

RECUEILLIR DES DONNÉES SUR L'UTILISATION DE VÉHICULES – L'EXPÉRIENCE DE L'ENQUÊTE SUR LES VÉHICULES AU CANADA

Adam Wroński¹

RÉSUMÉ

Lancée en 1999, l'Enquête sur les véhicules au Canada (EVC) vise à recueillir des renseignements sur l'utilisation des véhicules automobiles immatriculés au Canada. La population cible de l'EVC comprend tous les véhicules routiers (à l'exception du matériel spécial, des remorques et des motocyclettes) immatriculés au Canada. Chaque trimestre, on prélève un échantillon de véhicules et on utilise un carnet de bord de sept jours pour recueillir des renseignements détaillés sur l'utilisation des véhicules. Le carnet comprend des questions sur le kilométrage, le nombre de passagers, les caractéristiques du véhicule, l'objet et la durée des déplacements, les caractéristiques démographiques du conducteur et des passagers et la consommation de carburant. Comme il s'agit d'une enquête à participation volontaire et qu'il faut sept jours pour remplir le carnet, tout est mis en œuvre pour obtenir un taux de réponse suffisant et éviter les erreurs de réponse. Dans la première partie, nous décrivons le plan d'enquête et les méthodes de collecte des données, de vérification et d'imputation. Puis, nous expliquons les difficultés liées aux différentes étapes de l'enquête. Enfin, nous présentons les résultats d'une recherche menée afin de réduire au minimum l'incidence des erreurs non dues à l'échantillonnage.

MOTS-CLÉS : Erreurs non dues à l'échantillonnage, Statistiques de transports, Utilisation de véhicules.

1. INTRODUCTION

Jusqu'à récemment, en l'absence de mesures régulières de l'activité des véhicules routiers, les statistiques sur l'activité du secteur des transports au Canada étaient incomplètes. Même si les véhicules routiers dominent le transport des passagers et des marchandises, on ne disposait d'aucune mesure du nombre total de véhicules-kilomètres, de passagers-kilomètres ni de tonnes-kilomètres.

L'Enquête sur les véhicules au Canada (EVC), qui a commencé à recueillir des données au premier trimestre de 1999, a été élaborée à la demande de Transports Canada pour combler cette lacune statistique. Elle vise à recueillir des renseignements sur l'utilisation des véhicules routiers immatriculés au Canada, qui constituent la population cible de l'enquête. Au début de chaque trimestre, on crée la base de sondage de l'EVC en prélevant un échantillon de véhicules dans les fichiers d'immatriculation des véhicules. La collecte des données se fait au moyen d'une interview téléphonique assistée par ordinateur (ITAO) et de questionnaires envoyés et retournés par la poste. L'ITAO sert à vérifier les coordonnées du répondant et à obtenir des renseignements sur diverses caractéristiques du véhicule. On utilise un carnet de bord de sept jours pour recueillir des renseignements détaillés sur l'utilisation du véhicule. Le carnet comprend des questions sur le kilométrage, le nombre de passagers, les caractéristiques du véhicule, l'objet et la durée de chaque déplacement, les caractéristiques démographiques du conducteur et des passagers et le carburant acheté.

¹ Statistique Canada, bureau 11-O, Immeuble R.H. Coats, Pré Tunney, Ottawa (Ontario) K1A 0T6, Canada, wronada@statcan.ca

Le plan d'enquête de l'ECV ne ressemble à aucun autre. Comme il s'agit d'une enquête à participation volontaire, elle s'accompagne d'une série de difficultés inédites. Dans la présente communication, nous entendons nous concentrer sur les caractéristiques distinctes de l'EVC, montrer leur incidence et présenter la recherche liée à ces difficultés. Dans la première partie, nous présentons une brève description du plan d'enquête en soulignant les enjeux propres à l'EVC. Dans la seconde partie, nous abordons certains aspects de la qualité qui préoccupent les spécialistes des méthodes d'enquête. Puis, nous présentons les résultats de certaines études axées sur ces aspects de la qualité. Enfin, nous formulons nos conclusions.

