

Plan d'échantillonnage de l'Enquête nationale sur la santé de la population

Jean-Louis Tambay et Gary Catlin*

Résumé

En 1994, Statistique Canada a commencé la collecte de données pour l'Enquête nationale sur la santé de la population (ENSP). Cette enquête-ménage a pour objet de mesurer l'état de santé des Canadiens et de favoriser une meilleure compréhension des facteurs déterminants pour la santé. Il s'agit d'une enquête longitudinale pour laquelle on recueille, tous les deux ans, des données sur des personnes choisis dans le panel. Le présent article décrit le plan d'échantillonnage de l'ENSP, ainsi que les motifs qui en ont déterminé l'élaboration. Il examine la répartition et la sélection représentative; les modifications adoptées pour le Québec et pour les territoires; de même que l'intégration de l'ENSP et de l'Enquête longitudinale nationale sur les enfants (ELNE). La section finale expose certaines questions méthodologiques que l'on devra traiter pour faciliter la conduite de cette enquête à l'avenir.

Mots-clés: collecte de données, enquêtes sur la santé, taille des échantillons

Introduction

L'Enquête nationale sur la santé de la population (ENSP) sert à recueillir des renseignements sur la santé de la population canadienne¹. La première collecte de données sur 12 mois a commencé en 1994; d'autres collectes auront lieu tous les deux ans. En plus de données transversales, l'enquête permettra de recueillir des données longitudinales auprès de personnes choisies dans le panel. Les rapports sur le premier cycle de collecte devraient être diffusés plus tard cette année. Le présent document décrit le plan d'échantillonnage et fournit des renseignements de base sur l'ENSP.

* Jean-Louis Tambay (613-951-6959) travaille à la Division des méthodes d'enquêtes ménages et Gary Catlin (613-951-3830) travaille à la Division des statistiques sur la santé à Statistique Canada, Ottawa, K1A 0T6.

Objectifs et contenu

Les objectifs généraux de l'ENSP sont les suivants:

- C faciliter l'élaboration des politiques gouvernementales en fournissant des données sur l'état de santé de la population;
- C fournir des données qui aideront à comprendre les facteurs déterminants pour la santé;
- C recueillir des données sur les corrélations entre la santé et les facteurs économiques, sociaux, démographiques, professionnels, et environnementaux;
- C éclaircir la relation entre l'état de santé et l'utilisation des services de santé, y compris les services spéciaux et les services traditionnels;
- C suivre un panel de personnes dans le temps pour refléter la dynamique de la santé et de la maladie;
- C donner aux provinces, aux territoires et à d'autres clients une capacité d'enquête sur la santé en vue de compléter le contenu et/ou l'échantillonnage;
- C permettre de faire des liens entre les données de l'enquête et des données administratives courantes, notamment les statistiques de l'état civil, les mesures environnementales, les variables liées aux collectivités, ainsi que l'utilisation des services de santé.

Le contenu de l'enquête est déterminé d'après les critères suivants:

- C Les renseignements doivent être pertinents par rapport aux objectifs des provinces et des territoires en matière de santé, et faciliter le contrôle de ces objectifs. S'il y a des objectifs généraux, par exemple, au niveau national, on peut déterminer le contenu de l'enquête à partir des politiques et des programmes.
- C Les renseignements ne doivent pas faire double emploi avec d'autres sources.
- C Pour favoriser une meilleure compréhension de la

santé et des facteurs déterminants à cet égard, l'enquête doit fournir des renseignements dans des domaines qui n'ont pas encore été étudiés de manière approfondie.

- C L'enquête doit être axée sur des comportements ou des conditions pouvant faire l'objet de prévention, de traitements ou d'interventions.
- C L'enquête doit recueillir des renseignements sur les conditions qui causent le plus de souffrance et/ou qui coûtent le plus aux personnes, au grand public ou au système de santé.
- C L'enquête doit recueillir des renseignements sur les facteurs liés à une bonne santé, et non pas seulement aux maladies.

Conformément à ces lignes directrices, le questionnaire inclut des questions sur l'état de santé, sur l'utilisation des services de santé, sur les facteurs de risque, ainsi que les caractéristiques socio-économique et démographique. Par exemple, on y mesure l'état de santé par des questions touchant la perception de soi du point de vue de la santé, de la capacité fonctionnelle, des conditions chroniques et des limitations d'activités. Le recours aux services de santé est mesuré par des questions concernant les visites faites à des dispensateurs de soins de santé, les soins hospitaliers, et l'utilisation de médicaments. Les facteurs de risque comportementaux incluent le tabagisme, la consommation d'alcool, et l'absence d'activité physique. De plus, la première enquête accorde une attention particulière aux facteurs psychosociaux susceptibles de se répercuter sur la santé, notamment le stress, l'estime de soi, et le soutien social. Les renseignements démographiques et socio-économiques incluent l'âge, le sexe, l'éducation, l'origine ethnique, le revenu du ménage, et la situation professionnelle.

La collecte de données a commencé en 1994 et se fera tous les deux ans. Le premier contact avec les ménages échantillonnés s'effectue en personne; l'entrevue assistée par ordinateur facilite toutes les étapes de la collecte. On a recueilli les renseignements de base sur tous les membres du ménage. Les renseignements touchant le comportement ou ceux fondés sur la perception de soi sont recueillis lors d'une entrevue privée avec un des membres du ménage choisi au hasard, et qui devient ainsi un membre du panel. L'ensemble des membres du panel composent l'échantillon longitudinal qui fera l'objet d'un rappel tous les deux ans pour une période maximale de 20 ans.

