

Utilisation de données administratives pour augmenter l'efficacité du plan de sondage de la nouvelle Enquête nationale sur les voyages

Chunsan (Charles) Choi, Amanda Halladay et Keven Bosa¹

Résumé

Dans le cadre du remaniement du Programme de la statistique du tourisme, Statistique Canada élabore l'Enquête nationale sur les voyages (ENV), qui recueillera de l'information relativement aux voyages effectués par les voyageurs canadiens. Cette nouvelle enquête remplacera l'actuelle Enquête sur les voyages des résidents du Canada, de même que la composante reliée aux voyages des résidents canadiens de l'Enquête sur les voyages internationaux. L'ENV tirera parti des bases de sondage communes de Statistique Canada et des outils de traitement communs, tout en maximisant l'utilisation des données administratives. Dans cette communication, on montrera comment les données administratives, comme celles provenant des fichiers de Passeport Canada, de l'Agence des services frontaliers du Canada et de l'Agence du revenu du Canada, pourraient servir à améliorer l'efficacité du plan de sondage de l'ENV.

Mots-clés : données administratives; base de sondage; plan de sondage; stratification.

1. Introduction

1.1 Contexte

L'Enquête nationale sur les voyages (ENV), dont la collecte a commencé en février 2018, fournira des statistiques sur les voyages des résidents canadiens au pays et à l'étranger. Les données de l'ENV serviront à mesurer la taille et la situation de l'industrie du tourisme au Canada, du point de vue du volume, des caractéristiques et des répercussions économiques des voyages des Canadiens au Canada et à l'étranger. Il s'agira d'une enquête permanente, qui fournira des données sur le revenu et les dépenses touristiques pour le système de comptabilité nationale du Canada. L'ENV remplacera deux composantes de l'actuel Programme de la statistique du tourisme de Statistique Canada : l'Enquête sur les voyages des résidents du Canada (EVRC), qui recueille des renseignements auprès des Canadiens concernant leurs voyages au pays; et le volet des résidents canadiens de l'Enquête sur les voyages internationaux (EVI), qui recueille des données auprès des Canadiens qui voyagent à l'étranger. Le volet des non-résidents du Canada de l'EVI, qui recueille des renseignements auprès des visiteurs étrangers qui voyagent au Canada, sera compris dans une nouvelle enquête appelée « Enquête sur les voyages des visiteurs (EVV) ». Cette nouvelle enquête est toujours en voie d'élaboration et ne sera pas menée avant au moins 2019. La composante étrangère de l'EVI couvrira la période entre la mise en œuvre de l'ENV en 2018 et la mise en œuvre de l'EVV. Le tableau 1-1 résume la restructuration du Programme de la statistique du tourisme.

¹ Chunsan (Charles) Choi, Statistique Canada, 100, promenade Tunney's Pasture, Ottawa (Ontario) Canada, K1A 0T6 (charles.choi@canada.ca); Amanda Halladay, Statistique Canada, 100, promenade Tunney's Pasture, Ottawa (Ontario) Canada, K1A 0T6 (amanda.halladay@canada.ca); Keven Bosa, Statistique Canada, 100, promenade Tunney's Pasture, Ottawa (Ontario) Canada, K1A 0T6 (keven.bosa@canada.ca).

Tableau 1-1**Enquêtes actuelles et nouvelles enquêtes du Programme de la statistique du tourisme**

Lieu de résidence	Destination du voyage	Enquête actuelle	Nouvelle enquête
Canada	À l'intérieur du Canada	EVRC	ENV
Canada	À l'extérieur du Canada	EVI	ENV
À l'extérieur du Canada	Canada	EVI	EVV