2. LE PLAN D'ENQUÊTE

2.1. Description générale

La population cible de l'EVC comprend tous les véhicules automobiles immatriculés au Canada, à l'exception des remorques, des motocyclettes, des véhicules hors route (motoneiges, motodunes, véhicules amphibies, etc.) et du matériel spécial (grues, balayeuses mécaniques, chasse-neige et rétrocaveuses, par exemple).

L'objectif principal de l'enquête consiste à produire, à l'échelle du Canada, des estimations annuelles et trimestrielles du nombre de véhicules-kilomètres et de passagers-kilomètres pour tous les véhicules compris dans le champ de l'enquête, par type de véhicule. Dans la mesure du possible, on calcule aussi des estimations provinciales. L'enquête produit des estimations annuelles et trimestrielles réparties selon le type de véhicule, l'âge et le sexe du conducteur et le moment de la journée.

Les résultats constituent la principale source de renseignements sur l'utilisation des véhicules routiers à l'intention des chercheurs, des analystes des politiques et les membres du public que ces questions intéressent.

Lorsqu'on a défini le plan d'enquête, on a prévu qu'un fardeau important aurait une incidence négative sur le taux de réponse. Il faut sept jours pour remplir le carnet, ce qui exige une certaine discipline de la part du répondant.

2.2. Base de sondage et plan d'échantillonnage

La base de sondage se compose des fichiers d'immatriculation de véhicules des treize secteurs de compétence (dix provinces et trois territoires), qui sont fournis chaque trimestre pour les fins de l'EVC.

Comme les fichiers d'immatriculation sont différents d'un secteur de compétence à l'autre, on traite chaque fichier de telle sorte qu'on puisse extraire des données et les fusionner avec celles d'autres fichiers. Le traitement permet de repérer les véhicules dans le champ de l'enquête, les immatriculations périmées et les dédoublements. En présence de plusieurs enregistrements valides pour un seul véhicule, il est particulièrement difficile d'établir à quel secteur de compétence il faut rattacher le véhicule.

Parfois, les fichiers d'immatriculation arrivent en retard, sont créés trop tôt ou n'arrivent pas du tout. Il s'agit évidemment d'une source potentielle d'erreurs de couverture.

On répartit les véhicules en quatre types. On distingue d'abord les autobus. Puis, on répartit les autres véhicules en trois types, selon leur poids brut : moins de 4 500 kg, de 4 500 kg à 15 000 kg et 15 000 kg et plus, qu'on appelle respectivement véhicules légers, camions et camions de « catégorie 8 ».

Le budget de l'enquête permet d'échantillonner au total, par trimestre, environ 5 000 unités dans les provinces et 2 500 unités dans les territoires. On utilise un plan d'échantillonnage à deux degrés. Chaque trimestre, trois mois avant la période de référence, on prélève un échantillon de véhicules (premier degré) puis, à chaque véhicule choisi dans l'échantillon, on attribue de façon aléatoire une période de sept jours de déclaration (deuxième degré).

On construit les strates en fonction du secteur de compétence, du type et de l'âge des véhicules. Au sein d'un même secteur de compétence, tous les véhicules d'un même type sont répartis en deux strates d'âge (anciens et récents) selon l'année du modèle. On établit le seuil de démarcation entre les véhicules anciens et récents en réduisant au minimum la variance de l'estimation du nombre de véhicules-kilomètres par province et par type. Chaque année, on rajuste ce seuil de démarcation.

