Composante du produit nº 11-522-X au catalogue de Statistique Canada
La série des symposiums internationaux de Statistique Canada : recueil

Article

Symposium 2008 :
Collecte des données : défis, réalisations et nouvelles orientations

L'interaction entre la collecte des données et le plan de sondage : un gain dans un processus de sélection complexe?

par Marie-Eve Tremblay et Karine Moisan

2009

Statistique Canada Statistics Canada

Canada

Recueil du Symposium 2008 de Statistique Canada
Collecte des données : défis, réalisations et nouvelles orientations

L'interaction entre la collecte des données et le plan de sondage : un gain dans un processus de sélection complexe?

Marie-Eve Tremblay et Karine Moisan [1]

Résumé

L'*Enquête sur la satisfaction des usagers à l'égard des services de santé et des services sociaux du Québec* a pour objectif de fournir des estimations sur la satisfaction selon trois lieux de consultation (hôpital, clinique médicale et CLSC). Puisqu'un usager peut avoir consulté de un à trois lieux, et que le questionnaire ne peut porter que sur un seul de ces lieux, un mécanisme de sélection aléatoire du lieu de consultation a été mis en place. Nécessitant des probabilités de sélection variables, ce mécanisme de sélection a la particularité d'avoir été modifié en cours de collecte pour s'adapter de plus en plus aux disparités régionales, quant à l'utilisation des services de santé et des services sociaux.

Mots clés : Plan de sondage, probabilités variables, vagues de collecte.

1. Introduction

1.1 Présentation générale de l'enquête et de ses objectifs

L'Enquête sur la satisfaction des usagers à l'égard des services de santé et des services sociaux du Québec, 2006-2007 a été menée par l'Institut de la statistique du Québec à la demande du ministère de la Santé et des Services sociaux du Québec et des Agences de la santé et des services sociaux des diverses régions sociosanitaires du Québec. L'ensemble de la méthodologie de l'enquête est présentée dans Neill et autres (2007). Cette enquête, qui en était à sa première édition, s'inscrit dans un programme d'enquête visant à alimenter la réflexion des divers acteurs impliqués dans l'organisation des services de santé en dressant un portrait statistique de la satisfaction et des attentes des usagers à l'égard des services de santé et des services sociaux. Plus spécifiquement, les principaux objectifs de l'enquête étaient :

- d'établir le profil des personnes ayant consulté un professionnel de la santé et des services sociaux au moins une fois dans un hôpital[2], une clinique médicale[3] ou un centre local de services communautaires[4] (CLSC) au cours des douze mois précédant l'enquête;
- de fournir des estimations de la satisfaction des usagers selon les lieux de consultation visés par l'enquête (hôpital, clinique médicale et CLSC) et la région de résidence;
- de mesurer les attentes des usagers selon la région de résidence;
- d'identifier les aspects jugés les plus satisfaisants et les moins satisfaisants à partir d'une analyse conjointe de la satisfaction et des attentes exprimées par les usagers;

---

[1]Marie-Eve Tremblay, Institut de la statistique du Québec, 200, chemin Sainte-Foy, Québec, (Québec), G1R 5T4, Canada (marie-eve.tremblay@stat.gouv.qc.ca); Karine Moisan, Institut de la statistique du Québec, 200, chemin Sainte-Foy, Québec, (Québec), G1R 5T4, Canada (karine.moisan@stat.gouv.qc.ca).

[2]Il est à noter que les personnes ayant consulté un professionnel de la santé et des services sociaux à la clinique externe ou à l'urgence d'un hôpital, ainsi que celles ayant été hospitalisées, sont visées par l'enquête, alors que celles ayant consulté dans un hôpital à vocation psychiatrique ne le sont pas.

[3]Les cliniques médicales privées, c'est-à-dire les cliniques où on doit débourser un certain montant pour obtenir des services, les cliniques dentaires, les bureaux de chiropraticiens, de physiothérapeutes, de massothérapeutes, d'ostéopathes, d'acupuncteurs, etc., sont exclues de l'enquête.

[4]Précisons que les services à domicile sont inclus alors que les services Info-Santé sont exclus.

