

**Avis méthodologique pour la combinaison de plusieurs cycles de
l'Enquête sur la santé dans les collectivités canadiennes**

Juin 2007

Avis méthodologique pour la combinaison de plusieurs cycles de l'Enquête sur la santé dans les collectivités canadiennes

France Lapointe et Robert Courtemanche
Direction de la méthodologie, de la démographie
et des enquêtes spéciales

Le 20 juin 2007

Pour tout renseignement concernant l'ISQ et les données statistiques qui y sont disponibles, s'adresser à :

Institut de la statistique du Québec
200, chemin Sainte-Foy
Québec (Québec)
G1R 5T4
Téléphone : (418) 691-2401
ou
Téléphone : 1 800 463-4090
(aucuns frais d'appel au Canada et aux États-Unis)

Site Web : www.stat.gouv.qc.ca

Introduction

Le Ministère de la santé et des services sociaux du Québec a demandé à l'Institut de la statistique du Québec un avis méthodologique concernant l'intégration de plusieurs cycles de l'ESCC pour l'étude de phénomènes peu fréquents au niveau national ou régional ou encore de sous-populations particulières. Les questions posées sont les suivantes :

Est-ce que la collecte continue permettra une meilleure couverture des indicateurs au niveau régional? C'est-à-dire, la collecte continue permettra-t-elle de couvrir des indicateurs mal documentés ou des phénomènes rares au niveau national et des régions?

1. Contexte

Ces questions se posent en particulier pour les cycles à venir de l'ESCC. En effet, à partir du cycle 4.1, la collecte de données sera étalée sur une période de deux années et suivie sans arrêt par le cycle 5.1. Pour les cycles antérieurs, soit les cycles 1.1, 2.1 et 3.1, l'enquête était bisannuelle avec une collecte étalée sur un période d'une année. Ainsi, l'échantillon québécois du cycle 1.1 de l'ESCC offrait un portrait des Québécois de 12 ans et plus en 2001; celui du cycle 2.1 présentait les Québécois de 12 ans et plus en 2003; enfin, l'échantillon du Québec du cycle 3.1 de l'ESCC permettait l'inférence à la population québécoise de 12 ans et plus visée en 2005.

À partir du cycle 4.1, des estimations de même qualité que celles produites avec les données des cycles 1.1, 2.1 ou 3.1 seront obtenues en combinant les données recueillies au cours de deux années. Au cycle 4.1 par exemple, ce seront les données issues de la collecte faite en 2007 et en 2008 qui permettront de telles estimations. En effet, ce n'est qu'après ces deux années de collecte que le cycle 4.1 sera complété et que l'échantillon aura la même ampleur que chacun des cycles précédents.

Puisque les collectes des cycles suivants se feront ensuite en continu, il est naturel d'évaluer la possibilité de regrouper les données de plus de deux années de collecte pour produire des statistiques qui ne seraient pas suffisamment précises si elles étaient calculées avec l'échantillon d'un seul cycle d'enquête.

Cela soulève cependant quelques questions supplémentaires.

- Y a-t-il des conditions nécessaires à la production de statistiques pertinentes à partir d'échantillons de cycles combinés?
- Quelle est la nature des paramètres étudiés lorsqu'on combine des cycles?
- Quelle est la fiabilité des estimations régionales et infrarégionales produites par la combinaison des cycles 4.1 et 5.1 de l'ESCC?

2. Conditions nécessaires à la combinaison

Absence d'effet de cycle :

Les besoins des ministères et des régions correspondent en général à la description de phénomènes présents sur une population finie et pour une période de temps donnée. Pour que les statistiques issues de la combinaison de cycles aient un sens et répondent à ces besoins d'information, il importe de pouvoir garantir qu'il n'y ait pas « d'effet de cycle »¹. En effet, on doit s'assurer qu'à un cycle ou à un autre, la même population finie est couverte et qu'on en mesure les mêmes paramètres. Cela suppose que d'un cycle à un autre :

1. la population cible est la même ;
2. le plan de sondage et plus généralement la méthodologie de l'enquête assurent la même mesure de cette population;
3. les questions et choix de réponses sont les mêmes d'un cycle à un autre.