Au premier degré, on répartit par la racine cubique le nombre de véhicules d'une province pour déterminer le nombre d'unités échantillonnées dans chaque province. Puis, on détermine le nombre d'unités échantillonnées par type de véhicule dans chaque province en appliquant encore une fois la règle de répartition par la racine cubique. Cette règle a été établie arbitrairement, à la demande de la clientèle, afin de calculer des estimations de base dans le cas des petits secteurs de compétence. On trouve empiriquement la répartition optimale entre les deux strates d'âge de véhicules en fonction de la dernière année pour laquelle on dispose de données. La répartition tient compte d'une taille minimale de l'échantillon dans chaque strate. Puis, on prélève systématiquement, par code postal, l'échantillon de véhicules dans le fichier d'immatriculation. On procède ainsi pour éviter de choisir un nombre élevé de véhicules de grandes entreprises pour un même trimestre et pour s'assurer que les véhicules choisis dans les petites provinces soient diversifiés. En vertu d'une autre mesure visant à réduire le fardeau de réponse, tout véhicule choisi reste dans l'échantillon pendant un seul trimestre et hors de l'échantillon pendant les trois trimestres suivants.

Au deuxième degré, on attribue de façon aléatoire, à chaque véhicule choisi, une période de sept jours de déclaration de l'utilisation du véhicule. Dans chaque strate, le premier jour de déclaration est réparti uniformément sur le trimestre (répartition systématique) pour que le nombre de réponses soit uniforme dans le temps. Ainsi, environ le même nombre de répondants commencent à remplir le carnet chaque jour de la semaine. Dans le cas des territoires, l'échantillonnage est moins complexe car les exigences en matière de données sont moindres.

2.3. Collecte des données

La collecte des données d'enquête a commencé le 1^{er} février 1999. Seuls huit fichiers d'immatriculation de véhicules ont été reçus assez tôt pour être compris dans l'échantillon. Depuis le 1^{er} octobre 1999, l'enquête porte sur des véhicules de toutes les provinces et de tous les territoires.

Dans le cas des provinces, la collecte des données s'effectue en deux temps. Tout d'abord, on mène une interview téléphonique assistée par ordinateur (ITAO) auprès des propriétaires des véhicules immatriculés choisis. Cette interview sert à recueillir des renseignements généraux sur le véhicule (état actuel, etc.) et à demander au répondant de remplir un carnet de bord. Puis, on lui envoie le carnet de bord par la poste. S'il est impossible de rejoindre le répondant par téléphone, on lui envoie par la poste le carnet de bord accompagné d'un bref questionnaire pour recueillir certains renseignements ordinairement recueillis durant l'ITAO.

Un carnet de bord de sept jours, propre à chaque type de véhicule, sert à recueillir des renseignements détaillés sur l'utilisation du véhicule. Le carnet comprend des questions sur le kilométrage, le type de carrosserie du véhicule, le type de carburant, l'objet et la durée des déplacements, le nombre de passagers, les caractéristiques démographiques du conducteur et des passagers et le carburant acheté, ainsi que des renseignements sur le transport de marchandises dangereuses. Dans le cas des véhicules de plus de 4 500 kg, on pose des questions supplémentaires sur la configuration du camion (type de remorque, etc.).

Dans le cas des territoires, la collecte des données consiste à envoyer aux répondants deux cartes postales, l'une au début, l'autre à la fin du trimestre. La première carte postale demande au répondant d'inscrire la lecture d'odomètre au début de la première journée du trimestre et de répondre à des questions sur les caractéristiques du véhicule. À tous les répondants qui retournent la première carte postale, on envoie une deuxième carte postale leur demandant d'inscrire la lecture d'odomètre au début du premier jour du trimestre suivant. Ces deux lectures d'odomètre permettent de calculer la distance parcourue par le véhicule au cours du trimestre.