1.2 Défis liés aux enquêtes actuelles sur les voyages

L'EVRC est un supplément direct de l'Enquête sur la population active (EPA) au Canada, ce qui signifie que les questions de l'EVRC sont posées aux répondants de l'EPA immédiatement après qu'ils aient soumis leur questionnaire (à un intervieweur par téléphone ou en ligne, grâce à un questionnaire électronique). Comme il s'agit d'un supplément de l'EPA, la durée de l'interview de l'EVRC est limitée à 15 minutes. C'est donc dire que certaines questions ne peuvent pas être posées à l'intérieur de cette limite de temps. Par exemple, on n'a pas suffisamment de temps pour poser aux répondants des questions concernant leurs voyages internationaux. En outre, étant donné que la base de sondage de l'EVRC est tirée directement de celle des répondants à l'EPA, la pondération est complexe en raison du plan de sondage compliqué à plusieurs degrés de l'EPA. L'EPA est une enquête obligatoire et comporte une priorité plus élevée, les répondants étant obligés de participer. Toutefois, l'EVRC est une enquête à participation volontaire. Par conséquent, lorsqu'il y a des contraintes de temps ou de ressources, les intervieweurs donnent la priorité à l'EPA, et ils doivent parfois limiter leurs efforts de collecte des données de l'EVRC.

Le principal défi de l'EVI est l'absence de base de sondage dont on peut tirer un échantillon pour le volet des résidents canadiens de l'enquête. Par conséquent, des cartes d'invitation sont distribuées à la frontière aux résidents du Canada qui reviennent de l'étranger, et les invitent à aller en ligne et à remplir volontairement le questionnaire de l'EVI. Une autre limite de l'EVI est sa dépendance opérationnelle à l'égard de l'Agence des services frontaliers du Canada (ASFC) en ce qui a trait à la distribution des cartes d'invitation. L'ASFC est un organisme fédéral responsable d'assurer la sécurité nationale à la frontière, grâce à un contrôle du flux de personnes et de biens qui entrent au Canada. Étant donné que la distribution des cartes d'invitation de l'EVI n'est pas une priorité pour l'ASFC, celles-ci ne sont pas distribuées comme elles le devraient, particulièrement pendant les heures de pointe dans les points d'entrée les plus occupés. En raison de ces limites, l'EVI a connu un faible taux de questionnaires renvoyés à Statistique Canada.

1.3 Architecture opérationnelle du Bureau

L'Architecture opérationnelle du Bureau (AOB) représente un examen exhaustif des services fournis par Statistique Canada, dont l'objectif est de favoriser un environnement d'intégration et d'efficacité. Dans le cas des enquêtes de Statistique Canada, cela peut être interprété comme des efforts accrus en vue de maximiser l'utilisation d'outils communs, de la collecte de données électroniques et de données administratives. Comme le Programme de la statistique du tourisme est en voie d'être remanié, des efforts sont faits pour s'assurer que toutes les enquêtes sur les voyages respectent le cadre de l'AOb. Dans le cas de l'ENV, cela signifie l'utilisation d'une base commune des ménages, d'un logiciel généralisé d'échantillonnage et d'estimation, d'un questionnaire électronique pour la collecte des données et de sources de données administratives permettant d'améliorer toutes les étapes de l'enquête, au besoin. La présente communication sera axée principalement sur les avantages potentiels de l'utilisation des données administratives pour la stratification de l'échantillon de l'ENV.

2. Utilisation de données administratives pour l'Enquête nationale sur les voyages**2.1 Stratégie de conception de l'enquête**

L'outil de collecte de données de l'ENV sera un questionnaire électronique, et la population cible sera constituée de tous les résidents du Canada âgés de 18 ans et plus ne vivant pas en établissement institutionnel dans les dix provinces. Une lettre d'invitation sera envoyée par la poste aux logements sélectionnés, puis les répondants devront aller en ligne pour remplir le questionnaire. La base de sondage servira à sélectionner les échantillons mensuels dans une base de

sondage commune des ménages gérée par Statistique Canada, appelée Fichier de l'univers des logements (FUL), dont on sait qu'elle comporte une excellente couverture de la population canadienne (~97 %). Le FUL représente une vue instantanée du Registre des adresses et il est mis à jour à partir des données du recensement de la population, des données fiscales et d'autres dossiers administratifs. Une autre base commune gérée par Statistique Canada, le Fichier des indicateurs socioéconomiques (FSE), comprend des données sociodémographiques pour environ 80 % des ménages du FUL. Ces données démographiques du FSE serviront de données auxiliaires dans la stratification du FUL. En outre, un fichier des détenteurs de passeports canadiens devrait être fourni par Passeport Canada et sera aussi utile pour la stratification, afin d'améliorer le taux de réussite (c.-à-d. les proportions de voyageurs sélectionnés dans l'échantillon). La stratégie de stratification consistera à fractionner la base de sondage (FUL) en strates, au moyen des variables de stratification suivantes : variable géographique (10 provinces), groupes de revenu du ménage préclassifiés (9) et indicateur de passeport, servant à déterminer les ménages comptant au moins un titulaire de passeport d'âge adulte.

