

ENQUÊTES SUR LA SANTÉ -- ÉCHANTILLONS TRANSVERSAUX

Aspects pouvant expliquer les différences entre des estimations obtenues à deux différentes occasions d'enquête

** Travail en cours (Février 2003)
STC/DMEM

Depuis 1994, la Division de la santé a produit via ses enquêtes, une série de fichiers de données provenant d'échantillons transversaux. Contrairement à des échantillons longitudinaux, ces échantillons ont la particularité d'être uniquement représentatif de l'année durant laquelle les données ont été recueillies. Les données transversales disponibles proviennent de l'Enquête nationale sur la santé de la population (ENSP) pour les années 1994-95, 1996-97 et 1998-99, puis de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) pour 2000-01. Dans les cas où une variable a été recueillie à plusieurs occasions, il est possible pour les analystes de produire des estimations transversales, et ainsi d'examiner la tendance de cette variable dans le temps. Inévitablement, on observera des différences dans ces estimations, et la source de ces différences peut être multiple. Ce document rapporte les différents aspects pouvant expliquer les différences entre les estimations mesurées à partir des différents fichiers transversaux de l'ENSP et l'ESCC. À noter que les comparaisons avec l'ENSP sont faites en référant au fichier Santé (par opposition au fichier Général), c'est-à-dire le fichier contenant les données recueillies auprès de la personne sélectionnée. Ce fichier ressemble beaucoup à ce qui est fait pour l'ESCC, tant en terme de contenu qu'au niveau de l'échantillonnage (i.e. qu'ils incluent tous deux une sélection de personne(s) dans chaque ménage).

Aspects méthodologiques

- Population cible:

L'ENSP (composante ménage) et l'ESCC couvrent la même population et font face aux mêmes exclusions. La seule différence vient du fait que l'ESCC ne couvre que les personnes âgées de 12 ans et plus, alors que l'ENSP couvre, de façon générale, toute la population. Les détails de couverture de l'ENSP sont donnés plus loin dans ce document. À cause de cette différence et pour faciliter les comparaisons, les indicateurs présentés dans le document réfèrent, lorsque possible, uniquement aux personnes de 12 ans et plus. À noter aussi qu'en terme de géographie, les deux enquêtes couvrent les 10 provinces et les territoires. Toutefois pour l'ENSP, les territoires sont couverts par une composante indépendante (composante Nord), et pour cette raison, les territoires ont été exclus lors des calculs fait dans ce document.

- Questionnaire:

Une différence dans la façon dont la question est formulée peut avoir un impact sur les estimations. La majorité des concepts mesurés par les enquêtes de l'ENSP et l'ESCC utilisent la même question à travers le temps; il faut toutefois s'en assurer à l'aide des questionnaires avant de faire toute interprétation. Il en va de même pour les variables dérivées qui pourraient avoir été construites différemment d'une occasion à une autre.

- Collecte

	ENSP 1994-95	ENSP 1996-97	ENSP 1998-99	ESCC 2000-01
Période de collecte	Juin 1994 à juin 1995	Juin 1996 à July 1997	Juin 1998 à juin 1999	Sept. 2000 à Oct. 2001
Mode (% par téléphone; 12+)	27,7%	98,9%	91,1%	53% ²
Taux de réponse (ménage; all ages)	88,7%	82,6%	87,6%	89,9%
Taux de réponse (personne; 12+)	95,8%	95,6%	98,4%	92,6%
Taux de réponse par procuration (12 +) ¹	4,2%	2,3%	2,4%	6,3%
Longueur d'interview (approximatif)	50 min.	50 min.	50 min.	45 min.

1 Certains modules ne peuvent pas être demandés par procuration. Voir les questionnaires pour connaître lesquelles.

2 La structure opérationnelle utilisée pour faire les interviews téléphoniques a changé lors de l'année 2000. Ainsi, pour l'ENSP, tous les interviews téléphoniques étaient faits à partir du domicile de l'intervieweur. Pour l'ESCC, une certaine portion était aussi faite du domicile de l'intervieweur, alors que l'autre portion était faite à partir des sites d'appels.

