou qui ont répondu uniquement à la composante générale du questionnaire en 1994-1995 et qui, en 1996-1997, ont répondu au moins à la composante générale, avaient été placées en établissement ou étaient décédées.

L'*échantillon transversal* pour un cycle particulier de collecte comprend l'échantillon longitudinal, les échantillons supplémentaires et, le cas échéant, l'échantillon de remise à niveau.

Certaines provinces achètent des *échantillons supplémentaires (acquisition d'unités d'échantillonnage additionnelles)* pour les besoins de l'analyse transversale pour augmenter la taille de l'échantillon provincial durant un cycle de collecte particulier.

Un *échantillon de remise à niveau* peut être ajouté à un cycle particulier de collecte pour améliorer la représentativité transversale des membres du panel longitudinal. Sans cette remise à niveau, l'échantillon transversal de l'ENSP ne représenterait pas comme il convient certains segments de la population au cours du temps (enfants et immigrants, par exemple).

Pour les besoins du présent article, les *persévérants* sont les 15 670 personnes visées par le fichier longitudinal complet. Les *décrocheurs* sont les 1 606 membres de l'échantillon longitudinal qui ont fourni des renseignements incomplets en 1994-1995 ou 1996-1997 ou qui n'ont pas répondu en 1996-1997 (autrement dit, qui ont refusé de participer ou qui n'ont pu être retracés).

du premier cycle ont été recueillies principalement par interviews en personne dans les logements sélectionnés. Par contre, les données des cycles suivants ont été, ou seront, recueillies principalement par téléphone. Aucune étude n'a été faite quant aux effets éventuels de cette modification de la méthode de collecte. Toutes les données sont recueillies par autodéclaration (ou par déclaration par personne interposée). Aucune mesure physique n'est effectuée dans le cadre de l'ENSP.

Pour accroître la portée analytique des données, on étudie à l'heure actuelle la possibilité de coupler les dossiers de santé externes que possèdent les provinces avec les enregistrements des réponses aux questions de l'ENSP concernant l'utilisation des services de santé. Ce couplage n'est effectué qu'avec l'autorisation des personnes interrogées.

Cycle 2 – nouveau contenu

En 1996-1997, on a ajouté à la composante santé de l'ENSP une série de questions sur les blessures causées par des mouvements répétitifs et une autre sur la dépendance à l'égard de l'alcool. La première figurera dorénavant dans chaque cycle de l'enquête, tandis que la seconde ne sera utilisée que périodiquement. Comme le deuxième cycle avait pour thème l'accès aux services de santé, on a posé des questions sur l'utilisation des services, les raisons de cette utilisation, les obstacles rencontrés et les raisons invoquées pour ne pas avoir eu recours aux services de santé ou pour l'avoir fait peu souvent. Les services étudiés incluent la mesure de la tension artérielle, le test de Papanicolaou, la mammographie, l'examen des seins, l'examen général, la vaccination antigrippale, les visites chez le dentiste et l'examen de la vue. D'autres questions ont aussi été posées sur l'assurance complémentaire couvrant les soins dentaires, les lunettes ou les lentilles cornéennes, la chambre privée ou semi-privée à l'hôpital et les médicaments d'ordonnance. Enfin, l'utilisation des services d'urgence a fait l'objet d'une question.

Pour réduire les frais de collecte, on a intégré aux sections appropriées du questionnaire du deuxième cycle des questions sur la promotion de la santé financées par Santé Canada. (En 1994-1995, on s'était servi d'un questionnaire supplémentaire

distinct.) Les questions supplémentaires portaient sur la santé sexuelle, la sécurité routière, les mesures prises pour améliorer la santé et les opinions concernant l'usage du tabac et la consommation d'alcool.

Au deuxième cycle, un supplément sur l'asthme, sous forme d'enquête distincte financée par Santé Canada, a fourni des renseignements sur la gravité des crises, les facteurs de risque, les méthodes de contrôle et de traitement, le recours aux services médicaux et l'éducation concernant l'asthme.

Dans le cadre d'une entente avec le ministère de la Santé de l'Alberta, des questions supplémentaires ont été posées dans cette province sur les sources de renseignements sur la santé, l'exposition au soleil, le soutien social, les opinions concernant les parents, les services de santé, la santé sexuelle, la violence et la sécurité personnelle et le sens de la débrouillardise. Enfin, des questions sur la qualité des services de santé destinés aux enfants et sur l'accès à ces services ont été posées uniquement pour les échantillons supplémentaires en Alberta et au Manitoba.

Plan d'échantillonnage

Le plan d'échantillonnage stratifié à deux degrés (grappes, logements) de l'ENSP¹ est fondé sur celui de l'Enquête sur la population active de Statistique Canada, sauf pour le Québec, où il se fonde sur celui de l'Enquête sociale et de santé de Santé Québec. La taille de l'échantillon de base pour chaque province a été déterminée selon la méthode de répartition de Kish⁴, qui a permis d'établir un juste équilibre entre les exigences de fiabilité aux niveaux national et provincial. Au moins 1 200 ménages ont dû être sélectionnés dans chaque province pour pouvoir produire des données suffisamment fiables selon le sexe et les grands groupes d'âge. Les personnes qui vivent dans les réserves indiennes, les bases des Forces canadiennes et certaines régions éloignées du Québec et de l'Ontario sont exclues de la composante des ménages de l'enquête. Des composantes distinctes ont été mises au point pour la région du Nord et pour les établissements de soins de longue durée.

Les données sont pondérées pour tenir compte du plan d'échantillonnage, de la correction pour la

non-réponse et de la stratification a posteriori. Dans le cas des plans d'échantillonnage complexes comme celui de l'ENSP, les formules « classiques » qui servent habituellement à calculer la variance ne conviennent pas et on doit recourir à des méthodes plus perfectionnées^{5,6}. On a appliqué la méthode *jackknife* pour calculer la variance des données du premier cycle et la méthode *bootstrap*, celle des données du deuxième cycle. Ces méthodes tiennent compte de l'effet dû à la complexité du plan d'échantillonnage. Les deux méthodes prévoient la division de l'échantillon en sous-groupes (échantillons répétés ou *replicates*) et le calcul d'une estimation ponctuelle à partir d'un échantillon composé d'un sous-ensemble d'échantillons répétés. On peut alors déterminer la variation ou la variance des estimations d'après le nombre résultant préétabli d'échantillons. Les tables de variabilité d'échantillonnage approximative ont été préparées selon ces méthodes pour les données des deux cycles.