- de recueillir des données complémentaires facilitant l'étude des liens entre la satisfaction des usagers et certaines grandes caractéristiques sociodémographiques et de santé.

1.2 Collecte des données et questionnaire

La collecte des données s'est déroulée de novembre 2006 à juin 2007 auprès de 38 389 répondants provenant de 16 régions sociosanitaires du Québec (ce nombre correspond à un taux de réponse pondéré de 56,9 %). Les répondants devaient remplir un questionnaire comportant trois parties distinctes. La première documente le profil de consultation des usagers au cours des douze mois précédant l'enquête, c'est-à-dire le nombre de fois où ils ont consulté un professionnel de la santé et des services sociaux, pour eux-mêmes ou pour une personne à charge[5], dans un hôpital, une clinique médicale ou un CLSC. La deuxième partie du questionnaire est consacrée à la satisfaction et aux attentes des répondants. La mesure de la satisfaction se traduit au total par 43 questions : 41 d'entre elles sont liées à des items (compétence de l'intervenant, délai raisonnable pour un rendez-vous, propreté des lieux, etc.), regroupés au sein de 12 sous-dimensions qui relèvent de 3 grandes dimensions de la qualité des services, alors que les 2 autres questions portent sur la satisfaction globale et sur l'opinion générale à l'égard du système de santé au Québec. Enfin, la troisième partie concerne certaines caractéristiques de santé ou encore certaines caractéristiques sociodémographiques du répondant.

2. Détermination des probabilités de sélection du lieu de consultation

2.1 Description du plan de sondage et des étapes de sélection

La population visée par l'enquête est composée de l'ensemble des personnes de 15 ans et plus, vivant dans un ménage privé au Québec, et qui ont consulté un professionnel de la santé et des services sociaux dans un hôpital, une clinique médicale ou un CLSC, pour elles-mêmes ou pour une personne à charge, au cours des douze mois précédant l'enquête.

La sélection de l'échantillon de l'enquête a été faite à partir d'une génération aléatoire de numéros de téléphone (GANT) permettant d'atteindre des ménages privés admissibles à l'enquête. Pour chacun des ménages ainsi joint, une personne de 15 ans et plus a été sélectionnée aléatoirement parmi les personnes ayant consulté un professionnel de la santé et des services sociaux dans l'un ou l'autre des lieux de consultation retenus pour l'enquête au cours des douze mois précédents.

Rappelons qu'un des objectifs de l'enquête est de fournir des estimations fiables par région sociosanitaire et par lieu (hôpital, clinique médicale et CLSC). Le plan de sondage a donc été développé de manière à pouvoir estimer une proportion de 15 % et plus avec un coefficient de variation d'au plus 10 % pour chacun des lieux de consultation d'une région. Si la région de chaque numéro de téléphone (donc de chaque répondant) est connue d'avance, les lieux consultés ne le sont pas. De ce fait, il est impossible de sélectionner un échantillon par lieu de consultation *a priori*. De plus, comme on demande au répondant de fournir son degré de satisfaction pour une expérience précise, celle-ci doit être sélectionnée aléatoirement puisque des contraintes de temps et de fardeau du répondant ne permettent pas de recueillir de l'information sur plus d'un lieu de consultation par répondant. En effet, la durée moyenne d'entrevue pour une seule consultation de référence s'établit à près de 20 minutes. Pour déterminer cette consultation, il est donc nécessaire d'introduire des mécanismes de sélection supplémentaires.