Composition de l'échantillon transversal:

Enquête	Composition	Provenance (bases utilisées)	Population Couverte	Nombre de personnes sélectionnées par ménage	Représentativité géographique	Particularité
ENSP 1994-95	Membres du panel + Échantillon acheté pour 4 provinces (ON, CB, NB, MAN)	Aréolaire (panel + achat ; 84%) + RDD (achat; 16%)	0+	1	National + provincial, et régional pour ON, CB, NB & MAN	
ENSP 1996-97	Membres du panel + Échantillon acheté pour 3 provinces (ON, AB, MAN)	Membres choisis en 1994-95 sont recontactés (panel; 19%) + RDD (achat; 81%)	2+, sauf ON, AB et MAN où c'est 0+	<ul style="list-style-type: none"> • Panel = pas de sélection (même personne qu'au cycle 1). • RDD ON = 1 personne de 12+ • RDD AB & MAN = 1 personne de 12+, et un enfant (0-11) lorsque possible. 	National + provincial, et régional pour ON, AB, MAN	
ENSP 1998-99	Membres du panel + Échantillon de remise à niveau	Membres choisis en 1994-95 sont recontactés (panel; 87%) + RDD (remise à niveau; 13%)	0+	1 (même personne qu'au cycle 1 pour le panel)	National + provincial	Échantillon de remise à niveau est composé de bébés (0-1 an), et d'immigrants nouvellement reçus. Tiré à partir de groupes de rotation sortants de l'EPA
ESCC 2000-01	Échantillon purement transversal	Aréolaire (82%) + Bases téléphoniques (18%) -- Le pourcentage varie d'une région à l'autre	12+	<ul style="list-style-type: none"> • Base aréolaire = 1 ou 2 dépendant de la composition du ménage • Bases téléphoniques = 1 	National + provincial + régional	

Note:

- Pour L'ENSP de 1996-97 & 1998-99, la partie d'échantillon formée des membres du panel pourrait être perçue comme un groupe de personnes plus coopératives étant donné qu'ils ont déjà fait l'engagement de faire partie d'un panel.
- Pour l'ESCC, le groupe d'âge 12-19 a été suréchantillonné au profit des 20-64, ce qui procurera des meilleures variances pour les estimations pour ce groupe d'âge. Ce suréchantillonnage a été effectué en sélectionnant une ou deux personnes par ménage selon la composition de celui-ci.
- La répartition de l'échantillon a été faite selon les besoins en représentativité de chaque occasion d'enquête. Par exemple, l'échantillon de l'ESCC a été réparti de façon à bien couvrir chacune des 136 régions socio-sanitaires, alors que l'échantillon de l'ENSP l'avait été, de façon générale, en s'assurant uniquement une bonne représentativité à l'échelle provinciale. La composition de l'échantillon est donc beaucoup plus "rural" pour l'ESCC que pour l'ENSP à cause de cette contrainte de couverture plus détaillée de l'ensemble du pays. La pondération contrôle toutefois cette plus forte surreprésentation du milieu rural pour l'ESCC (voir tableau ci-dessous).

Pourcentage de l'échantillon et de la population vivant en milieu rural (12+ & provinces seulement)

	ENSP 1994-95	ENSP 1996-97	ENSP 1998-99	ESCC 2000-01
Échantillon (% rural)	23,4%	21,2%	23,2%	26,4%
Population (% rural pondéré)	16,8%	17,5%	18,5%	18,3%

- Taille d'échantillon (répondants de 12 ans et plus des fichiers maîtres)

Province	ENSP 1994-95	ENSP 1996-97	ENSP 1998-99	ESCC 2000-01
CANADA (exclus les territoires)	17 626	73 402	15 249	129 018
Terre-Neuve	918	868	875	3 870
Île-du-Prince-Édouard	899	829	844	3 651
Nouvelle-Écosse	911	882	943	5 319
Nouveau-Brunswick	1 111	929	948	4 996
Québec	2 581	2 521	2 593	22 667
Ontario	5 187	39 010	4 148	39 278
Manitoba	1 420	11 816	1 021	8 470
Saskatchewan	1 005	942	980	8 009
Alberta	1 310	14 203	1 384	14 456
Colombie-Britannique	2 284	1 402	1 513	18 302

NOTE: La différence dans les tailles d'échantillon se reflètera évidemment sur la précision des estimations produites avec les différents fichiers de données.

- Pondération:

- Saisonnalité:

Pour l'ENSP, la pondération n'a jamais inclus d'ajustement spécifiquement pour contrôler la saisonnalité. Toutefois, la collecte est effectuée en périodes égales (trimestres) couvrant de façon plus ou moins équivalente les quatre saisons. Pour ce qui est de l'ESCC, la collecte avait aussi été planifiée pour répartir équitablement l'échantillon entre les quatre saisons, toutefois, des problèmes opérationnels survenus durant la collecte ont débalancé le tout. Pour remédier à ce débalancement, un ajustement contrôlant pour la saison a été incorporé dans la pondération.

- Post-stratification:

La post-stratification a pour but de rétablir les sommes de poids de façon à ce qu'elles correspondent exactement à la population estimée. La post-stratification est faite indépendamment à l'intérieur de chaque région/province, et ce pour un certain nombre de groupes d'âge-sexe. Ces groupes ont été définis différemment au cours des occasions d'enquête, tel que présenté ci-dessous.