La population visée par l'ENSP englobe les membres de ménages résidant dans chaque province et territoire, sauf les habitants des réserves indiennes, des bases des Forces canadiennes ou de certaines régions éloignées. Un volet institutionnel, documenté ailleurs², traite des résidents à long terme de centres hospitaliers et d'établissements de soins pour bénéficiaires internes.

L'enquête de 1994 visait à produire des estimations provinciales transversales. Établie initialement à 22 000 ménages, la taille de l'échantillon est passée à 26 000 ménages après l'achat, par les provinces, d'unités d'échantillonnage dans le but de permettre la production d'estimations infraprovinciales. Il y a eu quatre périodes de collecte, soit en juin, août, et en novembre 1994, et en mars 1995. Pour chaque période, l'enquête a porté sur un ensemble distinct de ménages.

Plan d'échantillonnage des ménages

Le plan d'échantillonnage des ménages a été établi selon quatre critères:

- C la taille visée pour les différents échantillons aux niveaux national, provincial et territorial;
- C la décision de choisir un membre par ménage pour constituer le panel longitudinal;
- C le choix de l'Enquête sur la population active (EPA) comme véhicule pour sélectionner l'échantillon;
- C la décision d'intégrer l'ENSP et l'Enquête longitudinale nationale sur les enfants (ELNE).

Les trois premiers critères ont déterminé respectivement la répartition de l'échantillon, l'application d'une technique pour accroître la représentativité (la «méthode d'exclusion») et le tirage des échantillons provinciaux à l'extérieur du Québec.

Règle de sélection des personnes interviewées

Certaines enquêtes sur la santé recueillent des renseignements sur un seul membre du ménage. C'est le cas de l'Enquête promotion santé Canada³ de 1990 et de la *New Zealand Household Health Survey* de 1992-1993⁴. Plusieurs autres enquêtes portent sur tous les membres du ménage, comme l'Enquête santé Canada⁵ de 1978-1979, l'Enquête sur la santé en Ontario⁶ de 1990, l'Enquête sociale et de santé de 1992-1993, effectuée au Québec⁷, et la *National Health Interview Survey*, aux États-Unis⁸.

Il y a certains avantages à faire des entrevues avec chaque membre d'un même ménage. Cela permet, par exemple, de comparer certaines caractéristiques quant à la santé des membres d'un même ménage. Par ailleurs, il n'en coûte pas beaucoup plus d'interviewer tous les membres d'un ménage que d'en interviewer un seul.

Toutefois, en plus d'accroître le fardeau de réponse du ménage, cette approche présente aussi des inconvénients. Par exemple, les fortes corrélations existant entre les membres d'un ménage pour certaines caractéristiques rendent ce type d'échantillon moins pertinent, à certains égards, qu'un échantillon de même taille tiré d'un plus grand nombre de ménages. De plus, dans une perspective longitudinale, il est plus complexe, du point de vue logistique, d'assurer le suivi de tous les membres d'un ménage échantillonné. En effet, les familles se divisent et forment de nouveaux ménages, ce qui vient augmenter la taille de l'échantillon.

L'ENSP constitue un compromis entre l'interview d'un seul membre d'un ménage et celle de tous ses membres. Dans le cadre de cette enquête, la plupart des renseignements demandés portent sur un seul membre du ménage, et on recueille un certain nombre de renseignements sur la santé de tous les membres du ménage, y compris les caractéristiques socio-économiques, le recours aux soins médicaux, les limitations d'activités, et les conditions chroniques. Cela permet d'interviewer en profondeur la personne sélectionnée, d'obtenir un échantillon désagrégé des caractéristiques du ménage, et de simplifier le suivi longitudinal. Chaque fois que la personne choisie aux fins du suivi longitudinal est à nouveau interviewée, on recueille les mêmes renseignements de base sur la santé de tous les membres du ménage dont fait partie cette personne au moment de l'interview.

Le fait de définir le panel comme étant constitué d'une seule personne par ménage et de recueillir un nombre limité de renseignements sur tous les membres du ménage est en désavantage au plan financier, puisqu'il faut contacter suffisamment de ménages pour obtenir le nombre requis de personnes au sein du panel. Cependant, cela permet d'obtenir davantage de réponses aux questions sur l'état de santé et la situation socio-économique.

L'approche adoptée par l'ENSP peut comporter un autre inconvénient, soit que le panel longitudinal risque

d'inclure un nombre démesurément élevé de membres faisant partie de petits ménages. En effet, les chances qu'une personne soit incluse dans le panel, sont inversement proportionnelles au nombre de personnes appartenant au ménage. On a atténué en partie ce problème en rejetant certains ménages dont aucun des membres n'étaient âgés de moins de 25 ans (cette mesure est précisée plus loin).

Avec l'approche adoptée pour l'ENSP, deux types de pondération sont nécessaires aux estimations. L'une d'elles, basée sur la probabilité inverse de sélectionner le ménage, sert à pondérer les réponses données par tous les membres du ménage. L'autre facteur de pondération, basé sur la probabilité inverse de sélectionner le ménage multipliée par la probabilité inverse de sélectionner le membre du ménage appartenant au panel, sert uniquement pour les réponses données par les personnes faisant partie du panel. Le deuxième facteur de pondération, qui dépend de la taille du ménage, est susceptible de fluctuer d'un ménage à l'autre et, par conséquent, d'entraîner une plus grande variabilité des estimations relatives au panel que si tous les membres du ménage en faisaient partie. Cependant, si tous les membres d'un ménage devaient participer au panel, la pondération du panel connaîtrait inévitablement certaines fluctuations à mesure que les ménages se divisent, ou que leur taille s'accroît ou diminue.