Comme il s'agit d'une enquête à participation volontaire et qu'il faut sept jours pour remplir le carnet, tout est mis en œuvre pour obtenir un taux de réponse suffisant et éviter les erreurs de réponse. D'abord, toutes les adresses des propriétaires de véhicules échantillonnés sont mises à jour d'après une base de données fédérale qui retrace les changements de propriété de véhicules. Puis, on cherche les numéros de téléphone en consultant les annuaires téléphoniques informatisés et en effectuant une recherche manuelle. Ensuite, pour chacune des unités choisies, on procède à l'ITAO environ deux semaines avant la période désignée. Au cours de cette interview, on demande au répondant le nombre approximatif de kilomètres parcourus au cours de la semaine précédant l'interview. Si le répondant refuse de remplir le carnet, on offre de lui envoyer une version abrégée du questionnaire qui sert à recueillir uniquement les lectures d'odomètre. La deuxième ITAO a lieu le premier jour de la période désignée pour la déclaration de sept jours. On demande au répondant de commencer à remplir le carnet et, le cas échéant, de poser des questions à ce sujet. La troisième ITAO est une interview de suivi pour rappeler au répondant de retourner le carnet par la poste et pour obtenir des données auprès des non-répondants en offrant de leur envoyer la version abrégée du questionnaire.

Pour accroître le taux de réponse dans le cas des véhicules dont on n'a pas rejoint le propriétaire par ITAO dans un délai d'une semaine, on reporte de sept jours la période de déclaration, jusqu'à concurrence de quatre reports. Aux répondants qu'on n'a pas rejoints après quatre semaines et à ceux dont on n'a pas trouvé le numéro de téléphone, on envoie le carnet par la poste. Les répondants qui ont oublié de commencer à remplir le carnet sont encouragés à le faire une ou deux semaines plus tard.

2.4. Traitement des données et estimation

Le traitement des données commence par la consolidation des renseignements recueillis grâce aux trois ITAO, aux questionnaires abrégés et aux carnets propres à chaque type de véhicule. Tous les renseignements sont fusionnés dans une base de données qui comprend deux volets : les renseignements sur le véhicule et les données sur l'utilisation du véhicule (déplacements au cours de la période de déclaration).

Puis, on effectue des vérifications et on signale les données incohérentes au moyen d'un drapeau. On procède ensuite à l'imputation. Au cours de ce processus, on remplace toutes les données erronées et on impute toutes les réponses partielles concernant les déplacements. Une réponse est considérée comme partielle lorsqu'on dispose du kilométrage sur sept jours d'après les lectures d'odomètre ou l'approximation fournie par le répondant lors de l'ITAO.

Au moyen de l'algorithme d'imputation, on cherche à calculer les éléments manquants en utilisant d'abord d'autres renseignements obtenus du même répondant, puis les données concernant des véhicules semblables sur le fichier en ce qui concerne le kilométrage, les caractéristiques géographiques et celles du véhicule, le sexe et l'âge du conducteur principal, le jour de la semaine et le moment de la journée. On crée un vaste ensemble de rapports post-imputation pour vérifier les résultats. Au terme du processus, à chaque véhicule choisi correspondent sept jours de déplacements, c'est-à-dire sept véhicules-jours. Pour les fins de l'estimation, les déplacements sur deux jours sont scindés à minuit.

Des essais portant sur les observations relatives aux véhicules-jours (essais à comparaisons multiples F et Waller-Duncan) ont montré que l'utilisation variait considérablement entre les jours ouvrables et les jours de congé. Dans la composante provinciale de l'enquête, les véhicules-jours sont donc stratifiés a posteriori en jours de congé et en jours ouvrables. On utilise le calage en fonction du nombre de véhicules-jours dans chaque strate établie a posteriori :

$$\sum_{i \in s_h} \sum_{k \in s_i^{(W)}} a_i a_k g_k^{(W)} = N_h M^{(W)}$$

$$\sum_{i \in s_h} \sum_{k \in s_i^{(H)}} a_i a_k g_k^{(H)} = N_h M^{(H)}$$

Dans cette équation, N_h est le nombre de véhicules compris dans la strate h ; c'est la moyenne des dénombrements effectués d'après les fichiers d'immatriculation au début et à la fin de la période de référence. $M^{(W)}$ et $M^{(H)}$ sont les nombres de jours compris dans les strates établies a posteriori des jours ouvrables et des jours de congé au cours de la période de référence. Les valeurs g_k sont les facteurs de