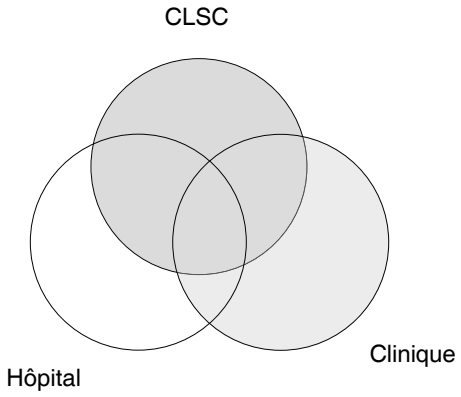
Ainsi, dans le cas où un usager a consulté plus d'une fois au cours des 12 derniers mois, il faut d'abord établir le lieu de la consultation de référence (hôpital, CLSC ou clinique). Ensuite, pour le lieu sélectionné, si l'usager a consulté à la fois pour lui-même et pour une personne à charge, le type de consultation est sélectionné aléatoirement avec des probabilités égales. Finalement, pour le lieu et le type de consultation sélectionnés, l'usager doit se référer à sa dernière expérience de consultation. Toutes les étapes de sélection décrites précédemment nous assurent un échantillon probabiliste des divers lieux et types de consultation.

---

[5]Par personne à charge, on entend un enfant de 14 ans et moins ou une personne ayant une incapacité qui l'empêche de prendre des décisions relatives à sa santé.

Pour sélectionner le lieu de consultation à enquêter, le profil de consultation de chaque répondant doit être établi. Comme un usager peut, au cours des douze derniers mois, avoir consulté un professionnel de la santé et des services sociaux dans l'un, deux ou trois des lieux de consultation retenus pour l'enquête, sept profils différents de consultation sont possibles; la figure 2.1-1 les illustre tous.

Figure 2.1-1
Résumé des différents profils de consultation possibles



Source : Institut de la statistique du Québec, *Enquête sur la satisfaction des usagers à l'égard des services de santé et des services sociaux du Québec, 2006-2007.*

Précisons d'emblée que pour obtenir un niveau de précision des estimations sensiblement égal pour chacun des lieux de consultation dans une région, le nombre de répondants visés devrait être à peu près le même dans chaque cas. Idéalement, le tiers des questionnaires d'une région devrait porter sur chacun des lieux de consultation. Comme les proportions de personnes ayant consulté un professionnel dans les trois lieux de consultation sont très différentes, et que la proportion de celles ayant consulté dans un CLSC est largement inférieure à la proportion ayant consulté dans une clinique ou un hôpital, cela conduit à sélectionner le lieu à enquêter selon des probabilités variables afin de rétablir l'équilibre, dans l'échantillon, entre les trois lieux de consultation.

2.2 Détermination des probabilités de sélection du lieu de consultation

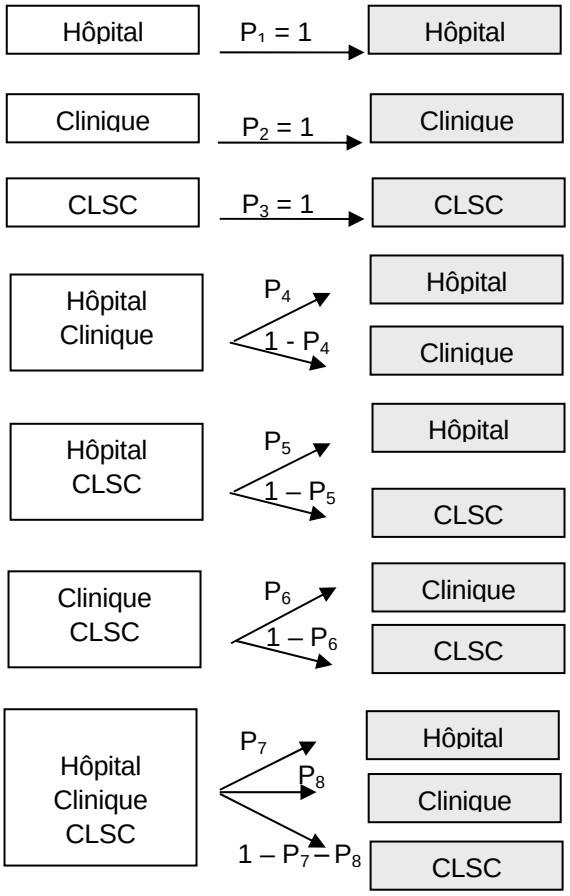
La figure 2.2-1 illustre la démarche de sélection du lieu de consultation : la colonne de gauche présente les divers profils de consultation possibles et la colonne de droite, le lieu sélectionné. Chaque lieu se retrouve dans quatre profils de consultation. Les probabilités de sélection associées à chacun des profils (P_1 à P_8) sont les paramètres à déterminer avant le début de la collecte.