Maintien d'un échantillon équilibré

Comme un seul membre de chaque ménage échantillonné répond au questionnaire détaillé sur la santé et, s'il y a lieu, devient membre de l'échantillon longitudinal, la probabilité qu'une

personne soit incluse dans cet échantillon est inversement proportionnelle au nombre de personnes qui composent le ménage. Autrement dit, les personnes qui vivent dans un ménage de petite taille — généralement des célibataires ou des personnes âgées — ont plus de chance d'être sélectionnées que les membres des ménages de plus grande taille (généralement des parents et leurs enfants). On pourrait certes apporter une correction en appliquant des coefficients de pondération appropriés, mais la surreprésentation ou la sous-représentation initiale de l'échantillon longitudinal produirait, pour certains domaines analytiques importants, un déséquilibre quant aux niveaux prescrits de fiabilité, déséquilibre qui, sans correction, se perpétuerait pendant les 20 années d'existence du panel longitudinal.

Pour corriger ce déséquilibre éventuel, au premier cycle, on a appliqué une méthode d'exclusion en vue d'accroître la représentation des parents et des enfants dans l'échantillon. Conformément à cette méthode, on soumet une partie de l'échantillon des ménages, variant normalement de 19 % à 40 %, à un tri de sélection. Après avoir dressé la liste des membres de chaque ménage sélectionné, on exclut de l'enquête les ménages qui ne comptent aucun membre de moins de 25 ans⁷.

Cette stratégie donne d'assez bons résultats si on compare l'échantillon obtenu selon cette méthode d'exclusion aux résultats obtenus par simulation sans appliquer cette méthode (tableau 1). En effet, elle a permis d'obtenir une meilleure représentation des enfants (de moins de 12 ans) et des jeunes (de 12 à 24 ans) si l'on prend pour référence la répartition réelle de la population établie d'après les données du Recensement de 1991. Qui plus est, cette amélioration n'a pas eu lieu au détriment des parents (généralement de 25 à 44 ans), dont la représentation s'est améliorée aussi. Les personnes âgées (de 65 ans et plus) ont continué d'être surreprésentées, mais dans une moindre mesure.

Puisque cette nouvelle répartition de l'échantillon se rapproche davantage de la répartition réelle de la population, la variance a diminué pour les domaines des enfants et des jeunes et augmenté pour celui des personnes âgées. Cependant, globalement, la

Tableau 1
Répartition selon l'âge des membres sélectionnés au hasard dans l'échantillon de l'Enquête nationale sur la santé de la population, avec et sans méthode d'exclusion, Canada, territoires non compris, 1994-1995

Groupe d'âge	Membres de l'échantillon de l'ENSP		Recensement de 1991
	Avec méthode d'exclusion	Sans méthode d'exclusion	
			%
Total	100.0	100.0	100.0
0-11	11.9	9.9	16.7
12-24	16.4	13.7	18.2
25-44	33.0	32.2	34.2
45-64	22.1	24.4	20.0
65+	16.5	19.8	11.0

Sources des données : Recensement de 1991 ; Enquête nationale sur la santé de la population, 1994-1995

Nota : Les données ayant été arrondies, leur somme peut ne pas correspondre aux totaux.

- 275 qui étaient décédées au deuxième cycle mais pour lesquelles on dispose de renseignements complets obtenus au premier cycle;
- 498 pour lesquelles on possède des renseignements partiels (réponses aux deux cycles, mais ne portant que sur la composante générale pour l'une ou pour les deux enquêtes; certaines sont décédées ou ont été placées en établissement).

Les 1 108 autres personnes n'ont pas répondu à l'enquête de 1996-1997.

On a créé trois fichiers longitudinaux, à savoir le fichier *complet* (15 670 personnes = 15 334 + 61 + 275), le fichier *partiel* (16 168 personnes = 15 670 + 498) et le fichier *carré* (17 276 personnes = 16 168 + 1 108).

On a fait le suivi auprès des personnes satisfaisant aux critères d'inclusion dans l'échantillon longitudinal et qui ont été placées en établissement, qui ont déménagé dans les Territoires du Nord-Ouest, au Yukon, dans une réserve indienne ou dans une base des Forces canadiennes, ou qui ont quitté provisoirement le Canada entre les deux cycles de l'enquête.

En tout, 2 840 personnes interrogées à l'occasion du premier cycle à titre d'unités d'échantillonnage supplémentaires achetées par certaines provinces pour augmenter l'effectif de l'échantillon n'ont fait l'objet d'aucun suivi en 1996-1997.

On dispose de données transversales recueillies durant le premier cycle (1994-1995) pour 58 439 personnes dans le cas de la composante générale du questionnaire et pour 17 626 personnes (de 12 ans et plus) dans le cas de la composante détaillée sur la santé (tableau 3). Au deuxième cycle, en 1996-1997, l'achat d'*échantillons supplémentaires (acquisition d'unités d'échantillonnage additionnelles)* par trois provinces pour les besoins de l'analyse transversale a porté à 210 377 le nombre de personnes pour lesquelles des renseignements étaient disponibles pour la composante générale du questionnaire (173 216 de 12 ans et plus) et à 81 804 pour la composante détaillée sur la santé (73 402 de 12 ans et plus). Par conséquent, l'*échantillon transversal* du deuxième cycle comprend l'échantillon longitudinal ainsi que les échantillons supplémentaires achetés par certaines provinces. Il n'y a pas eu de *remise à niveau de l'échantillon* au deuxième cycle.