En d'autres mots, pour combiner les données, on doit d'abord pouvoir garantir le contrôle de la « mesure » d'un cycle à l'autre. Statistique Canada, qui s'est déjà penché sur la question de la combinaison des cycles pour l'avenir², a intégré au plan d'échantillonnage des cycles futurs certaines mesures de contrôle. Par exemple, à l'avenir, on s'assurera que la répartition de l'échantillon de l'ESCC entre les deux bases de sondage utilisées, la base aréolaire et la base téléphonique, soit la même d'un cycle à un autre. Cette approche devrait permettre le respect des deux premières conditions à la combinaison des cycles.

Les analyses faites par l'ISQ au sujet de la comparabilité des statistiques produites aux cycles 1.1 et 2.1³ ont montré que les changements dans la répartition de l'échantillon entre les bases de sondage entre les cycles compromettent la comparabilité des estimations. Cela soulève la question de l'intégration des données obtenues d'achats d'échantillons supplémentaires faits par des régions sociosanitaires pour les cycles à venir⁴. À moins que ces achats ne soient eux-mêmes faits en harmonisant la répartition de l'échantillon supplémentaire entre les bases avec celle de l'échantillon principal, les données obtenues des échantillons supplémentaires ne pourraient pas être considérées pour la combinaison de cycles.

¹ On dira qu'il y a « effet de cycle » si les différences entre deux cycles de l'ESCC sont expliquées par des changements dans la mesure des paramètres d'un cycle à un autre.

² Voir par exemple THOMAS, Steven. (2006). *L'intégration des cycles de l'Enquête sur la santé dans les collectivités canadiennes*, article présenté au Symposium 2006 de Statistique Canada, à paraître.

³ CÔTÉ, Luc, Robert COURTEMANCHE et Bernard CARON, avec la collaboration de C. DAVELUY, C. ROBITAILLE, R. BARALDI et M. BORDELEAU (2005). *Comparabilité entre les cycles 1.1 et 2.1 de l'Enquête sur la santé dans les collectivités canadiennes : impact du changement apporté à la répartition de l'échantillon selon la base de sondage*, Québec, Institut de la statistique du Québec, 69 p.

⁴ Depuis la première version de la présente note, un avis méthodologique sur les achats d'échantillons supplémentaires a été produit par l'Institut de la statistique du Québec et envoyé à Statistique Canada. (Réf. CÔTÉ, Luc, Robert COURTEMANCHE (2007). *Enquête sur la santé dans les collectivités canadiennes : Incidence des achats d'échantillon*, Québec, Institut de la statistique du Québec, 6 p.). Cet avis précisait qu'à l'avenir ces achats devraient se faire de façon à minimiser les impacts sur la comparabilité des estimations régionales. Cet avis a été accepté par Statistique Canada.

Absence d'effet de temps :

En plus de l'absence « d'effet de cycle », la combinaison des données de l'ESCC 4.1 et 5.1 repose sur certaines hypothèses concernant « l'effet du temps »⁵. En effet, si le phénomène étudié était sujet à des effets conjoncturels durant la période couverte par les cycles 4.1 et 5.1, la combinaison des données ne serait pas indiquée. L'interprétation de statistiques se rapportant à ce phénomène, produites avec les données combinées, serait alors ardue. En fait, la combinaison des cycles 4.1 et 5.1 est envisageable dans tous les cas où une valeur moyenne a un sens. C'est-à-dire lorsqu'elle représente une réalité de la population. Or, s'il a eu un changement dans le temps, la moyenne ne représentera pas bien la situation présente dans la population antérieurement au changement, au cycle 4.1, ni celle qui prévaudrait en 5.1.