Méthode d'exclusion

On a appliqué une méthode «d'exclusion» pour accroître la représentativité du panel. Comme seulement un membre de chaque ménage échantillonné a été sélectionné pour faire l'objet d'une interview en profondeur et pour participer au panel longitudinal, la possibilité qu'une personne soit incluse dans le panel serait en rapport inverse avec le nombre de personnes appartenant à ce ménage. Les membres de ménages de grande taille, notamment les parents et les enfants à charge, auraient donc tendance à être sous-représentés dans le panel, alors que les membres de petits ménages, qui sont souvent célibataires ou âgés, seraient sur-représentés.

Dans le cadre de l'approche d'exclusion, une partie de l'échantillon est soumise à un tri de sélection afin d'exclure de l'enquête les ménages ne comportant personne de moins de 25 ans. Afin de maintenir la taille voulue de l'échantillon, on applique le facteur d'extension $1/(1-P_d)$, P_d représentant le pourcentage de

ménages que l'on s'attend à rejeter, à la taille des échantillons. P_d a généralement été calculé au niveau des provinces, mais appliqué au niveau des strates d'individus au sein des provinces. Par conséquent, on a restauré la taille des échantillons provinciaux, mais pas celle des échantillons stratifiés selon la répartition proportionnelle. La taille des échantillons a augmenté pour les strates qui comprennent un plus faible pourcentage de ménages n'ayant personne de moins de 25 ans, et a diminué pour les strates comportant un plus grand pourcentage de ménages n'ayant personne de moins de 25 ans.

Avant d'adopter l'approche d'exclusion, on a étudié plusieurs autres techniques. On aurait pu accroître la chance relative de sélectionner certains types sous-représentés de personne pour le panel, par exemple les enfants du ménage. Cependant, comme la sous-représentation posait un problème pour les ménages de grande taille, cette solution ne faisait que déplacer le problème vers d'autres membres de ces ménages, vers les parents, par exemple. On aurait aussi pu augmenter la représentation dans les régions où étaient concentrés les ménages de grande taille. Malheureusement, exception faite des immeubles résidentiels, les ménages de taille semblable n'étaient pas suffisamment concentrés, pour qu'on obtienne des résultats satisfaisants. Enfin, on a rejeté des méthodes plus traditionnelles permettant d'accroître la représentativité par sous-échantillonnage, car elles auraient nécessité deux séries de visites aux ménages échantillonnés, une option jugée non pratique. Au Québec, cependant, l'accessibilité des résultats d'une enquête provinciale sur la santé permettait un double échantillonnage (collecte initiale effectuée parmi un échantillon plus important grâce au groupement de ménages selon des caractéristiques observées, les sous-échantillons étant ensuite pris indépendamment dans chaque groupe)⁹.

La technique d'exclusion utilisée pour l'ENSP offre un meilleur rapport coût-champ d'observation que les autres règles d'exclusion envisagées. Par exemple, on a songé à rejeter les ménages ne comportant personne de moins de 20 ans, à rejeter ceux comportant moins de trois membres, et à combiner ces deux types d'exclusion. Pour des raisons financières et opérationnelles, le pourcentage de ménages rejetés au tri par une technique d'exclusion a normalement varié entre 25% et 30% en Ontario, entre 37,5% et 40% dans des régions urbaines à l'extérieur de l'Ontario, et entre 25% et 30% dans les régions rurales. Les

pourcentages ont été inférieurs dans les régions rurales, où il en coûte davantage pour contacter les ménages. Ils l'ont été également en Ontario parce que les nombreuses unités d'échantillonnage achetées par cette province ne comportaient pas d'exclusions (tableau 1). Comme les strates d'appartements renferment une forte concentration de petits ménages, on a réduit la taille de ces échantillons plutôt que d'appliquer la méthode d'exclusion. L'approche d'exclusion n'a pas été appliquée dans les régions éloignées, où il coûte cher de contacter les ménages, et son utilisation a été limitée dans les régions où le nombre d'unités d'échantillonnage achetées était important.

Répartition initiale de l'échantillon

Le budget de l'ENSP permettait un échantillon de 22 000 ménages. Il fallait au moins 1 200 ménages par province et par territoire pour assurer la fiabilité des estimations selon le sexe et le groupe d'âge. Sous réserve de cette restriction, on a obtenu la taille des divers échantillons provinciaux et territoriaux en recourant à la méthode de répartition de Kish, qui met en équilibre les besoins de fiabilité aux niveaux national et régional¹⁰. Selon cette méthode, l'échantillon a été réparti proportionnellement parmi $1/(0,804W_h^2 + 1/12^2)$, où W_h représente la proportion des ménages dans la province ou le territoire h ($h=1, \dots, 12$) selon le recensement de 1991. On pouvait augmenter l'échantillon en achetant d'autres unités d'échantillonnage.

Pour obtenir des estimations plus précises, il est préférable que les provinces soient stratifiées sur une base géographique, et parfois sur la base de caractéristiques socio-économiques, en régions relativement homogènes appelées strates. Dans chaque province, l'objectif consistait à répartir la taille de l'échantillon en strates proportionnellement aux données démographiques du recensement de 1991. La répartition proportionnelle a été choisie pour les raisons suivantes: a) elle promettait de meilleurs résultats pour les estimations provinciales des ratios et des pourcentages; b) elle pouvait produire des échantillons autopondérés (c'est-à-dire des unités ayant toutes le même poids d'échantillonnage) et donc plus faciles à analyser; c) elle constituait un compromis acceptable entre des plans où l'information auxiliaire, en corrélation avec des variables d'étude, n'était pas disponible et ceux où la multitude de caractéristiques étudiées concernait des ensembles différents de

variables auxiliaires; d) elle pouvait simplifier le recours à un plan d'échantillonnage polyvalent.