Les ministères de la Santé de l'Alberta, du Manitoba et de l'Ontario ont financé l'acquisition d'unités d'échantillonnage supplémentaires au deuxième cycle. Ces échantillons supplémentaires ont été sélectionnés de façon à fournir des estimations d'une fiabilité précise au niveau de la région sanitaire. Ils ont été sélectionnés selon la méthode de composition aléatoire (CA) et les entrevues ont été menées par téléphone. La

Tableau 3

Échantillons transversaux de l'Enquête nationale sur la santé de la population, Canada, territoires non compris, 1994-1995 et 1996-1997

	1994-1995			1996-1997		
	Total	Longitudinal (principal)	Supplémentaires (achetés)	Total	Longitudinal (principal)	Supplémentaires (achetés)
Composante générale						
Total	58 439	49 121	9 318	210 377	44 439	165 938
Enfants (moins de 12 ans)	11 477	9 616	1 861	37 161	8 419	28 742
Adultes (12 ans et plus)	46 962	39 505	7 457	173 216	36 020	137 196
Composante santé						
Total	17 626	14 786	2 840	81 804	15 681	66 123
Enfants (moins de 12 ans)	8 402	1 571	6 831
Adultes (12 ans et plus)	17 626	14 786	2 840	73 402	14 110	59 292

Source des données : Enquête nationale sur la santé de la population, 1994-1995 et 1996-1997

† Le Fichier santé transversal de l'ENSP de 1994-1995 ne comprend pas les données sur les enfants qui avaient moins de 12 ans et qui ont été interviewés dans le cadre de l'Enquête longitudinale nationale sur les enfants et les jeunes de 1994-1995.

... N'ayant pas lieu de figurer

stratification des échantillons sélectionnés par CA se fonde sur des groupes de circonscriptions téléphoniques. Dans les trois provinces concernées, pour chaque ménage échantillonné, on a recueilli les renseignements de la composante générale sur tous les membres et on a sélectionné au hasard une personne de 12 ans et plus pour répondre au questionnaire détaillé sur la santé. En Alberta et au Manitoba, dans la mesure du possible, on a également sélectionné un enfant de moins de 12 ans pour la composante santé.

Taux de réponse au deuxième cycle

Au deuxième cycle, le taux de réponse longitudinal a atteint 93,6 %. En ce qui concerne l'analyse transversale, pour la composante santé, le taux de réponse a été de 93,1 % pour les personnes faisant partie de l'échantillon longitudinal et de 75,8 % pour les personnes de 12 ans et plus faisant partie des échantillons supplémentaires sélectionnés par CA, ce qui correspond à un taux de réponse global de 79,0 %. Ces chiffres ne tiennent pas compte des enfants supplémentaires sélectionnés par CA en Alberta et au Manitoba. Pour les ménages sélectionnés par CA comptant un enfant de moins

de 12 ans) et dans lesquels un membre adulte avait déjà répondu au questionnaire, le taux de réponse est de 98,2 %.

On peut ventiler ces taux de réponse transversaux en taux de réponse des ménages et en taux de réponse des personnes sélectionnées dans l'échantillon longitudinal (tableau 4). Pour la partie longitudinale de l'échantillon, le taux de réponse est de 94,3 % pour les ménages et de 98,7 % pour les personnes sélectionnées au hasard. Pour la partie supplémentaire de l'échantillon sélectionnée par CA, le taux de réponse est de 80,0 % pour les ménages et de 94,8 % pour les personnes sélectionnées au hasard. Globalement, le taux de réponse des ménages est de 82,6 % et celui des personnes sélectionnées au hasard dans l'échantillon longitudinal, de 95,6 %.

Maintien du taux de réponse

Comme la participation des membres de l'échantillon longitudinal à l'enquête pourra durer jusqu'à 20 ans, il est indispensable de prendre des mesures pour soutenir leur intérêt et s'assurer de leur coopération. Aussi, l'équipe de l'ENSP a-t-elle élaboré plusieurs stratégies pour maintenir et améliorer les taux de réponse. De la documentation, comme des dépliants et des lettres, fait prendre conscience aux participants de l'importance et des avantages de l'enquête et répond aussi à certaines questions typiques, comme celles sur la nature confidentielle des renseignements personnels (voir *Confidentialité et diffusion*).

Étant donné l'engagement de longue durée que l'on attend de ces personnes, on a jugé qu'un premier contact en personne serait souhaitable. Par conséquent, en 1994-1995, l'entrevue du premier cycle s'est faite chez le participant. Les entrevues des cycles suivants ont été et seront généralement faites par téléphone, mais certaines seront faites sur place parce que quelques personnes interrogées n'ont pas le téléphone ou préfèrent cette forme d'entrevue (environ 5 % au cycle 2).

La mise à l'essai du questionnaire est essentielle pour n'importe quelle enquête, mais elle revêt une importance particulière dans le cas des enquêtes longitudinales. Ainsi, on recourt à des spécialistes

Tableau 4
Taux de réponse transversaux à l'Enquête nationale sur la santé de la population, Canada, territoires non compris, 1996-1997

Niveau de réponse	Échantillon longitudinal (principal)	Échantillon sélectionné par composition aléatoire (CA)	Total (principal + CA)
		%	
Taux de réponse global†	93,1	75,8	79,0
Ménages	94,3	80,0	82,6
Personnes sélectionnées au hasard (sauf les enfants sélectionnés par CA)	98,7	94,8	95,6

Source des données : Enquête nationale sur la santé de la population, 1996-1997

† On calcule le taux de réponse global pour les personnes sélectionnées au hasard dans l'ensemble des ménages en multipliant le taux de réponse des ménages qui ont participé à l'enquête par le taux de réponse des personnes sélectionnées appartenant à ces ménages.

Confidentialité et diffusion

L'Enquête nationale sur la santé de la population (ENSP) est réalisée aux termes de la *Loi sur la statistique*, qui garantit que les renseignements recueillis demeureront confidentiels. Tous les renseignements fournis à Statistique Canada, qu'ils proviennent d'une enquête, du Recensement ou de toute autre source, sont confidentiels. L'ENSP, tout comme d'autres enquêtes de Statistique Canada, doit relever le défi de diffuser autant de données que possible tout en respectant le caractère confidentiel des renseignements.