Ainsi, les phénomènes pour lesquels on peut supposer qu'il n'y a pas de changement durant la période couverte par les cycles 4.1 et 5.1 pourraient être étudiés avec les données combinées. De même, pour tous ceux dont le changement au cours de cette période est peu important ou prend la forme d'une tendance générale, on pourrait considérer la combinaison des cycles. Des tests statistiques faits avec les données provinciales pourraient aider à vérifier l'hypothèse de changement peu important et l'examen des données des cycles antérieurs pourrait révéler une tendance générale; cela pourrait aider à prendre une décision éclairée. Enfin, les risques de « mauvaises » décisions découlant d'une combinaison injustifiée devraient aussi être mis dans la balance.

3. Nature des paramètres étudiés par la combinaison

Pour les cycles 1.1, 2.1 et 3.1, la majorité des statistiques produites correspondaient à des prévalences estimées chez la population cible *pour une année donnée*. Par exemple, la prévalence estimée de fumeurs au Québec au cycle 1.1 nous informait sur la proportion de fumeurs chez les Québécois de 12 ans et plus en 2001.

La collecte des données du cycle 4.1 se fait quant à elle en continu sur une période de deux ans. Ainsi, les données de l'échantillon du cycle 4.1, fourniront un état de situation « moyen » pour les années 2007 et 2008. La collecte du cycle 5.1 suivra immédiatement celle de 4.1, sans interruption des opérations. La combinaison des données des cycles 4.1 et 5.1 permettra donc de connaître le portrait moyen de l'état de santé des Québécois de 12 ans et plus de 2007 à 2011. Si une information de cette nature répond aux besoins des ministères et régions, on considérera la combinaison des cycles 4.1 et 5.1 dans la mesure où les conditions présentées plus haut sont respectées.

En contrepartie, on comprendra qu'il est plus difficile de parler de prévalence moyenne d'un phénomène en 2001 et en 2003. C'est que la discontinuité des données des cycles 1.1, 2.1 ou 3.1, associée aux interruptions d'un an entre chaque cycle de collecte de données, représente un obstacle à une interprétation simple et naturelle des données issues d'une combinaison de cycles antérieurs au cycle 4.1. Cette difficulté s'ajoute aux problèmes de comparabilité des données des cycles antérieurs révélés par une étude particulière sur la question (voir par exemple Côté, L. et autres, 2005).

⁵ On dira qu'il a « effet du temps » si les différences entre deux cycles s'expliquent par des changements dans le temps des paramètres de santé.

4. Fiabilité des estimations régionales et infrarégionales issues de la combinaison des cycles 4.1 et 5.1

Des calculs approximatifs de précision des estimations qui pourraient être faits avec les données recueillies entre 2007 et 2011 par l'ESCC ont été faits. On a supposé que les cycles 4.1 et 5.1 fourniraient tous deux des échantillons semblables à celui du cycle 3.1 de l'enquête, excluant les achats d'échantillons supplémentaires faits alors par le Bas-St-Laurent, la région de Montréal et celle de Laval. Par ailleurs, tous les calculs sont faits à partir des renseignements tirés du Fichier de microdonnées à grande diffusion (FMGD) du cycle 3.1 de l'ESCC; les effets de plan régionaux de ce fichier ont notamment été pris en considération. On a ainsi estimé la taille attendue des échantillons régionaux de l'échantillon combiné des cycles 4.1 et 5.1 et calculé la précision approximative des estimations qu'on en tirerait.

Estimations régionales

Pour chaque région, on a obtenu la prévalence minimale des phénomènes pouvant être étudiés avec un coefficient de variation d'au plus 15 %. On considère en effet que de telles estimations seraient fiables. Certaines sous-populations définies par le genre ou le groupe d'âge ont aussi été examinées. Le tableau 1 résume ces résultats. En comparaison, les mêmes calculs ont été faits en considérant seulement un cycle d'enquête (voir le tableau 3 à l'annexe).

La question étant de savoir si on pourrait couvrir des indicateurs mal documentés ou des phénomènes plus rares au niveau régional en combinant les données de deux cycles, un seuil minimal d'intérêt pour une prévalence de 10 % a été visé et le critère de fiabilité correspond ici à l'obtention d'une estimation ayant un coefficient de variation d'au plus 15 %.