Comme mentionné précédemment, les usagers consultent moins dans les CLSC que dans les deux autres lieux. Il faut donc une sélection qui favorise ce lieu lorsqu'il n'est pas le seul à avoir été consulté. Ainsi, en modifiant les probabilités de sélection, on peut augmenter le nombre de répondants (n) pour certains lieux de consultation, en l'occurrence le CLSC, mais il y a une perte de précision possible associée à la variabilité des probabilités de sélection. La taille d'échantillon efficace ($n_{efficace}$) permet de tenir compte de ces deux phénomènes. Elle est définie comme : $n_{efficace} = \frac{n}{e.p.}$ où $e.p.$ est l'effet de plan mesurant la perte de précision par rapport à un plan de sondage aléatoire simple. La proportion de répondants d'un lieu de consultation donné provenant de chacun des quatre profils de consultation possibles est modifiée lorsque les probabilités de sélection sont changées. Plus cette distribution s'éloigne de celle que l'on a dans la population, plus l'effet de plan augmente; cela contribue à diminuer la taille de l'échantillon efficace pour produire les estimations. En définitive, notre sélection vise donc à égaliser le plus possible les tailles d'échantillons efficaces projetées des trois lieux de consultation pour 15 régions sociosanitaires, dans le but d'obtenir des précisions semblables dans tous les lieux et pour toutes les régions.

Mentionnons qu'aucune sélection du lieu de consultation n'a été faite dans la région du Nord-du-Québec, en raison d'une organisation des services différente dans cette région.

La détermination des probabilités de sélection peut se faire au moyen de divers critères que l'on cherche à optimiser. Ainsi, on cherche à la fois à minimiser la somme des $n_{efficaces}$ manquants pour les trois lieux de consultation, à minimiser les écarts entre ces $n_{efficaces}$ manquants et à maximiser le nombre de combinaisons région et lieu de consultation pour lesquelles la précision voulue est atteinte.

Figure 2.2-1
Mécanisme de sélection du lieu de consultation



Source : Institut de la statistique du Québec, *Enquête sur la satisfaction des usagers à l'égard des services de santé et des services sociaux du Québec, 2006-2007*

Cela dit, l'optimisation des probabilités doit se faire en tenant compte de certaines contraintes. Premièrement, pour tous les usagers ayant consulté un seul lieu de consultation au cours des 12 derniers mois, aucune sélection ne peut être faite, puisque ce lieu est automatiquement sélectionné. Ce phénomène peut avoir un impact important. Par exemple, dans la région du Bas-Saint-Laurent, on observe dans la population que seulement 17 % des usagers ayant consulté à l'hôpital n'avaient consulté aucun autre lieu de consultation, mais la proportion de répondants dont la consultation de référence a eu lieu à l'hôpital qui n'avaient consulté aucun autre lieu atteint 54 %. Deuxièmement, lorsque les probabilités sont modifiées trop fortement pour favoriser le CLSC, les tailles d'échantillons efficaces pour la clinique et l'hôpital deviennent trop faibles, et il vaut mieux atteindre les objectifs de précision dans deux des trois lieux de consultation plutôt qu'avoir des tailles d'échantillons efficaces semblables, mais trop faibles pour les trois lieux. Enfin, il faut toujours laisser une chance minimale d'échantillonner chacun des lieux possibles pour conserver le caractère probabiliste de la sélection. Ainsi, aucun des P_4 à P_8 ne peut prendre les valeurs 0 ou 1.

Une fois le travail d'optimisation des probabilités de sélection complété par les méthodologistes, l'information est transmise aux responsables de la collecte des données afin d'être implantée dans le logiciel de type ITAO. Le tableau 2.2-1 présente l'ensemble des probabilités de sélection ayant servi lors de la première vague de collecte.