La répartition de l'échantillon comporte trois étapes : une répartition selon la province, selon la variable du passeport et selon le groupe de revenu. Pour la première étape, l'échantillon total sera affecté à chaque province au moyen d'une méthode de répartition par puissance prenant la forme suivante :

$$\gamma_p = \frac{N_p^q}{\sum_p N_p^q},$$

où N_p représente la population de chaque province et $q=1/3$. Pour la deuxième étape, une répartition fondée sur la variable du passeport sera effectuée. La méthode particulière de cette répartition continue d'être explorée, étant donné que l'on n'a pas encore reçu l'ensemble de données final de Passeport Canada. Enfin, pour la troisième étape, une répartition de Neyman sera effectuée pour affecter l'échantillon à chacun des groupes de revenu. À partir de chaque logement échantillonné, on sélectionnera un adulte pour remplir le questionnaire de l'ENV.

2.2 Utilisation actuelle des données administratives

Plusieurs sources de données administratives sont déjà utilisées pour l'EVI et l'EVRC, et seront adoptées par l'ENV et utilisées pour les mêmes fins. Par exemple, l'EVRC utilise les estimations démographiques mensuelles du recensement pour le calage en fonction des totaux de contrôle nécessaires pour le sexe, le groupe d'âge et la province. De même, l'EVI utilise des données administratives fournies par l'ASFCA, appelées dénombrements à la frontière, qui comportent une vaste gamme de statistiques pour tous les voyageurs canadiens et internationaux dont l'entrée ou le retour au Canada est autorisé (Statistique Canada, 2016). Ces dénombrements à la frontière servent de totaux de contrôle à l'étape de la pondération de l'EVI.

2.3 Sources de données fiscales pour la variable du revenu

On a observé un rapport entre le revenu du ménage et la fréquence des voyages, à partir des enquêtes précédentes sur les voyages. On sait que les ménages qui ont un revenu plus élevé sont plus susceptibles de voyager au pays et à l'étranger. Ainsi, dans un effort en vue d'obtenir un plus grand nombre de voyageurs dans l'échantillon pour l'ENV, on utilise le revenu du ménage comme variable de stratification. Autrement dit, les ménages provenant des strates ayant des revenus plus élevés auront, en général, une probabilité plus grande d'être sélectionnés.

La principale source de données sur le revenu des ménages est le Fichier maître des particuliers T1 (FMP T1), qui est fourni par l'Agence du revenu du Canada et qui comporte un enregistrement par personne. Le revenu du ménage peut être obtenu à partir du FMP T1 en additionnant tous les enregistrements du même logement, puis en les fusionnant avec le FUL, afin d'obtenir une variable de revenu au niveau du logement. Si un logement ne correspond pas à un enregistrement du FMP T1, le revenu du logement peut être obtenu d'une autre source, à savoir le FSE. Le revenu qui est disponible dans le FSE provient du fichier sur la famille T1, qui est aussi fourni par l'Agence du revenu du Canada et qui est mis à jour chaque année. Le revenu du FMP T1 a préséance sur le revenu du FSE parce qu'il est plus à jour.

Afin d'évaluer la qualité du FMP T1, un couplage d'enregistrements a été effectué pour fusionner le fichier avec le FUL en utilisant la variable AR_UID et le numéro d'assurance sociale (NAS). L'AR_UID est un identificateur de

logement unique du FUL et du FSE. Les résultats ont montré que 68 % des enregistrements de la base de sondage (c.-à-d. le FUL) correspondaient au FMP T1. En outre, 16 % des enregistrements du FUL ont été mis en correspondance avec le FSE pour obtenir le revenu du ménage pour les enregistrements qui n'étaient pas appariés avec le FMP T1. Ainsi, le revenu du ménage est disponible pour 84 % des ménages de la base de sondage (FUL).