On a d'abord voulu déterminer si les échantillons combinés permettront un examen valable d'une prévalence d'au moins 10 % pour chacune des régions sociosanitaires. La première colonne du tableau 1 révèle que ce sera possible de produire de telles estimations pour toutes les régions. Par exemple, l'échantillon combiné du Bas-Saint-Laurent permettrait de produire une bonne estimation (CV inférieur à 15 %) pour les phénomènes présents chez au moins 3 % des personnes de 12 ans et plus vivant dans cette région. En comparaison, avec seulement le cycle 4.1 de l'ESCC (voir le tableau 3 en annexe), l'échantillon régional attendu du Bas-Saint-Laurent fournira des estimations de même qualité pour des prévalences d'au moins 5 %.

Un examen plus complet du tableau 1 révèle qu'à l'exception de la région 2410 Nord-du-Québec, les données régionales de l'échantillon combiné permettraient une étude valable des hommes séparément des femmes et des personnes âgées de 12 à 44 ans séparément des 45 ans et plus. Les régions de Montréal (2406), de la Montérégie (2416) et de la Capitale Nationale (2403) pourraient également étudier des groupes d'âges plus fins : les 12-24 ans, les 25-44 ans, les 45-64 ans et les 65 ans et plus.

Tableau 1. Prévalence minimale des phénomènes pouvant être étudiés, avec la combinaison des cycles 4.1 et 5.1 de l'ESCC, au niveau régional avec un coefficient de variation d'au plus 15 %, selon la région sociosanitaire, le groupe d'âge et le sexe.

Régions sociosanitaires	Prévalence minimale du phénomène étudié								
	Tous	hommes	femmes	12-44 ans	12-24 ans	25-44 ans	45 ans et plus	45-64 ans	65 ans et plus
2401 Bas-Saint-Laurent	3%	6%	5%	6%	14%	9%	5%	8%	12%
2402 Saguenay – Lac-Saint-Jean	3%	7%	6%	7%	16%	10%	6%	9%	14%
2403 Capitale Nationale	2%	4%	4%	4%	10%	6%	4%	6%	9%
2404 Mauricie – Centre-du-Québec	3%	6%	5%	5%	13%	8%	5%	8%	12%
2405 Estrie	3%	7%	6%	7%	17%	11%	6%	10%	15%
2406 Montréal-Centre	1%	2%	2%	2%	5%	3%	2%	3%	5%
2407 Outaouais	3%	6%	5%	6%	15%	9%	5%	9%	13%
2408 Abitibi – Témiscamingue	3%	7%	6%	6%	16%	10%	6%	9%	14%
2409 Côte-Nord	3%	7%	6%	7%	16%	10%	6%	9%	14%
2410 Nord-du-Québec	6%	13%	11%	12%	28%	18%	11%	17%	25%
2411 Gaspésie – Îles-de-la-Madeleine	4%	8%	7%	7%	18%	11%	7%	11%	16%
2412 Chaudière – Appalaches	3%	6%	5%	6%	14%	9%	5%	8%	12%
2413 Laval	2%	5%	4%	5%	12%	7%	4%	7%	10%
2414 Lanaudière	3%	6%	5%	6%	14%	9%	5%	8%	12%
2415 Laurentides	3%	5%	5%	5%	13%	8%	5%	8%	11%
2416 Montérégie	2%	4%	3%	4%	9%	5%	3%	5%	8%

Certains indicateurs de santé d'intérêt portent sur des populations particulières. Des calculs ont également été faits pour l'examen des sous-populations suivantes :

- les femmes de 15 à 44 ans;
- les hommes de 40 ans et plus;
- les hommes ou les femmes de 60 ans et plus;
- les hommes ou les femmes de 65 ans et plus.