L'ENSP a recours à des fichiers maîtres qui contiennent toutes les données. Pour chaque cycle, le fichier maître comprend un Fichier général transversal et un Fichier santé transversal. Un fichier longitudinal a en outre été créé à l'issue du deuxième cycle. Des fichiers de microdonnées à grande diffusion (FMGD) sont préparés d'après ces fichiers pour être utilisés à l'extérieur de Statistique Canada sans qu'on puisse identifier les personnes qui ont participé à l'enquête. Ces fichiers doivent être approuvés par un comité interne de Statistique Canada avant leur diffusion. Des FMGD transversaux sur CD-ROM et sur disquette ont été diffusés pour les deux premiers cycles de l'enquête (1994-1995 et 1996-1997)^{8,9}. Ces FMGD sont en vente. Les chercheurs universitaires et les étudiants y ont accès grâce à l'Initiative de démocratisation des données (IDT) de Statistique Canada. Les FMGD sont également fournis aux bénéficiaires du Programme national de recherche et de développement en matière de santé mené conjointement par Santé Canada et Statistique Canada.

Plusieurs mesures ont été prises lors de la création du FMGD du cycle 1 afin de répondre aux normes de confidentialité. On a examiné les dénombrements univariés et les combinaisons de variables au niveau de détail géographique le moins agrégé dans chaque province pour repérer les enregistrements solitaires dans l'échantillon. En outre, on a examiné la répartition des coefficients de pondération pour les provinces qui ont acheté des échantillons supplémentaires, afin de s'assurer qu'il soit impossible d'identifier des strates particulières (donc, des données à un niveau assez peu agrégé de détail géographique). On a examiné le Fichier général pour voir s'il était possible de recréer des ménages d'après les données. Ces procédures ont mené à la suppression ou au regroupement de variables pour créer le FMGD transversal approuvé de l'ENSP. Dans certains cas, on a introduit dans le FMGD des variables dérivées au lieu des réponses particulières à des questions individuelles. L'Indice de l'état de santé qui remplace les 31 questions distinctes d'après lesquelles il est calculé en est un exemple.

Une méthode plus complexe a présidé au respect de la confidentialité des renseignements personnels dans le FMGD transversal de 1996-1997. Cette situation tenait principalement au fait que le FMGD de 1994-1995 avait déjà été diffusé et que les échantillons transversaux des deux cycles se chevauchaient, puisque les membres du panel longitudinal continuent d'être observés. En plus des méthodes utilisées pour les données de 1994-1995, on a comparé les facteurs de pondération appliqués aux données des deux cycles pour déterminer s'il était possible d'identifier des personnes ou des ménages. En outre, on a procédé à une étude plus détaillée des tables à trois dimensions utilisées pour examiner 13 variables clés au niveau de l'unité géographique la plus petite pour laquelle des données seraient diffusées pour chaque province afin de repérer les combinaisons solitaires de variables associées

à de fortes proportions. Qui plus est, pour empêcher les utilisateurs de créer indépendamment des enregistrements longitudinaux à partir des deux FMGD transversaux, on a apparié les enregistrements du FMGD de 1994-1995 à ceux proposés pour le fichier de 1996-1997 en se fondant sur 12 variables d'appariement, pour déterminer les proportions de paires réelles (les mêmes personnes) et de faux appariements (personnes différentes) par rapport au nombre total d'appariements. Les enregistrements qui ont donné lieu à des paires réelles ont été examinés pour voir si des combinaisons particulières de variables permettraient d'identifier les personnes qui ont fourni les renseignements. Enfin, on a examiné les nouvelles variables introduites dans l'enquête en 1996-1997, pour déterminer si le risque d'enfreindre la confidentialité des renseignements avait augmenté depuis le premier cycle. Comme en 1994-1995, ces procédures ont mené à la suppression ou au regroupement de certaines variables.

Les préoccupations quant au respect de la nature confidentielle des renseignements personnels touchent aussi les outils d'estimation de la variance mis à la disposition des utilisateurs. Puisque les renseignements détaillés sur le plan de sondage qu'exigent les méthodes d'estimation de la variance permettent d'identifier des régions géographiques à un niveau de détail très poussé, d'autres méthodes sont à l'étude. Cependant, il faut aussi s'assurer que les coefficients de pondération eux-mêmes ne permettent pas aux utilisateurs d'identifier de petites régions géographiques ou d'apparier des FMGD provenant de cycles différents.

La nature longitudinale de l'enquête, ainsi que la quantité de données recueillies, augmenteront la probabilité d'identifier des enregistrements solitaires, donc, des personnes particulières, à mesure que les données s'accumuleront. Étant donné le travail que demande la création du FMGD transversal et l'incertitude actuellement associée à la création d'un FMGD longitudinal après deux cycles de collecte de données seulement, il est probable qu'après le troisième cycle, la vérification du respect de la nature confidentielle des données transversales soit encore plus complexe et qu'il soit impossible de préparer un FMGD longitudinal.

Puisque le contenu des FMGD est limité et que les utilisateurs pourraient ne pas avoir les moyens de payer pour la production de données personnalisées compte tenu de leurs budgets de recherche restreints, il faut trouver un moyen d'offrir un accès plus général aux données. À cet égard, un service conçu pour offrir l'accès à distance aux fichiers maîtres a donné d'assez bons résultats. Statistique Canada fournit aux utilisateurs autorisés des faux fichiers de l'ENSP, c'est-à-dire des fichiers qui ont la même structure que les fichiers maîtres, mais qui contiennent des données fictives. Les utilisateurs préparent leurs propres programmes informatiques, les mettent à l'essai en se servant des faux fichiers, puis les transmettent à Statistique Canada par courrier électronique. Les membres de l'équipe de l'ENSP exécutent les programmes sur les fichiers maîtres à l'interne, vérifient les données de sortie pour s'assurer du respect des normes de confidentialité, puis les envoient aux utilisateurs par courrier électronique. En ce qui concerne l'estimation de la variance, les utilisateurs peuvent recourir au service d'accès à distance pour soumettre des programmes fondés sur les poids afin qu'on les applique aux fichiers maîtres. D'autres moyens de permettre aux chercheurs d'utiliser les données de l'ENSP tout en veillant au respect des normes de confidentialité sont à l'étude (par exemple, dans certaines conditions, des chercheurs pourraient consulter les fichiers maîtres de l'ENSP aux bureaux régionaux ou au bureau central de Statistique Canada).

de la santé et à des groupes de discussion en profondeur pour élaborer et perfectionner les questions de l'ENSP, et on met le questionnaire à l'essai deux fois avant chaque cycle. Ces essais permettent de déterminer si les personnes interrogées comprennent les questions qu'on leur pose. Les intervieweurs font par la suite des commentaires dont on tient compte lors de la rédaction finale du questionnaire. Leurs commentaires guident en outre la présentation des pages écrans de l'application d'interview assistée par ordinateur (IAO) qui fait, elle aussi, l'objet de nombreuses mises à l'essai.