calage. Le poids de deuxième degré pour le k^{e} véhicule-jour est $a_k = M_i / m_i$, où M_i est le nombre de jours compris dans la période de référence et m_i est le nombre de jours compris dans la période de référence pour lesquels les données sont disponibles. Le poids de premier degré a_i est fondé sur quatre catégories de réponse possibles à l'intérieur de la strate h : n_{hr} – répondants dans le champ de l'enquête (comprend les enregistrements imputés), n_{hs} – non-répondants dont on a établi qu'ils sont dans le champ de l'enquête, n_{ho} – répondants hors du champ de l'enquête, et véhicules pour lesquels on ne dispose d'aucun renseignement. Pour le i^{e} répondant, $a_i = N_h / (n_{ho} + n_{hr} + n_{hs}) \times (n_{hr} + n_{hs}) / n_{hr}$; pour les non-répondants, $a_i = 0$ et pour les répondants hors du champ de l'enquête, $a_i = N_h / (n_{ho} + n_{hr} + n_{hs})$. La pondération répétée suppose que les caractéristiques statistiques des répondants dans le champ de l'enquête sont représentatives des non-répondants qui sont effectivement dans le champ de l'enquête. Aux véhicules pour lesquels on ne dispose d'aucun renseignement, on attribue les mêmes caractéristiques statistiques qu'à tous les véhicules qui sont effectivement dans le champ ou hors du champ de l'enquête.

Ce calage assure que chaque véhicule-jour déclaré à l'intérieur d'une strate compte pour une part égale dans les estimations. Les estimations des dénombrements en fonction des caractéristiques des véhicules et toutes les estimations relatives aux territoires sont fondées sur un plan d'échantillonnage simple à un seul degré.

L'EVC cherche également à intégrer les taux d'imputation dans ses calculs de la précision des estimations. Chaque trimestre, on effectue une simulation pour tenir compte de l'incidence de l'imputation. Dans une population constituée des données obtenues auprès des répondants, on prélève un échantillon initial, puis on prélève des sous-échantillons pour représenter les répondants et les non-répondants compris dans l'échantillon initial. Puis, on impute les non-répondants selon la stratégie d'imputation de l'EVC. On calcule les estimations de coefficients de variation (c.v.) de l'échantillon initial et de l'échantillon initial avec imputation des non-répondants. Cette méthode nous permet de visualiser la variation du c.v. en fonction de l'importance de l'imputation. On utilise alors un modèle de régression linéaire pour quantifier la dépendance du c.v. par rapport au taux d'imputation pour estimer le c.v. avec imputation de la non-réponse.

3. LES DIFFICULTÉS

Les résultats trimestriels étaient à la disposition du public dès le quatrième trimestre de 1999. La plupart des estimations étaient très proches des approximations éclairées des spécialistes, d'où le risque d'une confiance excessive dans les estimations de l'enquête qui peuvent être biaisées à cause d'erreurs non dues à l'échantillonnage. Comme on ne dispose pas de données administratives ou statistiques auxquelles on pourrait comparer les résultats de l'EVC, rien n'indique qu'il peut exister des biais. Les erreurs non dues à l'échantillonnage peuvent avoir une forte incidence sur les estimations; par conséquent, elles constituent aujourd'hui la principale difficulté pour les méthodologistes de l'EVC.

La principale source de toutes les difficultés tient au faible taux de réponse. Pour l'année 2000, environ 38 % seulement des véhicules-jours déclarés contiennent tous les éléments nécessaires sans défauts ou avec seulement des défauts mineurs (réponse complète). Comme la méthode de collecte actuelle est très complexe et très coûteuse (plus de 50 % du coût total de l'enquête) et qu'il ne peut y avoir d'autres suivis ni de mesures d'encouragement des répondants, on n'a trouvé aucune recette pour améliorer le taux de réponse.