Cependant, la répartition proportionnelle n'est pas «optimale», qu'il s'agisse de minimiser le coût pour une fiabilité donnée, ou de maximiser la fiabilité pour un coût donné (les coûts de collecte étant supérieurs dans les strates rurales et éloignées) dans le cas d'une enquête axée sur la mesure d'une seule série de caractéristiques connexes, comme le revenu. Par ailleurs, la répartition proportionnelle ne tient pas compte des exigences liées aux estimations infraprovinciales: la taille des échantillons est souvent insuffisante pour produire des estimations fiables dans certaines régions.

Achat d'unités par les provinces et autres modifications de l'échantillonnage

Quatre provinces ont décidé d'augmenter leur échantillon afin de répondre à certains critères de fiabilité pour certaines sous-populations précises. Dans chaque cas, l'échantillonnage additionnel a été attribué à certaines régions sanitaires (secteurs géographiques infraprovinciaux que les provinces utilisent à des fins administratives). Ces unités d'échantillonnage ajoutées ne devraient pas être intégrées à l'échantillon longitudinal.

Ainsi, l'Ontario a acquis d'autres unités d'échantillonnage dans ses régions sanitaires afin d'obtenir des estimations d'une exactitude donnée pour deux ou trois groupes d'âge-sexe par région. Le Manitoba, pour sa part, a demandé des échantillons de 450 ménages à Winnipeg et de 225 ménages dans les autres régions sanitaires. Dans ces deux provinces, les régions peu peuplées du Nord ont été fusionnées pour maintenir à un niveau raisonnable le nombre d'unités additionnelles. Le Nouveau-Brunswick, quant à lui, a acquis des unités d'échantillonnage additionnelles pour accroître la représentation des régions sanitaires 4, 5 et 7.

Par ailleurs, la Colombie-Britannique a demandé l'ajout de 850 ménages uniquement pour la région sanitaire de Prince George. Comme il n'y avait pas assez d'intervieweurs locaux pour absorber le volume supplémentaire, la plupart des ménages ajoutés ont été contactés par composition aléatoire (CA). Bien que les personnes rejointes par CA hésitent souvent à donner des renseignements qui permettraient de les retracer, cette stratégie a pu être mise en oeuvre, car

il n'y avait aucune exigence longitudinale attachée aux unités ajoutées. Par contre, les exigences relatives à l'échantillon ordinaire ont été appliquées à la partie non visée par la CA.

La taille de tous les échantillons a également été accrue par le nombre de ménages qu'on s'attendait à voir rejetés à la sélection par la méthode d'exclusion.

Afin de réduire le fardeau de réponse et le coût de la collecte dans les deux territoires, on a décidé d'intégrer les échantillons et les questionnaires de l'ENSP à ceux de l'Enquête longitudinale nationale sur les enfants (ELNE)¹¹. Comme il fallait au moins 1 500 ménages par territoire pour donner l'échantillon voulu d'enfants aux fins de l'ELNE, on a eu recours à une méthode d'exclusion semblable à celle utilisée dans les provinces pour soustraire 300 de ces ménages à la collecte de l'ENSP. Par conséquent, les questions de l'ENSP figurant dans le questionnaire intégré ont été posées à seulement 1 200 ménages par territoire.

Le tableau 1 donne la taille prévue des échantillons par province et par territoire. Cette taille est indiquée pour les ménages à interviewer et pour ceux dont on prévoit qu'ils seront rejetés à la sélection, après un bref contact par méthode d'exclusion. Les chiffres représentent le nombre de logement privés qui étaient admissibles, et ce, avant que ne soient éliminées les non-réponses (estimées à près de 10%).

Tirage de l'échantillon

Dans chaque province, sauf au Québec, l'ENSP a fait appel au plan d'échantillonnage polyvalent mis au point pour le remaniement de 1994 de l'Enquête sur la population active (EPA)^{12,13}. Le plan de base de l'EPA est un échantillon stratifié à plusieurs degrés composé de logements sélectionnés parmi les grappes. Pour les fins du plan, chaque province est répartie en trois types de régions: grands centres urbains, villes urbaines, et régions rurales. Des strates géographiques ou socio-économiques sont formées au sein de chaque grand centre urbain, les grappes renfermant environ 150 à 250 logements. Certains centres urbains ont des strates distinctes pour les appartements et pour les secteurs de dénombrement (SD) du recensement où le revenu moyen par ménage est élevé. On choisit ensuite, dans chaque strate, six grappes ou immeubles résidentiels (parfois 12 ou 18 appartements) par une méthode d'échantillonnage aléatoire avec probabilité proportionnelle à la taille (PPT), la taille correspondant

Tableau 1

Taille de l'échantillon de l'Enquête nationale sur la santé de la population

	Répartition initiale	Unités d'échantillonnage acquises	Taille de l'échantillon (ménages)		
			À interviewer	Rejetés à la sélection	Total
Terre-Neuve	1 220	...	1 221	171	1 392
Île-du-Prince-Édouard	1 201	...	1 199	223	1 422
Nouvelle-Écosse	1 270	...	1 270	246	1 516
Nouveau-Brunswick	1 243	180	1 423	234	1 657
Québec	3 584	...	3 479*	...	3 479
Ontario	4 817	2 183	7 001	1 021	8 022
Manitoba	1 307	493	1 800	324	2 124
Saskatchewan	1 287	...	1 288	257	1 545
Alberta	1 674	...	1 674	305	1 979
Colombie-Britannique (CA non comprise)	1 996	61	2 057	448	2 505
Total partiel	19 599	2 917	22 413	3 229	25 642
Colombie-Britannique unités acquises pour CA	...	788	788	...	788
Yukon	1 200	...	1 200	300	1 500
Territoires du Nord-Ouest	1 200	...	1 200	300	1 500
Total	21 999	3 705	25 601	3 829	29 430