2.4 Données sur les passeports de Passeport Canada

Il est difficile de s'assurer que les voyageurs canadiens et internationaux sont représentés dans l'échantillon mensuel de l'ENV. En 2010, moins de 3 % des Canadiens ont voyagé à l'étranger (sauf aux États-Unis) au cours d'un mois donné. De même, les données de l'EVRC de novembre 2013 montrent qu'environ 30 % des Canadiens ont voyagé au Canada au cours de ce mois. Dans un effort en vue de sélectionner un plus grand nombre de voyageurs internationaux dans l'échantillon, une variable de passeport est calculée, afin d'identifier les ménages dans lesquels au moins un adulte est titulaire d'un passeport canadien valide. Seuls les titulaires de passeport adultes sont pris en compte parce que la population cible de l'ENV est constituée des résidents canadiens âgés de 18 ans et plus.

Étant donné qu'un passeport est requis pour les voyages internationaux, on espère que la probabilité d'échantillonner des voyageurs internationaux (et peut-être aussi des voyageurs canadiens) soit plus élevée pour les ménages qui comptent au moins un titulaire de passeport d'âge adulte.

Une étude pilote de l'ENV a été menée en février et mars 2016, afin de vérifier la réponse au questionnaire électronique, ainsi que toutes les procédures opérationnelles. Pour cette étude pilote, Passeport Canada a fourni un échantillon aléatoire simple d'un million d'enregistrements de Canadiens titulaires d'un passeport valide (à partir de son fichier comptant environ 23 millions d'enregistrements). Pour l'ENV principale en 2018, Statistique Canada prévoit avoir un fichier complet de tous les titulaires de passeport canadien. Le fichier d'échantillon de Passeport Canada a été apparié aux logements de la base de sondage (c.-à-d. le FUL), puis on a créé un indicateur de passeport pour identifier les ménages comptant au moins un titulaire de passeport d'âge adulte. Cet indicateur a été fixé à 1 s'il y avait au moins un titulaire de passeport d'âge adulte vivant dans un logement donné, et à 0 autrement. Cette stratégie de stratification permet d'augmenter la probabilité de compter des voyageurs internationaux dans l'échantillon de l'ENV. Tout comme le FMP T1, la qualité du fichier de Passeport Canada peut être évaluée au moyen d'un couplage d'enregistrements avec le fichier d'échantillon. Pour l'échantillon d'un million d'enregistrements utilisés dans l'ENV pilote, environ 85 % des enregistrements d'adultes peuvent être mis en correspondance avec la base de sondage de l'ENV. Les variables d'appariement utilisées pour ce couplage étaient les suivantes : AR_UID, prénom, nom de famille, sexe et date de naissance (Yang, 2015).

2.5 Résultats préliminaires de l'étude pilote

Conformément à la description de la section 2.4, l'étude pilote de l'ENV a été stratifiée au moyen de l'indicateur de passeport. Environ la moitié de l'échantillon de l'étude pilote a été affectée à des strates dans lesquelles cet indicateur désignait la présence d'un titulaire de passeport d'âge adulte dans le ménage. Pour la partie de l'échantillon où l'indicateur de passeport était de 0, 9,3 % des répondants ont déclaré au moins un voyage international, et 31,8 %, au moins un voyage au Canada. Pour la partie de l'échantillon où l'indicateur de passeport était de 1, 11,9 % des répondants ont déclaré au moins un voyage international, et 35,6 % ont déclaré au moins un voyage au Canada. Comme il fallait s'y attendre, la proportion de voyageurs, tant internationaux que canadiens, semble être légèrement plus grande dans les ménages où un adulte est titulaire d'un passeport. Les répercussions sont probablement amoindries par le fait que les strates où l'étiquette de passeport est égale à 0 ne signifie pas qu'il n'y a pas de membre adulte du ménage qui est titulaire d'un passeport valide, mais plutôt qu'aucun adulte du ménage n'a été apparié avec l'échantillon d'un million d'enregistrements fourni par Passeport Canada. Un échantillon de moins de 5 % du fichier de tous les titulaires de passeport canadien a servi à la stratification, ce qui signifie que peut-être 95 % des titulaires de passeport vivent dans des logements où l'indicateur de passeport dérivé est de 0. Cela signifie aussi que la majorité des logements où l'indicateur de passeport est de 0 devraient être remplacés par une valeur de 1. Pour toutes les raisons indiquées, les différences mentionnées ci-dessus devraient être plus importantes si l'on utilise le fichier complet des passeports, et encore plus pour les voyageurs internationaux.