La difficulté est d'autant plus grande que pour près de la moitié des jours (de la période de référence) à l'égard desquels on obtient une réponse complète, les véhicules n'étaient pas utilisés. En outre, 30 % des véhicules-jours déclarés nécessitent une imputation importante (réponse partielle) parce que les données déclarées sont incohérentes ou qu'on dispose uniquement de l'approximation fournie par le répondant en ce qui concerne le nombre de kilomètres parcourus durant la semaine précédant l'ITAO. Pour la plupart des estimations concernant les caractéristiques des déplacements, le taux d'imputation (partie imputée de

l'estimation) se situe entre 40 % et 60 %. Au niveau des strates, les dénombrements de véhicules et les estimations du kilométrage présentent des taux d'imputation de 0 % à 6 %. Il est donc essentiel que la méthode d'imputation ne crée aucun biais.

Une source potentielle de biais dû à l'imputation est l'approximation fournie par le répondant en ce qui concerne le nombre de kilomètres parcourus durant la semaine précédant l'ITAO. À l'étape de l'imputation, on utilise beaucoup ce renseignement pour établir le nombre approximatif de kilomètres parcourus durant la semaine. On n'a pas encore effectué d'étude comparative des approximations fournies par les répondants (données de rappel) et des lectures d'odomètre pour en évaluer la qualité. Comme on recueille les données d'une semaine précise (et non d'une semaine moyenne), on suppose uniquement que les approximations fournies par les répondants présentent la même distribution que celle des lectures d'odomètre. Les approximations fournies par les répondants sont corrigées par le ratio de la moyenne de ces approximations à la moyenne des lectures d'odomètre dans le cas des véhicules pour lesquels on dispose des deux mesures. Lorsqu'on calcule ces moyennes, on prend le type de véhicule, la province, les caractéristiques démographiques du conducteur (si les répondants sont assez nombreux) pour fonder la correction sur des véhicules semblables. Mais quelle est l'incidence de cette méthode sur le biais?

4. LA RECHERCHE ET SES RÉSULTATS

On a effectué des études supplémentaires pour résoudre ces difficultés.

4.1. Étude de la méthodologie de collecte

Cette étude visait à trouver des méthodes de collecte moins coûteuses que la méthode actuelle. Pour le deuxième trimestre de 2000, au lieu de l'échantillon habituel d'environ 5 000 véhicules, on a prélevé un échantillon de 8 000 véhicules qu'on a réparti de façon aléatoire en deux échantillons distincts, selon les règles suivantes : chacun des deux échantillons devait avoir à peu près la même taille au sein de chaque strate et toutes les unités ayant le même code postal à six caractères devaient être regroupées dans l'un des échantillons. On a adopté ces règles pour éviter de communiquer plus d'une fois avec les répondants dont plusieurs véhicules ont été choisis.

Pour l'un des échantillons, on a ensuite recueilli les données selon la méthode de collecte actuelle de l'EVC. Dans le cas de l'autre échantillon, au lieu de mener l'ITAO habituelle, on a envoyé directement les questionnaires aux répondants par la poste, sans contact préalable et sans ITAO de suivi.

Pour les fins de notre étude, les renseignements recueillis auprès des deux échantillons ont été vérifiés et imputés séparément. Puis, pour chaque méthode, on a procédé à l'estimation et on a comparé les résultats.

Dans le cas de la collecte effectuée par envoi postal, le taux de réponse est inférieur à la moitié du taux de réponse obtenu selon la méthode habituelle (environ 20 % des répondants dans l'échantillon prélevé ont fourni tous les renseignements demandés et 10 % ont fourni des réponses partielles). Un taux de réponse aussi faible risque de se traduire par des estimations biaisées.

L'analyse des réponses révèle que si l'on n'effectue pas d'ITAO, le profil des répondants s'en trouve modifié comme suit :

- Il y a moins de réponses indiquant que le véhicule n'est pas utilisé. Dans l'échantillon enquêté selon la méthode actuelle, 24,3 % des dates présentent une valeur nulle pour le nombre de kilomètres parcourus par jour, contre seulement 20,9 % dans le cas l'échantillon enquêté selon la méthode de l'envoi postal.
- Selon la méthode actuelle, on recueille davantage de réponses auprès des propriétaires qui font une grande utilisation de leurs véhicules.
- En l'absence de l'ITAO, on obtient des estimations beaucoup plus faibles des déplacement-kilomètres chez les jeunes conducteurs (âgés de moins de 25 ans).