* L'échantillon du Québec est inférieur à celui de la répartition initiale parce que 100 unités ont été mises de côté afin de régler des problèmes éventuels de couverture de la base de sondage.

au nombre de ménages. Le plan de l'EPA prévoit un renouvellement du sixième de l'échantillon tous les mois.

Les autres villes et régions rurales de chaque province sont stratifiées sur une base géographique d'abord, puis selon les caractéristiques socio-économiques. Il s'agit généralement de régions visées à la fois par la Commission de l'assurance-chômage et par l'EPA comme régions économiques. Dans la plupart des strates, six grappes (habituellement des SD du recensement) sont sélectionnées au moyen de la méthode probabilité proportionnelle à la taille (PPT). Dans certains cas caractérisés par la faible densité de la population, on obtient un plan à trois degrés en sélectionnant d'abord deux ou trois unités primaires d'échantillonnage (UPE), qui correspondent habituellement à des groupes de SD, et en divisant chaque UPE en grappes, dont six sont échantillonnées. La sélection se fait à chaque degré au moyen de la méthode PPT.

Une fois terminées les opérations de listage des grappes de l'échantillon, on obtient l'échantillon des

logements. Les taux d'échantillonnage étant déterminés d'avance, il existe souvent un écart entre la taille prévue de l'échantillon et les chiffres effectivement obtenus. Le rendement de l'échantillon, par exemple, est parfois excessif. Pour réduire le coût de la collecte, on corrige un rendement excessif en éliminant une partie des unités sélectionnées à l'origine, et en modifiant les poids du plan d'échantillonnage. Cette opération, habituellement réalisée à un niveau agrégé, est appelée stabilisation de l'échantillon. De plus, on augmente la taille requise des échantillons, en ménages, pour tenir compte des logements, l'expérience ayant prouvé que 15% de tous les logements ne sont pas occupés par des ménages faisant partie du champ d'observation (certains logements étant vacants ou occupés sur une base saisonnière, d'autres étant occupés par des ménages non visés par l'enquête). Le plan d'échantillonnage donne environ 60,000 ménages pour l'EPA. Pour les fins d'enquêtes requérant des échantillons moins importants, on «réserve» habituellement de un à six renouvellements par province, un renouvellement correspondant au sixième de l'échantillon total. La stabilisation de l'échantillon sert à maintenir celui-ci au

niveau souhaité; c'est le cas, notamment, lorsqu'on réserve deux renouvellements, même si la taille d'échantillon voulue représente seulement 1,5 renouvellement.

Cette méthode d'échantillonnage adoptée pour l'EPA a été modifiée pour répondre aux besoins de l'ENSP. D'abord, le plan de l'EPA ne reflétait pas les besoins de l'ENSP pour ce qui est de la répartition infraprovinciale, notamment à cause de l'acquisition d'autres unités d'échantillonnage infraprovinciales. Demander un nombre fixe de renouvellements dans l'ensemble d'une province se serait avéré insuffisant pour certaines régions, et inefficace pour d'autres. Il fallait donc pouvoir déterminer le nombre de renouvellements aux niveaux infraprovinciaux.

Ensuite, l'ENSP devrait répondre à des exigences d'échantillonnage supplémentaires au niveau des grappes. Pour estimer la variance, il fallait répartir les grappes de chaque strate en deux échantillons répétés ou plus (c.-à-d. en sous-échantillons sélectionnés d'une manière identique, mais indépendante). Il fallait aussi répartir l'échantillon parmi les quatre périodes de collecte. Par souci d'économie, il était toutefois préférable de visiter chaque grappe au cours d'une seule période de collecte. Aussi fallait-il que le nombre de grappes retenues par strate soit de huit ou d'un autre multiple de quatre.

Par suite de ces modifications apportées au plan de l'EPA, l'échantillon de grappes de l'ENSP peut être considéré comme un échantillon stratifié répété dont les strates sont constituées de groupes de strates initiales 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 une incidence appréciable sur les résultats de l'enquête.

Intégration avec l'Enquête longitudinale nationale sur les enfants

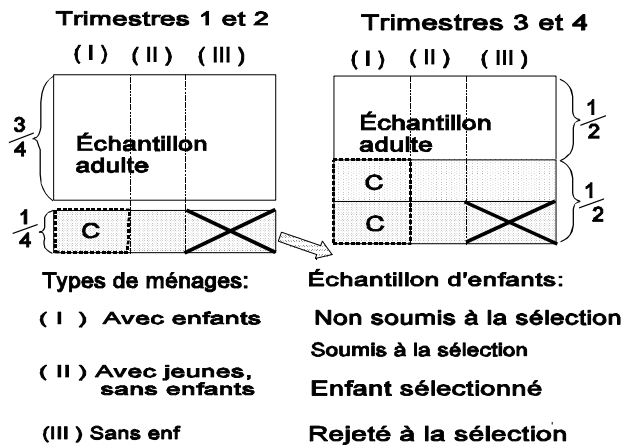
L'Enquête longitudinale nationale sur les enfants (ELNE) est une enquête-ménage qui suivra dans le temps un échantillon d'environ 25 000 enfants de moins de 12 ans. L'échantillon a été obtenu à partir de ménages comprenant des enfants actuellement ou récemment visés par l'EPA. La collecte de données initiales a eu lieu en décembre 1994 et en février 1995, et le suivi des enfants sélectionnés doit avoir lieu tous les deux ans.