2.6 Sources possibles de données administratives en vue d'améliorations futures

L'ASFC a un fichier administratif pour la grande majorité des Canadiens qui sont entrés au Canada à tous les postes frontaliers. Si Statistique Canada peut obtenir ce fichier administratif, les Canadiens qui reviennent d'un voyage international pourront être identifiés plus facilement, et on n'aura pas besoin du fichier de Passeport Canada. Ainsi, le plan d'échantillonnage peut être amélioré du point de vue de la stratification dans le cas des voyageurs internationaux. Idéalement, Statistique Canada aimerait obtenir ce fichier, étant donné que cela améliorerait considérablement l'efficacité de l'échantillonnage pour le Programme de la statistique du tourisme.

Parmi les autres sources de données administratives possibles qui peuvent aider à améliorer le Programme de la statistique du tourisme figurent les données de cartes de crédit au niveau de la transaction. À l'heure actuelle, Statistique Canada n'a pas accès à ces données, mais aimerait explorer la possibilité de les obtenir et examiner les améliorations possibles qu'elles apporteraient au Programme de la statistique du tourisme. Par exemple, les modèles de dépenses des personnes ayant des caractéristiques similaires à celles de répondants seraient utiles pour imputer les dépenses non déclarées. Il serait aussi possible d'identifier les voyageurs, selon le lieu des transactions, si ces données étaient disponibles. Un examen plus approfondi des utilisations possibles et de la disponibilité des données des cartes de crédit doit être effectué, mais peu importe l'utilisation prévue, Statistique Canada respectera la *Loi sur la protection des renseignements personnels* et s'engage à protéger les droits et la vie privée des personnes, ainsi que leurs renseignements personnels.

3. Conclusion

L'élaboration de l'ENV répond à l'objectif de l'AOB de regrouper les processus et d'uniformiser les systèmes, afin de réaliser des économies, tout en maintenant des normes élevées de qualité. L'ENV comporte des améliorations par rapport aux enquêtes existantes du Programme de la statistique du tourisme, grâce à l'utilisation d'un plus grand nombre de données administratives (comme les dossiers fiscaux et un fichier administratif fourni par Passeport Canada), l'utilisation d'outils communs, la mise en œuvre d'un questionnaire électronique pour la collecte des données et l'intégration d'un logiciel généralisé pour procéder à l'échantillonnage et à l'estimation. Il existe de nombreuses utilisations possibles des sources de données administratives dans les enquêtes sur le tourisme comme l'ENV. Le plan de sondage peut être amélioré du point de vue du processus de stratification, grâce aux données de Passeport Canada, de l'ASFC, et de l'Agence du revenu du Canada. Les dénombrements aux postes frontaliers fournis par l'ASFC, de même que les totaux démographiques projetés, peuvent servir au processus de calage, afin d'améliorer la qualité des estimations. D'autres sources de données administratives restent à explorer, comme le fichier des passages frontaliers et les données de cartes de crédit au niveau de la transaction, et elles présentent un grand potentiel du point de vue de l'amélioration des enquêtes sur les voyages à Statistique Canada. On examinera la faisabilité de l'utilisation de ces fichiers de données administratives à l'avenir, afin d'améliorer encore davantage le Programme de la statistique du tourisme.

Bibliographie

Statistics Canada (2016), International Travel Survey: Frontier Counts (ITS), available at: <http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5005&lang=en&db=imdb&adm=8&dis=2> (accessed May 19, 2015)

Yang, Ming Jie (2015), "NTS Record Linkage Project", unpublished report, Ottawa, Canada: Statistics Canada.