L'application informatique à l'origine du questionnaire permet de personnaliser les questions selon la personne interrogée et de faire ainsi état des contacts antérieurs. Dans le cas des membres de l'échantillon longitudinal, on intègre aux questions le nom de la personne, ainsi que les pronoms personnels et possessifs propres à son sexe. Les données recueillies antérieurement sont également intégrées pour permettre le suivi. Par exemple, si une personne a déjà fourni les renseignements sur son niveau de scolarité à l'occasion d'un cycle antérieur, on ne demande pas ce renseignement une deuxième fois. On ne pose à cette personne que des questions au sujet des études qu'elle aurait faites depuis la dernière entrevue, en les adaptant au niveau de scolarité mentionné auparavant.

Le dépistage est essentiel dans le cas des enquêtes longitudinales. Les efforts en vue de repérer les participants à l'ENSP qui ont déménagé entre deux cycles sans en avertir Statistique Canada portent généralement fruit. Seulement 1,7 % de ces personnes n'ont pu être retracées pour le deuxième cycle.

On prend contact avec les membres de l'échantillon longitudinal qui ont été placés en établissement de soins de longue durée ou qui ont déménagé dans le Nord du Canada et on inclut les données qui les concernent dans le fichier longitudinal. Inversement, on suit ceux qui ont participé à la composante des établissements de santé, qui sont sortis de ces établissements ou ceux qui ont participé à la composante de la région du

Nord et qui ont déménagé dans l'une des 10 provinces. On classe les données ainsi recueillies dans les fichiers longitudinaux respectifs.

Un « cinquième trimestre » de collecte de données a été ajouté après le premier cycle uniquement pour assurer le suivi des personnes qui n'avaient pas répondu à l'enquête.

Correction pour la non-réponse et stratification a posteriori

Malgré les efforts déployés pour la contrer, une certaine part de non-réponse reste inévitable. Ses effets sur les données d'enquête peuvent être doubles : en premier lieu, elle diminue la taille effective de l'échantillon et, par conséquent, augmente la variance. En deuxième lieu, elle peut mener à des estimations biaisées, si les personnes qui n'ont pas répondu diffèrent de celles qui ont répondu en ce qui a trait aux caractéristiques évaluées^{10,11}. En fait, au deuxième cycle, la répartition des *persévérants* et des *décrocheurs* dans l'échantillon longitudinal diffère considérablement selon le sexe, le groupe d'âge, la situation professionnelle, le revenu du ménage et la présence d'au moins un problème de santé chronique (tableau B en annexe). Le taux de décrochage est plus élevé pour les hommes que pour les femmes : 53,3 % contre 46,7 %. La proportion de personnes qui ont cessé de participer à l'enquête est plus forte pour les groupes d'âge moyen (18 à 24 ans, 25 à 44 ans et 45 à 64 ans) que pour les groupes d'âge moins avancé et plus avancé. Les poids d'échantillonnage ont en outre été corrigés en vue de tenir compte de l'effet de la non-réponse^{12,13}.

En guise d'étape finale de la pondération longitudinale, on a procédé à des rajustements selon la province, le groupe d'âge (0 à 11 ans, 12 à 24 ans, 25 à 44 ans, 45 à 64 ans, 65 ans et plus) et le sexe, afin que l'échantillon pondéré soit représentatif de la population estimative de 1994 (stratification a posteriori).

On s'est appuyé sur les données du premier cycle pour corriger les données pour la non-réponse. On s'est servi de l'algorithme CHAID (Chi-square Automatic Interaction Detection) offert par le logiciel *Knowledge Seeker* pour créer des catégories de

pondération pour la correction de la non-réponse fondées sur des variables considérées comme étant de bons prédicteurs de la non-réponse¹². On a formé de la sorte des catégories de pondération pour les caractéristiques pour lesquelles la différence entre les personnes qui ont répondu à l'enquête et celles qui n'y ont pas répondu était la plus marquée et pour lesquelles la taille de l'échantillon était suffisante pour exécuter le test de signification du chi carré. Les variables particulières incluent le revenu, l'âge, le sexe, la race, le lieu de naissance, le fait d'être propriétaire/locataire du logement, la présence d'enfants/de jeunes dans le ménage, la taille du ménage et plusieurs variables géographiques, comme la province et la région urbaine/rurale. Les variables utilisées pour créer les catégories pour la correction de la non-réponse diffèrent d'une province à l'autre. Les corrections pour tenir compte de la non-réponse au deuxième cycle de l'enquête réussissent, dans la plupart des cas, à réduire le biais parmi les membres persévérants de l'échantillon en regard d'un sous-ensemble de variables de l'ENSP (tableau C en annexe). Les variables utilisées pour la correction pour la non-réponse, ou les variables qui leur sont corrélées, devraient, en principe, être celles qui produisent la réduction la plus importante du biais chez les membres persévérants de l'échantillon.

Incohérence des déclarations au fil du temps

L'intégration de données antérieures au stade de la conception du questionnaire peut créer des difficultés d'un genre particulier. Plus précisément, on peut observer des incohérences après qu'une même personne ait répondu aux questions à plusieurs reprises et les réponses par personne interposée ne font qu'accentuer le phénomène.