On est en train d'intégrer l'ELNE et l'ENSP, car les deux enquêtes portent sur des caractéristiques semblables des enfants. Dans les territoires, les enquêtes feront appel à un échantillon commun de ménages, ainsi qu'au même questionnaire. L'intégration dans les provinces se limite à la collecte de données communes sur les enfants et à l'utilisation d'un même logiciel d'interview assistée par ordinateur.

Dans les provinces, ENSP fournit un échantillon de 4 000 à 5 000 enfants à l'ELNE dans les provinces, ce qui permet à cette enquête de réduire la taille des échantillons. Pour ce faire, on pose les questions détaillées à tous les enfants (jusqu'à concurrence de quatre) d'un même ménage échantillonné aux fins de l'ENSP. Après la collecte de l'ENSP, les données détaillées sur les enfants sont traitées par l'ELNE et utilisées pour les estimations de cette enquête.

Vu les impératifs liés au calendrier des activités, aucun enfant n'a pu être retenu pour le panel de l'ENSP avant la troisième période de collecte (ou trimestre). Cela a faussé la représentativité saisonnière des enfants au sein du panel et a réduit la taille de l'échantillon correspondant. Afin d'accroître le rendement de l'échantillon des enfants, sans nuire à la représentativité saisonnière des autres membres du ménage au cours des deux derniers trimestres, on a décidé de réattribuer à ces trimestres une partie de l'échantillon de l'ENSP. Comme cette décision a été prise après les opérations d'échantillonnage ci-dessus, la réattribution a été appliquée à des ménages au sein des grappes plutôt qu'à des grappes entières.

La figure 1 montre comment on a remanié la répartition de l'échantillon aux fins de l'intégration. Le carré de gauche représente une grappe affectée au trimestre 1 ou 2, et celui de droite, une grappe affectée au trimestre 3 ou 4. Les ménages sont classés par type comme suit: (I) ménages comptant des enfants (de moins de 12 ans); (II) ménages avec jeunes («jeunes» désignant les moins de 25 ans); et (III) ménages sans enfants ni jeunes. L'échantillon comprend des ménages «adultes» et des ménages «enfants». Dans les ménages «adultes», seules les personnes de 12 ans ou plus peuvent être sélectionnées pour faire partie du panel. Les procédures appliquées aux ménages «enfants» varient selon le type de ménage. S'il y a des enfants (ménages de type I), on a choisi au hasard l'un d'eux pour l'intégrer au panel. S'il n'y a pas d'enfants, on rejette le ménage à la sélection (applicable aux ménages



Santé Québec a fourni des renseignements permettant de classer son échantillon en quatre types de ménage: les ménages à un seul membre; les ménages avec enfants; les autres ménages comptant des jeunes (de moins de 25 ans); les autres (plus d'un membre, sans enfants ni jeunes). Aux fins de l'ENSP, on a classé par imputation aléatoire les ménages qui n'ont pas répondu à l'ESS selon la distribution recueillie pour les ménages interviewés par l'ESS au sein de la même grappe.

La taille de l'échantillon de l'ENSP a d'abord été répartie parmi les quatre catégories de densité urbaine. Afin d'éviter la concentration de l'échantillon à Montréal, la répartition a été proportionnelle à $1/(2W_h^2 + 1/4^2)$, où W_h représente la part de la population pour la catégorie h , $h=1, 2, 3, 4$. Dans chaque catégorie, on a tenté d'obtenir un sous-échantillon de l'ESS qui serait, en ce qui concerne le membre du panel sélectionné, proportionnel aux populations des quatre types de ménage. Pour ce faire, on a tiré de l'ESS un nombre suffisant de ménages pour donner le rendement voulu de ménages avec enfants (le groupe le plus sous-représenté), après quoi on a retiré la part excédentaire des trois autres groupes de ménages. On a donc tiré un échantillon supérieur de près de 50% à la répartition initiale. Cependant, après en avoir supprimé les deux tiers des ménages à un seul membre, la moitié des autres ménages sans enfants ni jeunes, et le sixième des ménages avec jeunes, mais sans enfants, on a presque atteint l'objectif. Les mêmes ratios ont été appliqués à toutes les catégories de densité urbaine, la distribution des types de ménages étant analogue.

Au Québec comme ailleurs, la représentation saisonnière, l'estimation de la variance et l'intégration avec l'ELNE ont eu une incidence sur le sous-échantillonnage. Ainsi, on a 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 avec enfants a été réparti, selon un rapport de 2 à 3, entre un échantillon «adultes» et un échantillon «enfants». Les ménages «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 enfants a lui aussi été réparti selon un rapport de 2 à 3, entre un échantillon «adultes» et un échantillon «enfants». Cela donnait aux enfants devenus membres de ces ménages entre l'ESS et l'ENSP, notamment les nouveaux-nés, une chance d'être inclus dans le panel. Lorsqu'aucun enfant n'était présent, on a choisi un membre âgé de 12 ans ou plus comme membre du panel dans le ménage «enfants» (on n'a fait aucune exclusion au Québec).

Le tableau 2 donne la ventilation prévue de l'échantillon selon les données de l'ESS. La faible surrepartition des ménages comptant un seul membre est voulue. En effet, ce groupe se caractérise par un taux de non-réponse plus élevé et il est le plus susceptible de s'accroître à mesure qu'évolue la composition des ménages.