Afin d'alléger le tableau 1, les résultats de ces calculs ne sont pas présentés en détails. Ceux-ci révèlent qu'en moyenne, les données régionales de l'échantillon combiné anticipé fourniraient de bonnes estimations pour des phénomènes présents chez environ 12 % des femmes de 15 à 44 ans ou chez environ 10 % des hommes de 40 ans et plus. Les phénomènes répandus chez environ 20 % des hommes ou 15 % des femmes de 60 ans et plus pourraient en outre être étudiés dans plusieurs régions. L'étude des hommes ou des femmes de 65 ans et plus serait quant à elle limitée à des phénomènes plus prévalents, de 20 % à 25 %. Ces résultats sont des moyennes; certaines régions seraient donc davantage limitées pour l'étude de ces sous-populations, le Nord-du-Québec notamment. D'autres tireraient davantage des données de cycles combinées; la région de Montréal ou la Montérégie en sont deux exemples.

Estimations infrarégionales

La question de la production d'estimations infrarégionales a aussi été examinée. On s'est intéressé aux estimations portant sur une entité géographique plus fine à l'intérieur d'une région sociosanitaire. De tels besoins en information statistique de santé ont déjà été soulevés antérieurement; ils ont notamment justifié l'achat d'échantillons supplémentaires par certaines régions pour des cycles antérieurs de l'ESCC.

Nous avons dû faire des hypothèses sur la composition des échantillons à venir des cycles 4.1 et 5.1 de l'enquête. Les calculs supposent comme précédemment que ces deux échantillons seraient semblables à celui du cycle 3.1 et que la représentation infrarégionale serait proportionnelle à la population de chaque sous-région. Le tableau 2 présente la proportion de la population régionale, représentée par une sous-région, nécessaire pour étudier des phénomènes répandus chez 10 %, 15 % ou 20 % des 12 ans et plus avec une bonne précision ($CV < 15\%$). Il importe de mentionner que ce critère de qualité pour les estimations est plus sévère que celui qu'a considéré Statistique Canada pour les achats infrarégionaux supplémentaires effectués aux cycles antérieurs. En effet, les tailles visées pour les sous-échantillons infrarégionaux correspondaient alors à l'obtention de statistiques ayant un coefficient de variation de moins de 25 %. Lorsque le coefficient de variation pour une estimation est de 25 %, la marge d'erreur correspondant à un intervalle de confiance à 95 % est égale à 50 % de la valeur de l'estimation. Pour une prévalence donnée, l'échantillon nécessaire pour répondre au critère de Statistique Canada est plus petit que celui permettant d'assurer le respect de notre critère de précision.

Tableau 2. Proportion de la population régionale représentée par la sous-région, nécessaire pour étudier un phénomène dont la prévalence est de 10 %, 15 % ou 20 % avec une précision relative (CV) de 15 %.

Régions sociosanitaires	Prévalence du phénomène d'intérêt		
	p=10%	p=15%	p=20%
2401 Bas-Saint-Laurent	26%	16%	11%
2402 Saguenay – Lac-Saint-Jean	30%	19%	13%
2403 Capitale Nationale	18%	11%	8%
2404 Mauricie – Centre-du-Québec	24%	15%	11%
2405 Estrie	32%	20%	14%
2406 Montréal-Centre	9%	6%	4%
2407 Outaouais	27%	17%	12%
2408 Abitibi – Témiscamingue	29%	18%	13%
2409 Côte-Nord	30%	19%	13%
2410 Nord-du-Québec	60%	38%	27%
2411 Gaspésie – Îles-de-la-Madeleine	35%	22%	15%
2412 Chaudière – Appalaches	26%	16%	11%
2413 Laval	21%	13%	9%
2414 Lanaudière	25%	16%	11%
2415 Laurentides	24%	15%	11%
2416 Montérégie	16%	10%	7%