Pour réduire au minimum ce genre d'incohérences, on a intégré aux questions du deuxième cycle des questions d'approfondissement fondées sur des données du premier cycle. Par exemple, aux personnes qui ont fait état d'un changement ayant trait à un problème de santé chronique qui normalement n'évolue pas, on a posé des questions pour confirmer le changement, pour en connaître les raisons et pour enregistrer les dates

pertinentes. Cependant, on a omis de demander des précisions dans les cas où une telle demande risquait d'enfreindre les règles de respect de la nature confidentielle des renseignements. Par exemple, on ne pose aucune question d'approfondissement si un problème de santé chronique particulier (comme le diabète ou l'épilepsie) a été enregistré précédemment, mais qu'une personne répondant par procuration déclare plus tard que la personne concernée ne souffre pas de ce problème. Pareillement, on n'a pas posé de questions pour vérifier les réponses par personne interposée au sujet de l'usage du tabac. Enfin, bien que des questions d'approfondissement permettent de valider les changements de valeur de variables comme la consommation d'alcool ou le poids, on n'en a pas posé étant donné leur nature délicate.

On s'est servi des données obtenues après avoir posé des questions d'approfondissement dans le fichier longitudinal de l'année courante, mais on n'a pas corrigé les données recueillies antérieurement. On a maintenu les incohérences dans le fichier afin que les analystes puissent les traiter de la façon qu'ils jugent la plus appropriée. Cependant, on a retenu et utilisé les données du deuxième cycle sur la date de naissance et le sexe quand la confirmation au deuxième cycle des renseignements recueillis au premier a révélé des incohérences.

Déclaration par personne interposée

On a posé des questions sur les mesures de prévention et sur l'usage du tabac dans le cadre de la composante santé du questionnaire. Le taux de réponse à ces questions par personne interposée pour l'échantillon longitudinal de personnes de 12 ans et plus est très faible (1,8 %), car on a fortement déconseillé cette forme de déclaration pour cette partie du questionnaire. Pour d'autres sections (dont celles sur la dépression et l'estime de soi), la déclaration par personne interposée n'était pas autorisée. En revanche, les questions sur les problèmes de santé chroniques et la limitation des activités ont été posées dans le cadre de la composante générale du questionnaire, où on acceptait des réponses par personne interposée au nom des membres du panel longitudinal.

Le taux de réponse par personne interposée a diminué au deuxième cycle, car les intervieweurs ont essayé de communiquer directement avec les membres du panel longitudinal pour leur poser les questions des composantes générale et santé. Néanmoins, certaines incohérences donnent à penser que la déclaration par personne interposée pourrait poser des problèmes. Les comparaisons préliminaires d'estimations sur la prévalence des problèmes de santé chroniques calculées d'après les données du premier et du deuxième cycles en se fondant sur la composante générale ainsi que sur la composante santé du questionnaire indiquent que la déclaration par personne interposée influencent les résultats transversaux ainsi que longitudinaux. (Les données sur les membres de moins de 12 ans de l'échantillon longitudinal sont entièrement fondées sur la réponse par personne interposée et ont donc été exclues de ces analyses.)

Étant donné ces résultats préliminaires, la solution à court terme choisie pour le troisième cycle a consisté à améliorer le guide de l'intervieweur et le manuel des procédures afin de décourager la réponse par personne interposée aux questions de la composante générale dans le cas de l'échantillon longitudinal.

Modifications des conventions de codage

Une autre difficulté que posent les données longitudinales survient quand les conventions de codage changent au cours du temps, par exemple, à la suite de modifications de la classification des maladies, des médicaments, des professions ou des industries, ou encore de la classification géographique. La stratégie générale adoptée pour l'ENSP consiste à s'appuyer sur le même système de classification pour tous les cycles de l'enquête. Autrement dit, si on adopte un nouveau système, les données chronologiques contenues dans le fichier longitudinal sont recodées. Suivant le temps, l'espace et le budget disponibles, on pourrait maintenir plus d'un système de classification si l'analyse le justifiait.

Si le recodage n'a pas lieu, les différences entre les systèmes de classification sont mentionnées dans la documentation sur l'enquête.

Conclusion

Les questions soulevées dans le présent article, ainsi que la foule d'autres problèmes qui se manifesteront sans aucun doute à l'occasion des prochains cycles de l'Enquête nationale sur la santé de la population, sont inhérents à la nature longitudinale de la collecte des données. Ce genre d'enquête longitudinale pose un défi aussi bien aux chercheurs qui analysent les données qu'aux concepteurs d'enquête qui élaborent le plan de sondage et recueillent, traitent et diffusent l'information. ●

Remerciements

Les auteurs remercient Yves Béland, Jiajian Chen, Leslie Geran, Jennifer Hubbard, Bryan LaFrance, Jackey Mayda, Georgia Roberts, Karen Roberts, Eric Sayre, Margot Shields, Kathryn Wilkins et Douglas Yeo pour leur collaboration.

Références

1. J.-L. Tambay et G. Catlin, « Plan d'échantillonnage de l'Enquête nationale sur la santé de la population », *Rapports sur la santé*, 7(1), 1995, p. 31-42 (Statistique Canada, n° 82-003 au catalogue).
2. Statistique Canada, « Enquête nationale sur la santé de la population - Cycle 2, 1996-1997 », *Le Quotidien* (n° 11-001 au catalogue), 29 mai 1998, p. 3-5.
3. Statistique Canada, *Aperçu de l'Enquête nationale sur la santé de la population, 1996-1997* (n° 82-567 au catalogue), Ottawa, Ministère de l'Industrie, 1998.
4. L. Kish, « Plans de sondage à usages multiples », *Techniques d'enquête*, 14(1), 1988, p. 19-33 (Statistique Canada, n° 12-001 au catalogue).
5. J.N.K. Rao, C.F.J. Wu et K. Yue, « Quelques travaux récents sur les méthodes de rééchantillonnage applicables aux enquêtes complexes », *Techniques d'enquête*, 18(2), 1992, p. 209-217 (Statistique Canada, n° 12-001 au catalogue).
6. K.F. Rust et J.N.K. Rao, « Variance estimation for complex surveys using replication techniques », *Statistical Methods in Medical Research*, 5, 1996, p. 283-310.