Tableau 2

Ventilation de l'échantillon au Québec selon le type de ménage

Type de ménage	Répartition de la population	Répartition échantillon*	
	%	Nombre	%
Total	100	3 479	100
Ménages B un seul membre	9	415	12
Ménages avec enfants	37	1 293	37
Enfant sélectionné	15	526	15
Autre personne sélectionnée	22	767	22
Autres ménages avec jeunes	29	1 033	30
Autres ménages sans enfants ni jeunes	24	738	21

* Les chiffres ayant été arrondis leur somme peut ne pas correspondre aux totaux indiqués.

Plan d'échantillonnage dans les territoires

Au Yukon et dans les Territoires du Nord-Ouest, l'ENSP et l'ELNE ont été réalisées comme enquête unique pour réduire le fardeau de réponse et le coût de la collecte, facteurs beaucoup plus élevés dans ces régions. Dans chaque territoire, on a sélectionné comme échantillon 1 500 ménages, dont 300 ont été rejetés à la sélection. Il s'agissait de donner un rendement suffisant d'enfants aux fins de l'ELNE. Cette démarche supposait que l'on tienne compte aux fins de l'ELNE des enfants des ménages faisant partie de l'échantillon (jusqu'à concurrence de trois enfants par ménage).

Dans les Territoires du Nord-Ouest, les ménages ont été sélectionnés au hasard dans chaque collectivité d'au moins 100 personnes. Certaines collectivités éloignées ont été exclues.

Au Yukon, à Whitehorse et dans trois collectivités de taille moyenne, on a eu recours à la composition aléatoire (CA). Ailleurs, les ménages ont été sélectionnés par PPT à partir d'un échantillon tiré de collectivités plus importantes. Neuf régions et collectivités ont été exclues à cause de leur petite taille ou du coût de l'enquête dans ces régions. On a eu recours à des interviews sur place chaque fois qu'il n'y a pas eu de CA. Les exclusions représentaient 2% de la population des Territoires du Nord-Ouest et 15% de la population du Yukon.

Ventilation prévue de l'échantillon par âge et sexe

Le tableau 3 présente la ventilation prévue par âge et sexe pour l'ensemble des échantillons provinciaux, sauf les unités d'échantillonnage acquises par la Colombie-Britannique et contactées par CA. Les chiffres sont approximatifs étant donné la variabilité liée au plan d'échantillonnage, à la méthode d'exclusion, à la sélection des personnes et au taux de non-réponse. La non-réponse peut en effet réduire d'environ 10% le rendement de l'échantillon.

Les chiffres ne reflètent pas deux changements survenus tard dans l'élaboration du plan. Ainsi, à l'Île-du-Prince-Édouard, la décision de ne pas déplacer l'échantillon vers les trimestres 3 et 4 a réduit le nombre de ménages avec enfants et, par conséquent, le nombre d'enfants au sein du panel. De plus, après le trimestre 1, on a découvert que les enfants de 12 ans n'étaient pas sélectionnés dans le panel. Des mesures compensatoires ont été prises pour les trimestres 3 et 4 exclusivement, pour que cette distorsion saisonnière de l'échantillon des enfants de 12 ans soit la même que pour les enfants de moins de 12 ans.

Les chiffres pour l'échantillon global sont basés sur le nombre total de personnes dans les ménages échantillonnés, tandis que ceux du panel sont basés sur le nombre de personnes choisies pour en faire partie (un par ménage). On donne les chiffres pour l'échantillon global parce que certaines questions de l'enquête sont posées pour tous les membres du ménage.

Tableau 3

Rendement estimé de l'échantillon selon l'âge et le sexe de l'Enquête nationale sur la santé de la population

	Total	Moins de 12 ans	12-24 ans	Âge		
				25-44	45-64	65 ans et plus
Chiffres pour l'échantillon global (tous les membres du ménage)						
Total	65 578	12 712	13 799	21 951	11 506	5 610
Hommes	32 500	6 513	6 988	10 657	5 851	2 491
Femmes	33 078	6 199	6 811	11 294	5 655	3 119
Chiffres pour le panel (un membre par ménage)						
Total	22 431	2 839	4 004	7 738	4 585	3 265
Hommes	10 808	1 458	1 982	3 824	2 246	1 298
Femmes	11 623	1 381	2 022	3 914	2 339	1 967

La représentativité des personnes choisies pour le panel a été accrue par la méthode d'exclusion à l'extérieur du Québec, et par le plan particulier au Québec. À l'extérieur du Québec, la méthode d'exclusion a réduit d'environ 37% la sous-représentation des panelistes appartenant à des ménages de type I (avec enfants). La méthode d'exclusion a aussi réduit de moitié la surreprésentation des ménages de type III (n'ayant personne de moins de 25 ans). Comme il a été mentionné précédemment, les résultats pour le Québec sont encore plus précis.

En ce qui concerne l'intégration avec l'ELNE, malgré l'exception que constitue l'Île-du-Prince-Édouard, on prévoit interviewer 4 746 enfants (2 434 garçons et 2 312 filles). Ce nombre est plus élevé que celui du panel, car l'ELNE doit traiter tous les enfants des ménages «enfants», jusqu'à concurrence de quatre, dans chaque province.