Tableau 2.2-1
Ensemble de probabilités de sélection pour la première vague de collecte

	Probabilités de sélection (%)		
Lieux consultés	Hôpital	Clinique	CLSC
Hôpital	100	0	0
Clinique	0	100	0
CLSC	0	0	100
Hôpital - Clinique	55	45	0
Hôpital - CLSC	30	0	70
Clinique - CLSC	0	20	80
Hôpital - Clinique- CLSC	22	18	60

2.3 Régionalisation des probabilités de sélection

Pour faciliter la gestion et le suivi sur le terrain, l'échantillon a été fractionné aléatoirement en trois lots correspondant à trois vagues de collecte pour lesquelles la collecte a débuté à des moments différents, soit en novembre 2006, janvier et avril 2007. La première vague contenait environ 25 % de l'échantillon initial de numéros de téléphone, la seconde 45 % et la dernière 30 %.

La répartition de l'échantillon selon plusieurs vagues offre la possibilité de suivre l'évolution des divers paramètres de collecte (taux d'admissibilité, de productivité et de réponse) et de les réviser si certaines hypothèses de départ s'avèrent erronées. Dans cette enquête, le taux d'admissibilité, c'est-à-dire la proportion de ménages où au moins une personne de 15 ans et plus a consulté pour un service de santé au cours des 12 derniers mois, n'était pas connu avec précision et nos hypothèses de départ devaient pouvoir être révisées si nécessaire. C'est donc pour procéder aux ajustements le plus rapidement possible que la première vague de collecte était de taille inférieure aux autres.

Par ailleurs, à la suite du prétest de l'enquête, nous avions seulement des estimations provinciales des proportions d'usagers ayant consulté chacun des lieux de consultation. Or, nous savions que ces proportions variaient entre les diverses régions sociosanitaires et les vagues de collecte nous ont donc permis de raffiner les paramètres de sélection du lieu de consultation en cours d'enquête. Ainsi, pour la vague 1 de l'enquête, les probabilités de sélection (présentées au tableau 2.2-1) qui sont introduites dans le questionnaire informatisé (ITAO) ont été déterminées au niveau provincial. Par la suite, l'analyse des données recueillies au cours de la première vague de collecte a permis de définir cinq jeux de probabilités de sélection (P1 à P8), chacun d'entre eux regroupant de une à sept régions dont les profils de consultation étaient assez semblables. Ce sont ces jeux de probabilités qui ont été utilisés pour la vague 2. L'analyse des données provenant de la vague 2 a permis d'en arriver à une série de probabilités spécifiques pour chaque région; ce sont ces dernières qui ont servi à la collecte des données de la vague 3. Cette façon de procéder apporte toutefois une contrainte supplémentaire à l'optimisation des probabilités de sélection, puisque lors du travail en prévision des vagues 2 et 3, on doit tenir compte des résultats obtenus aux vagues précédentes, vagues pour lesquelles on ne peut évidemment plus modifier les probabilités de sélection.

En ayant adapté progressivement les probabilités de sélection aux disparités régionales, le contrôle sur le nombre d'entrevues efficaces complétées par mission n'en a été que meilleur pour chacune des régions, et ceci constitue un avantage important de l'utilisation des vagues de collecte. Nous appellerons ce processus la régionalisation des probabilités de sélection.

2.4 Interaction entre la collecte des données et le plan de sondage

Dans tout ce processus de régionalisation des probabilités de sélection, une interaction constante est nécessaire entre le plan de sondage et les résultats de la collecte de données. Ainsi, pour prévoir le nombre de répondants dans

chacun des lieux de consultation, il faut d'abord prévoir le nombre de répondants (tous lieux confondus) par région. Les hypothèses de départ doivent constamment être revues à la lumière des résultats obtenus. Ce travail a été fait par les responsables de la collecte de l'Institut qui ont développé des outils permettant d'estimer le nombre attendu de répondants en séparant les dossiers en diverses catégories selon leur code de réponse au moment de la projection. Ainsi, certains codes de réponses présentent davantage de chance d'être convertis en répondants que d'autres. À titre d'exemple, il est vraisemblable de croire que parmi les dossiers codés « refus par manque de disponibilité au niveau du ménage » (dossiers pour lesquels la sélection de la personne n'a pas été faite), on obtiendra proportionnellement moins de répondants en fin de collecte que parmi les dossiers codés « rendez-vous ferme avec la personne sélectionnée » ou encore « entente prise avec une tierce personne ». En émettant des hypothèses sur les taux d'admissibilité, de productivité et de réponse pour chacun des codes de réponse intermédiaires, les responsables de la collecte de données sont en mesure d'estimer le nombre de répondants attendus pour chacune des vagues de collecte, et ce, pour toutes les régions sociosanitaires.