Groupes d'âge utilisés pour la post-stratification:

- ENSP 1994-95: 12-24, 25-44, 45-64, 65+ (pas d'enfants au cycle 1)
- ENSP 1996-97: 2-11, 12-24, 25-44, 45-64, 65+ (sauf pour les provinces ayant un achat d'échantillon où le groupe 0-1 a été ajouté)
- ENSP 1998-99: 0-11, 12-24, 25-44, 45-64, 65+ (une étape de pré-poststratification a été appliquée aux groupes 0-3 & 4-11, à l'échelle du Canada*sexe)
- ESCC 2000-01: 12-19, 20-29, 30-44, 45-64, 65+

Note: Pour des estimations du nombre total de personnes par groupe d'âge, plus le groupe d'âge se rapproche de l'intervalle utilisé pour la post-strate, plus la variance sera petite (par exemple, à tailles d'échantillon égales, une estimation pour le nombre de 12-17 aura un CV beaucoup plus petit avec l'ESCC qu'avec l'ENSP puisque ce groupe d'âge est presque le même qu'une des post-strates, i.e. 12-19.

- Imputation:

L'ENSP n'a jamais eu recours à l'imputation pour aucun des trois premiers cycles concernés dans ce document. Toute valeur manquante est codée en conséquence, sans être remplacée par une autre valeur dans les fichiers de données. De son côté, l'ESCC a dû imputer quelques variables en raison d'un taux de réponse par procuration trop élevé. Dans le cas de réponse par procuration, plusieurs questions n'étaient pas demandées étant donné leurs caractères plutôt privés ou personnels. Conséquemment, un taux élevé de non-réponse à ces questions a été observé. L'imputation a donc été utilisée pour obtenir des données pour ces questions non répondues et résultant d'une interview par procuration. L'article de St-Pierre et Béland (2002) présente la mise en situation complète de même que la méthode utilisée.

Référence: St-Pierre, M. & Béland, Y. (2002). Imputation of Proxy Respondents in the Canadian Community Health Survey. *Proceedings of the Survey Methods Section*. Statistical Society of Canada.

- Méthode de calcul de la variance (bootstrap):

Le bootstrap est utilisé pour toutes les occasions d'enquête, toutefois, certains détails techniques diffèrent d'une occasion à l'autre.

- ENSP 1994-95: incorpore seulement la post-stratification
- ENSP 1996-97: incorpore seulement la post-stratification
- ENSP 1998-99: incorpore la non-réponse (ménage et personne) et la post-stratification
- ESCC 2000-01: incorpore dans le bootstrap tous les ajustements à partir de l'ajustement de non-réponse ménage

- Variabilité échantillonnale

Le fait que l'information soit recueillie auprès d'un échantillon, et non auprès de toute la population, implique que les résultats obtenus seront tous sujets à une variabilité échantillonnale. La variabilité attachée à chaque estimation produite peut dans certains cas expliquer la différence entre les estimations obtenues aux différentes occasions d'enquête. Pour vérifier si une différence est vraiment significative, donc pas uniquement due à la variabilité des estimations, on doit procéder à des tests statistiques. Par exemple, un test de Student pourra vérifier si deux agrégats diffèrent significativement ou non un de l'autre.

Aspects contextuels

- Changement dans les normes relatives à la santé

Certaines variables sont dérivées en se référant à une certaine norme. Par exemple, selon sa valeur, l'indice de masse corporel permet de déclarer qu'une personne est obèse si son indice est supérieur à une certaine norme. Similairement, certaines normes cliniques sont utilisées pour déterminer si une personne souffre d'une certaine maladie ou d'un problème de santé chronique. Ces normes peuvent être appelées à changer dans le temps selon les avancements dans le domaine de la santé.

Par exemple, le critère utilisé pour déclarer une personne diabétique a été modifié dans les années 90. Selon les critères de l'Organisation mondiale de la Santé (OMS) établis en 1985, le diabète est défini par un taux de glycémie à jeun égal ou supérieur à 7,8 mmol/L et/ou par un taux de glycémie égal ou supérieur à 11,1 mmol/L deux heures après l'épreuve d'hyperglycémie. En 1997, l'*American Diabetes Association* a adopté le taux de glycémie à jeun comme critère principal et a réduit son taux de 7,8 à 7,0 mmol/L. Les lignes directrices de pratique clinique 1998 pour le traitement du diabète au Canada ont par la suite adopté ce changement. Ce changement pourrait en théorie avoir un effet sur l'incidence (et prévalence) du diabète au Canada. Il est donc important d'être au courant des changements adoptés aux directives cliniques pour le diagnostic d'une maladie ou problème de santé chronique.

- Changement réel dans la population

Après avoir examiné tous les aspects méthodologiques, il reste que les différences observées entre deux occasions d'enquête peuvent demeurer le fruit de la réalité. La santé est un domaine très dynamique et en constante évolution; les différents indicateurs de la santé sont donc sujet aux fluctuations.