7. J.-L. Tambay et C. Mohl, *Improving sample representativity through the use of a rejective method*, American Statistical Association, Proceedings of the Survey Research Methods Section, 1995, p. 29-38.
8. Statistique Canada, *L'enquête nationale sur la santé de la population, 1994-1995*, Fichiers de microdonnées à grande diffusion (n° 82F0001 au catalogue), Ottawa, Ministre de l'Industrie, 1995.
9. Statistique Canada, *L'enquête nationale sur la santé de la population, 1994-1995*, Fichiers de microdonnées à grande diffusion (documentation, nos 82M0009XCB, CD-ROM et 82M0009GPF au catalogue), Ottawa, Ministre de l'Industrie, 1995.
10. L. Swain et D. Dolson, « Current issues in household survey nonresponse at Statistics Canada », *Statistics in Transition*, 3(3), 1998, p. 439-467.
11. L. Swain et D. Dolson, « Current issues in household survey nonresponse at Statistics Canada », Nonresponse in Survey Research: Proceedings of the Eighth International Workshop on Household Survey Nonresponse (1997), 1998, p. 1-21. (Nota : Ceci est une version abrégée de la référence n° 10.)
12. J.-L. Tambay, I. Schiopu-Kratina, J.E. Mayda *et al.*, « Traitement de la non-réponse du cycle deux de l'enquête nationale sur la santé de la population », *Techniques d'enquête*, 24(2), 1998, p. 159-169 (Statistique Canada, n° 12-001 au catalogue).
13. D.M. Stukel, C.A. Mohl et J.-L. Tambay, *Weighting for cycle 2 of Statistics Canada's National Population Health Survey*, Statistical Society of Canada, Proceedings of the Survey Methods Section, 1997, p. 111-116.
14. J.N.K. Rao et D.R. Thomas, « Chi-squared tests for contingency tables », dans *Analysis of Complex Surveys*, C.J. Skinner, D. Holt et T.M.F. Smith (s. la dir. de), New York, Wiley, 1989, p. 89-114.

Annexe

Tableau A

Résumé du contenu, Enquête nationale sur la santé de la population, 1994-1995 et 1996-1997

	Cycle 1 (1994-1995)	Cycle 2 (1996-1997)		Cycle 1 (1994-1995)	Cycle 2 (1996-1997)
Contenu de base			Contenu supplémentaire (acheté) intégré au questionnaire de l'ENSP		
Incapacité au cours des deux dernières semaines	O	O			
Utilisation des services de santé	O	O	Enquête promotion santé		
Limitation des activités	O	O	Régime/nutrition	N	O
Problèmes de santé chroniques	O	O	Taille et poids	N	O
Caractéristiques sociodémographiques	O	O	Auto-examen des seins	N	O
Niveau de scolarité	O	O	Allaitement	N	O
Population active	O	O	Grossesse	N	O
Revenu	O	O	VIH	N	O
État de santé perçu par la personne	O	O	Usage du tabac	N	O
Santé de la femme	O	O	Consommation d'alcool	N	O
Tension artérielle	O	O	Santé sexuelle	N	O
Taille et poids	O	O	Sécurité routière	N	O
État de santé	O	O	Insécurité alimentaire	N	O
Activités physiques	O	O			
Blessures résultant de mouvements répétitifs	N	O	Questions supplémentaires achetées par certaines provinces		
Blessures	O	O	Sens de la débrouillardise (Alberta)	O	O
Consommation de médicaments	O	O	Sens de la débrouillardise (Manitoba)	O	N
Usage du tabac	O	O	Renseignements sur la santé	N	O
Consommation d'alcool	O	O	Bronzage et expositions aux rayons UV	N	O
Santé mentale	O	O	Soutien social	N	O
Soutien social	O	O	Opinions au sujet des parents	N	O
Sentiment de cohésion	O	N	Services de santé	N	O
Dépendance à l'égard de l'alcool	N	O	Santé sexuelle	N	O
			Sécurité personnelle et violence	N	O
			Services de santé pour enfants	N	O
Contenu thématique					
Psychosocial			Contenu supplémentaire (acheté), enquêtes distinctes		
Stress	O	N	Enquête promotion santé	O	N
Problèmes courants	O	N	Enquête sur l'asthme	N	O
Événements récents	O	N			
Facteurs de stress chez les enfants et les adultes ("traumatismes")	O	N			
Stress au travail	O	N			
Estime de soi	O	N			
Contrôle	O	N			
Accès aux services de santé					
Tension artérielle	N	O			
Test de Papanicolaou	N	O			
Mammographie	N	O			
Examen des seins	N	O			
Auto-examen des seins	N	O			
Allaitement	N	O			
Examen général	N	O			
Vaccination antigrippale	N	O			
Visites chez le dentiste	N	O			
Examen des yeux	N	O			
Services d'urgence	N	O			
Assurance	N	O			

Tableau B

Caractéristiques de référence (cycle 1) des persistants et des décrocheurs, échantillon longitudinal, Enquête nationale sur la santé de la population, 1994-1995