Un processus itératif est donc mis en place afin d'utiliser le plus possible ces projections dans la détermination des probabilités de sélection du lieu de consultation. Ainsi, au cours de la vague 1, des projections en nombre de répondants sont faites en cours de collecte. À l'aide des effets de plan estimés pour chaque lieu de consultation, ces projections sont converties en tailles d'échantillons efficaces attendues par région et lieu de consultation. C'est sur cette base que les probabilités de sélection peuvent être régionalisées pour la seconde vague de collecte. Ce processus se répète intégralement pour la troisième vague. À la fin de cette dernière, les projections ne pouvant servir à raffiner les probabilités de sélection puisque la collecte se termine, elles ont plutôt servi à prendre les décisions relatives à la gestion de la fin de la collecte. Au moyen de cette approche, les régions où il apparaissait que quelques efforts supplémentaires permettraient d'atteindre les objectifs pour les trois lieux de consultation étaient privilégiées à d'autres régions où un des lieux de consultation était tellement sous-représenté qu'il aurait fallu un très grand nombre de répondants pour atteindre les objectifs. Soulignons que de telles décisions n'ont été prises que dans les toutes dernières semaines de la collecte, qui a duré près de sept mois, afin d'en minimiser les impacts et d'éviter d'introduire des biais. Toute cette démarche nécessite une collaboration étroite entre les méthodologistes et les responsables de la collecte des données.

3. Résultats

3.1 Résultats obtenus

Les effets de plans moyens obtenus par région au moyen des probabilités de sélection variables sont de 1,9 pour la clinique, 1,7 pour l'hôpital et 1,3 pour le CLSC. Environ 500 répondants efficaces étaient visés pour chacun des lieux dans chacune des régions sociosanitaires, et cet objectif a été atteint dans 38 cas sur 45. Certaines régions, comme le Saguenay-Lac-Saint-Jean, présentaient des proportions d'usagers ayant consulté le CLSC si faibles (environ 27 %) que les probabilités de sélection variables et la régionalisation de ces mêmes probabilités n'ont pas permis d'atteindre les objectifs pour le CLSC. En effet, pour cette région, seulement 370 répondants efficaces ont été obtenus pour le CLSC contre 648 pour l'hôpital et 608 pour la clinique. Lorsque les proportions d'usagers ayant consulté le CLSC sont plus importantes, comme c'est le cas pour la région de Laval par exemple avec environ 44 %, alors les résultats sont beaucoup plus probants comme en témoigne le nombre de répondants efficaces par lieu de consultation qui est de 502 pour l'hôpital, 544 pour la clinique et 543 pour le CLSC.

3.2 Gain apporté par la régionalisation des probabilités de sélection

Afin de vérifier si le fait de modifier les probabilités de sélection en cours de collecte a été bénéfique, une simulation a été effectuée. Elle consiste à estimer les tailles d'échantillons efficaces que nous aurions obtenues si les probabilités de sélection provinciales avaient été utilisées tout au long de la collecte, c'est-à-dire si aucune régionalisation des paramètres n'avait été faite. Les résultats de cette simulation sont présentés aux figures 3.2-1, 3.2-2 et 3.2-3. Ces histogrammes comparent les tailles d'échantillons efficaces réellement obtenues et celles obtenues par la simulation.