À propos du plan à venir

Les objectifs parfois contradictoires des estimations longitudinales et transversales soulèvent des questions touchant le plan d'échantillonnage pour les futurs passages de l'ENSP. Après un certain nombre d'années, le panel longitudinal, bien qu'il soit une source d'estimations transversales fiables pour 1994, deviendrait inefficace pour les estimations transversales. D'ici 2004, plus du sixième de la population sera née ou arrivée au Canada depuis 1994¹⁴. Ces nouveaux habitants feront l'objet des estimations transversales seulement s'ils vivent avec une personne qui était admissible au panel en 1994. La fiabilité diminuera également en raison de l'attrition, c'est-à-dire de la réduction des échantillons par suite de décès, de la non-réponse, du passage hors du champ d'observation et de situations dont on ne peut assurer le suivi (par exemple, des personnes qui ont déménagé et dont on ne connaît pas la nouvelle adresse).

On a proposé deux approches pour remettre à niveau la taille des échantillons et maintenir la fiabilité des estimations transversales. Une approche consistait à ramener au niveau de 1994 les échantillons des passages à venir, de façon à compenser l'attrition.

L'autre approche, soit celle qui a été adoptée, constituait à «mettre en réserve» la taille de l'échantillon remise à niveau pour un passage, et à transférer cette «réserve» à la taille de l'échantillon remise à niveau pour le passage suivant. Selon cette approche, la taille des échantillons sera moins importante en 1996, 2000, 2004, etc., et plus importante en 1998, 2002, 2006, etc., mais la taille moyenne restera la même. Grâce à cette méthode, on disposera d'un échantillon plus important dans la deuxième série d'années pour améliorer les estimations transversales. Les années pour lesquelles l'échantillon sera moins important suffiront à l'échantillon longitudinal, mais les estimations transversales correspondantes seront moins fiables.

Le suivi des personnes faisant partie du panel de 1994 pose des problèmes opérationnels et méthodologiques. Certaines personnes ont déménagé et sont difficiles à retracer. Malheureusement, ces personnes se démarquent fréquemment de la population générale (par exemple, elles tendent davantage à être jeunes, de sexe masculin et sans emploi). Par conséquent, il faudra retracer le plus grand nombre possible de personnes ayant déménagé pour minimiser le biais qu'entraînerait leur non-réponse. En 1994, on a demandé aux membres du panel de fournir le nom et l'adresse de personnes susceptibles de connaître leurs allées et venues, pour le cas où elles déménageraient. D'autres mesures sont prises entre les passages de l'enquête pour identifier ceux qui ont déménagé. Par exemple, on enverra une lettre demandant aux personnes d'indiquer leur nouvelle adresse si elles ont déménagé, ou si elles prévoient déménager. Par ailleurs, on a pris des arrangements avec Postes Canada pour obtenir l'adresse de ceux qui ont déménagé. On a également élaboré un système de suivi informatisé pour transmettre les «cas» à l'intervieweur ou au bureau régional susceptible de retrouver le plus facilement ces personnes. On examine aussi d'autres questions, notamment la forme que prendra l'intégration du ELNE, l'éventuelle correction des variations saisonnières de l'échantillon par période de collecte, ainsi que le suivi de ceux qui n'ont pas répondu en 1994.

Remerciements

Les auteurs désirent remercier M.P. Singh, Ph. D. pour ses suggestions utiles.

Références

1. Catlin G., Will P., "Enquête nationale sur la santé de la population: premiers faits saillants", *Rapports sur la santé* (Statistique Canada, nE 82-003 au catalogue) 1992; 4(3):313-319.
2. Mohl C., "National Population Health Survey Institutional Sample", Statistics Canada. Household Survey Methods Division. Ottawa; 1995. Unpublished.
3. Santé et Bien-être social Canada. "Rapport technique: Enquête promotion santé Canada". Publié sous la direction de, Stephens T, Fowler GD. Ottawa. Ministère des Approvisionnements et Services Canada, 1990.
4. Brown, D. "The 1992-93 New Zealand Household Health Survey". *The Survey Statistician*. 1994;30:10-12.
5. Statistique Canada et Santé et Bien-être Canada. "*La Santé des Canadiens: Rapport de l'Enquête Santé Canada*". (Statistique Canada, n° 82-538F au catalogue) 1981.
6. Ontario Ministry of Health. *Ontario Health Survey Highlights*. Toronto; 1992.
7. Courtemanche R, Tarte F. "Plan de sondage de l'enquête Santé Québec". Cahier Technique 87-02, Montréal, Enquête Santé Québec, 1987.
8. Adams PF, Benson V. "Current estimates from the National Health Interview Survey". National Center for Health Statistics. *Vital Health Stat*. 1991; 10(181).
9. Cochran WG. *Sampling Techniques*. 3rd ed. New York: John Wiley & Sons; 1977.
10. Kish L. "Plans de sondage à usage multiple". *Techniques d'enquête* (Statistique Canada, n°12-001 au catalogue) 1988; 14(1):19-33.
11. Montigny G. "L'enquête longitudinale nationale sur les enfants (ELNE)". *Rapports sur la santé* (Statistique Canada, n° 82-003 au catalogue) 1993; 5(3):317-320.
12. Singh MP, Drew JD, Gambino JG, Mayda F. "*Méthodologie de l'enquête sur la population active du Canada 1984-1990*" (Statistique Canada, n° 71-526 au catalogue) 1990.
13. Singh MP, Gambino J, Laniel N. "Research Studies for the Labour Force Survey Sample Redesign". *American Statistical Association - 1994 Proceedings of the Section on Survey Research Methodology*. Alexandria, VA. In press.
14. Statistique Canada. *Projections démographiques pour le Canada, les provinces et les territoires 1993-2016* (Statistique Canada, n° 91-520 au catalogue) 1994.