Les calculs effectués suggèrent, par exemple, que seul l'échantillon combiné correspondant à la sous-région « Sainte-Foy-Sillery-Laurentien » de la Capitale Nationale permettra de produire une estimation pour une prévalence d'au moins 10 % dont le coefficient de variation serait inférieur à 15 %. En effet, cette sous-région représente environ 20 % des 12 ans et plus de la Capitale-Nationale. Avec l'échantillon combiné des cycles 4.1 et 5.1, les prévalences d'intérêt devraient être d'au moins 15 % pour la plupart des sous-régions de la Capitale-Nationale. Toutefois, pour les sous-régions de Charlevoix et de Portneuf, on devrait se contenter d'étudier des indicateurs de santé correspondant à des prévalences beaucoup plus grandes, en l'occurrence supérieures à 20 %, puisque ces sous-régions comptent respectivement environ 5 % et 7 % de la population régionale visée. En effet, cette proportion de la population est inférieure au seuil minimal de 8 % nécessaire pour étudier une prévalence de 20 %.

5. Conclusion

L'intégration de plusieurs cycles de l'ESCC, permettra, dans certains cas, d'étudier des indicateurs de santé mal documentés ou des phénomènes rares aux niveaux national et régional. En moyenne, avec les échantillons régionaux combinant deux cycles, on pourra documenter des indicateurs de santé présents chez environ 3 % de la population cible ou plus. Il faut que les prévalences d'intérêt soit d'au moins 6 % pour que les échantillons régionaux d'un seul cycle offrent des estimations de même précision.

Pour ce faire, il faudra cependant s'assurer de bien contrôler les « effets de cycle » en conservant la même formulation et choix de réponses pour les questions, en s'assurant de couvrir la même population cible et en s'assurant que la méthodologie d'enquête est la même à chaque cycle. En ce sens, les achats d'échantillons supplémentaires ne respectant pas ces conditions ne seraient pas considérés pour la combinaison.

Il sera également nécessaire de choisir avec précaution la méthode d'intégration des données afin de permettre la production d'estimations correspondant aux besoins en information statistique.

Enfin, rappelons qu'idéalement, les phénomènes étudiés par la combinaison des cycles ne devraient pas avoir été sujets à des effets conjoncturels entre 2007 et 2011 ou devraient avoir peu changé dans ce laps de temps.

ANNEXE

Tableau 3. Prévalence minimale des phénomènes pouvant être étudiés, avec le cycle 4.1 de l'ESCC, au niveau régional avec un coefficient de variation d'au plus 15 %, selon la région sociosanitaire, le groupe d'âge et le sexe.

Régions sociosanitaires	Prévalence minimale du phénomène étudié								
	Tous	hommes	femmes	12-44 ans	12-24 ans	25-44 ans	45 ans et plus	45-64 ans	65 ans et plus
2401 Bas-Saint-Laurent	5%	11%	10%	11%	25%	16%	10%	10%	22%
2402 Saguenay – Lac-Saint-Jean	6%	13%	11%	12%	28%	18%	11%	11%	25%
2403 Capitale Nationale	4%	8%	7%	8%	19%	12%	7%	7%	16%
2404 Mauricie – Centre-du-Québec	5%	11%	9%	10%	23%	15%	9%	9%	21%
2405 Estrie	7%	14%	12%	13%	29%	19%	12%	12%	26%
2406 Montréal-Centre	2%	4%	4%	4%	10%	6%	4%	4%	9%
2407 Outaouais	6%	12%	10%	11%	26%	17%	10%	10%	23%
2408 Abitibi – Témiscamingue	6%	12%	11%	12%	27%	18%	11%	11%	24%
2409 Côte-Nord	6%	13%	11%	12%	27%	18%	11%	11%	25%
2410 Nord-du-Québec	12%	23%	20%	22%	43%	31%	20%	20%	40%
2411 Gaspésie – Îles-de-la-Madeleine	7%	14%	12%	14%	30%	20%	13%	13%	27%
2412 Chaudière – Appalaches	5%	11%	10%	11%	24%	16%	10%	10%	22%
2413 Laval	4%	9%	8%	9%	21%	13%	8%	8%	19%
2414 Lanaudière	5%	11%	9%	11%	24%	16%	10%	10%	21%
2415 Laurentides	5%	10%	9%	10%	23%	15%	9%	9%	21%
2416 Montérégie	3%	7%	6%	7%	16%	10%	6%	6%	15%