Figure 3.2-1
Tailles d'échantillons efficaces obtenus à l'enquête et en supposant le maintien des probabilités de sélection provinciales tout au long de la collecte pour le lieu de consultation CLSC

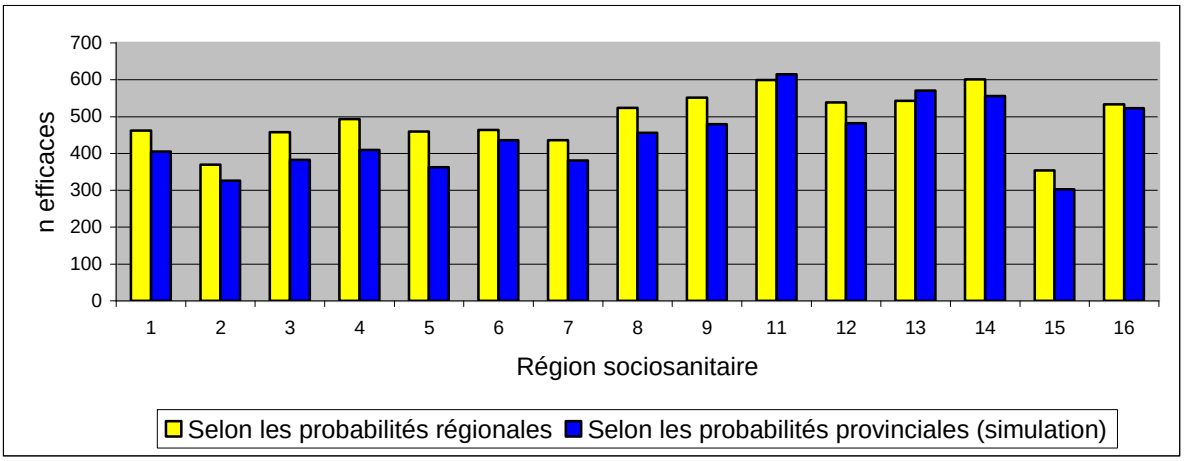


Figure 3.2-2
Tailles d'échantillons efficaces obtenus à l'enquête et en supposant le maintien des probabilités de sélection provinciales tout au long de la collecte pour le lieu de consultation hôpital

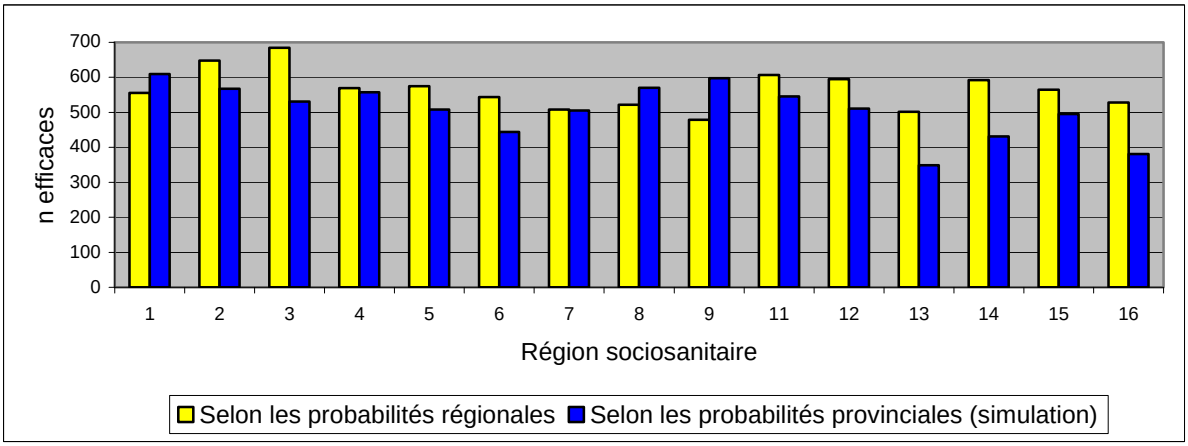
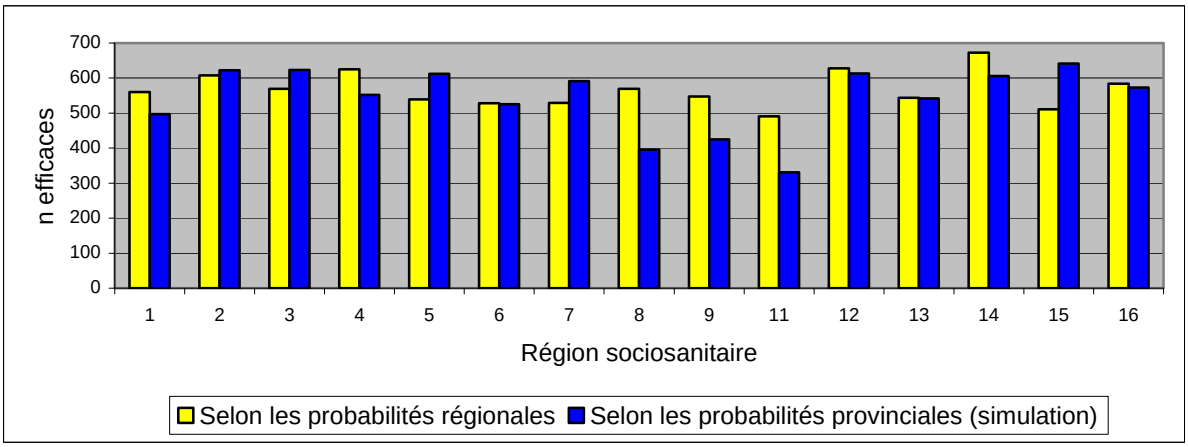