Caractéristiques personnelles	Persévérants		Décrocheurs		Chi carré
	Taille de l'échantillon	%	Taille de l'échantillon	%	
Sexe					
Masculin	7 209	49,1	835	53,3	5,3*
Féminin	8 461	50,9	771	46,7	
Groupe d'âge					
0-11	1 908	16,9	114	11,5	30,6***
12-17	1 047	9,0	90	8,1	
18-24	1 434	8,6	228	12,5	
25-44	5 226	33,3	602	36,1	
45-64	3 509	20,7	378	22,3	
65+	2 546	11,6	194	9,5	
État matrimonial					
Marié(e)	7 416	49,2	806	50,9	ns
Célibataire	5 638	40,5	587	40,0	
Marié(e) antérieurement	2 614	10,3	211	9,1	
Situation professionnelle					
Travaille à l'heure actuelle	7 288	47,0	810	49,0	18,3***
A travaillé au cours des 12 derniers mois, mais ne travaille pas à l'heure actuelle	1 092	6,2	135	7,6	
N'a pas travaillé au cours des 12 derniers mois	4 741	25,2	481	28,1	
Sans objet	2 415	21,7	148	15,3	
Revenu du ménage					
Inférieur	1 116	5,3	162	8,4	21,1***
Moyen-inférieur	2 240	12,5	243	15,6	
Moyen	4 661	30,5	457	30,4	
Moyen-supérieur	5 192	36,0	439	31,7	
Supérieur	1 843	15,8	158	14,0	
Niveau de scolarité					
Pas de diplôme d'études secondaires	4 744	33,2	518	34,2	ns
Diplôme d'études secondaires	1 963	14,8	221	15,2	
Certaines études postsecondaires	3 214	23,3	331	20,7	
Diplôme collégial ou universitaire	3 819	28,7	405	29,8	
Vit seul(e)					
Non	12 749	90,7	1 336	91,6	ns
Oui	2 921	9,3	270	8,4	
État de santé autodéclaré					
Mauvais	386	1,8	38	2,9	ns
Passable	1 379	7,0	105	7,6	
Bon	3 834	23,7	302	25,6	
Très bon	5 712	36,0	399	36,5	
Excellent	4 359	31,4	280	27,3	
Au moins un problème de santé chronique					
Non	7 182	50,1	811	54,2	5,7*
Oui	8 471	49,9	775	45,8	
A fait une dépression au cours des 12 derniers mois					
Non	12 290	94,5	885	93,0	ns
Oui	781	5,5	68	7,0	

Source des données : Enquête nationale sur la santé de la population, fichier longitudinal carré, 1994-1995 et 1996-1997

Nota : Les répartitions n'incluent pas la catégorie « données manquantes ». Sauf dans le cas du revenu du ménage, de l'état de santé tel que déclaré par la personne interrogée et de la dépression, variables pour lesquelles les proportions de données manquantes étaient respectivement de 5 %, 3 %, et 9 %, la proportion de valeurs manquantes dans le fichier longitudinal carré est inférieure à 1 %. Les proportions sont calculés d'après des données pondérées. Le test du chi carré porte sur 500 coefficients de pondération bootstrap et inclut la correction de deuxième ordre de Rao-Scott¹⁴ tenant compte du plan de sondage complexe.

* $p \leq 0.05$

** $p \leq 0.01$

*** $p \leq 0.001$

Tableau C

Caractéristiques de référence de l'ensemble des membres de l'échantillon longitudinal du cycle 1 de l'Enquête nationale sur la santé de la population (1994-1995) et des persévérants du cycle 2 (1996-1997), après et avant correction des coefficients de pondération pour tenir compte de la non-réponse au cycle 2

Caractéristiques personnelles	Tous les membres de l'échantillon longitudinal du cycle 1 (1994-1995)		Cycle 2 (1996-1997) persévérants uniquement		
			Après correction pour tenir compte de la non-réponse		Avant correction pour tenir compte de la non-réponse
	Taille de l'échantillon	%	Taille de l'échantillon	%	%
Sexe					
Masculin	8 044	49,5	7 209	49,5	49,1
Féminin	9 232	50,5	8 461	50,5	50,9
Groupe d'âge					
0-11	2 022	16,3	1 908	16,3	16,9
12-17	1 137	8,9	1 047	9,0	9,0
18-24	1 662	9,0	1 434	8,9	8,6
25-44	5 828	33,6	5 226	33,6	33,3
45-64	3 887	20,8	3 509	20,8	20,7
65+	2 740	11,4	2 546	11,4	11,6
État matrimonial					
Marié(e)	8 222	49,4	7 416	49,4	49,2
Célibataire	6 225	40,5	5 638	40,4	40,5
Marié(e) antérieurement	2 825	10,1	2 614	10,1	10,3
Situation professionnelle					
Travaille à l'heure actuelle	8 098	47,2	7 288	47,6	47,0
A travaillé au cours des 12 derniers mois, mais ne travaille pas à l'heure actuelle	1 227	6,3	1 092	6,3	6,2
N'a pas travaillé au cours des 12 derniers mois	5 222	25,5	4 741	25,1	25,2
Sans objet	2 563	21,0	2 415	21,1	21,7
Revenu du ménage					
Inférieur	1 278	5,6	1 116	5,4	5,3
Moyen-inférieur	2 483	12,8	2 240	12,5	12,5
Moyen	5 118	30,5	4 661	30,3	30,5
Moyen-supérieur	5 631	35,6	5 192	35,9	36,0
Supérieur	2 001	15,6	1 843	15,9	15,8
Niveau de scolarité					
Pas de diplôme d'études secondaires	5 262	33,3	4 744	33,0	33,2
Diplôme d'études secondaires	2 184	14,8	1 963	14,8	14,8
Certaines études postsecondaires	3 545	23,0	3 214	23,4	23,3
Diplôme collégial ou universitaire	4 224	28,8	3 819	28,8	28,7
Vit seul(e)					
Non	14 085	90,8	12 749	90,7	90,7
Oui	3 191	9,2	2 921	9,3	9,3
État de santé autodéclaré					
Mauvais	424	1,9	386	1,8	1,8
Passable	1 484	7,0	1 379	7,0	7,0
Bon	4 136	23,8	3 834	23,8	23,7
Très bon	6 111	36,1	5 712	36,1	36,0
Excellent	4 639	31,1	4 359	31,3	31,4
Au moins un problème de santé chronique					
Non	7 993	50,6	7 182	50,2	50,1
Oui	9 246	49,4	8 471	49,8	49,9
A fait une dépression au cours des 12 derniers mois					
Non	13 175	94,4	12 290	94,5	94,5
Oui	849	5,6	781	5,5	5,5

Source des données : Enquête nationale sur la santé de la population, fichiers longitudinaux carré et complet, 1994-1995 et 1996-1997

Nota : Les répartitions n'incluent pas la catégorie « données manquantes ». Sauf dans le cas du revenu du ménage et de la dépression, variables pour lesquelles la proportion de données manquantes est de 5 %, la proportion de valeurs manquantes dans le fichier longitudinal complet est inférieure à 1 %. Les proportions sont calculées d'après des données pondérées. Tous les coefficients de pondération incluent l'étape de la stratification a posteriori.