Figure 3.2-3
Tailles d'échantillons efficaces obtenus à l'enquête et en supposant le maintien des probabilités de sélection provinciales tout au long de la collecte pour le lieu de consultation clinique



Rappelons que les objectifs poursuivis étaient entre autres d'obtenir un nombre suffisant de répondants dont la consultation de référence avait eu lieu au CLSC et d'obtenir des tailles d'échantillons efficaces les plus semblables possibles pour les trois lieux de consultation, afin de produire des estimations de précision semblable. Le tableau 3.2-1 confirme donc qu'en moyenne, les tailles d'échantillons efficaces régionales ont augmenté pour tous les lieux de consultation, mais surtout pour le CLSC, ce qui était l'objectif principal du mécanisme de sélection. De plus, on remarque que la régionalisation des probabilités de sélection a permis d'égaliser davantage les tailles d'échantillons efficaces entre les régions puisque les écarts-types sont inférieurs lorsque l'on utilise les probabilités de sélection régionales. Il y a donc lieu de conclure que, bien que les gains ne soient pas spectaculaires, ils sont toutefois appréciables dans l'atteinte des objectifs.

Tableau 3.2-1
Moyenne et écart-type des tailles d'échantillons efficaces par lieu de consultation selon les probabilités de sélection régionales et provinciales

	Moyenne (écart-type) des $n_{efficaces}$		
Probabilités de sélection	CLSC	Hôpital	Clinique
Régionales	493 (73)	565 (55)	567 (49)
Provinciales (simulation)	446 (92)	507 (76)	543 (94)

4. Conclusion

D'une manière plus générale, l'approche présentée dans cet article peut être utilisée dans diverses situations, notamment lorsque, dans le cadre d'une seule enquête, on recherche diverses sous-populations dont les prévalences varient dans la population. L'utilisation des probabilités de sélection variables permet alors d'obtenir un nombre suffisant de répondants appartenant à chacune des sous-populations pour réaliser des estimations sur cette sous-population. De plus, l'utilisation des données de la collecte pour régionaliser certains paramètres peut s'avérer un atout important lorsque l'on fait face à un manque d'information initiale fiable pour un niveau géographique fin. L'interaction entre le plan de sondage et la collecte des données nécessaire à un tel processus permet une plus grande implication des méthodologistes dans la collecte des données et une plus grande contribution des responsables de la collecte au plan de sondage.

Bibliographie

Neill, G., Tremblay, M.-E., Végiard, S., Lavoie, A. et Moisan, K. (2007). Enquête sur la satisfaction des usagers à l'égard des services de santé et des services sociaux du Québec, 2006-2007: description et méthodologie. Quebec, Canada: Institut de la statistique du Québec.