Le fait de mener l'ITAO permet de compenser (du moins en partie) la sous-représentation des véhicules très utilisés, des jeunes répondants et des véhicules non utilisés. Dans le cas de la collecte effectuée par envoi postal, on obtient moins de renseignements permettant de procéder à l'imputation.

De plus, dans les petites provinces, une augmentation de la taille de l'échantillon alourdirait le fardeau de réponse des propriétaires de camions et d'autobus. On savait déjà que certains répondants trouvaient excessif le nombre de questionnaires reçus de l'EVC. En élargissant l'échantillon, on n'obtiendrait donc pas un nombre de réponses proportionnellement plus élevé.

Aussi l'étude a-t-elle permis de conclure que la collecte par envoi postal ne pouvait qu'avoir une incidence négative sur la qualité.

4.2. Analyse de l'incidence de l'imputation

En comparant les données imputées aux données non imputées, on a découvert que les déplacements imputés étaient habituellement plus longs que les déplacements déclarés et que les véhicules pour lesquels on dispose uniquement de l'approximation fournie par le répondant enregistraient davantage de déplacements. Les écarts sont présentés dans la figure 1 et dans le tableau 1. Le tableau 2 ci-dessous montre les aspects à l'égard desquels l'imputation a la plus forte incidence sur la distance moyenne parcourue.

Figure 1. Distribution de la durée des déplacements selon la méthode de collecte

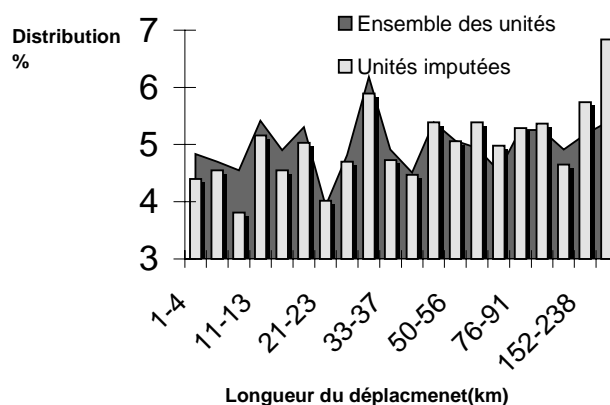


Tableau 1. Distance moyenne parcourue selon le type d'imputation

| Type d'imputation | Par journée | Par déplacement |
|--|-------------|-----------------|
| Aucune (réponse complète) | 59 km | 30 km |
| Partielle (réponse partielle) | 63 km | 25 km |
| Total (d'après l'approximation fournie par le répondant) | 82 km | 42 km |

Tableau 2. Aspects à l'égard desquels la distance moyenne parcourue par jour présente de grands écarts entre les données déclarées et les données imputées

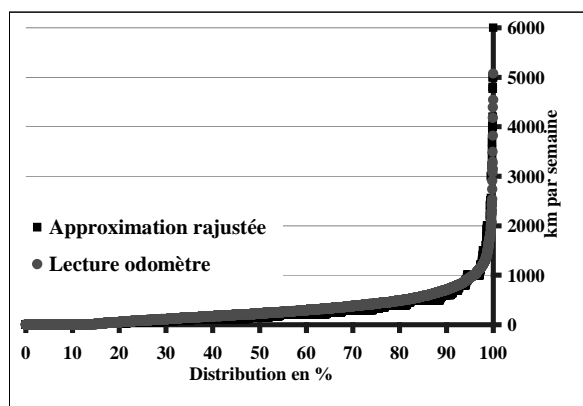
| Aspect | Distance moyenne | |
|--|------------------|---------|
| | Déclarée | Imputée |
| Conducteur masculin | 25 km | 33 km |
| Conducteur âgé de moins de 20 ans | 21 km | 29 km |
| Conducteur âgé de plus de 85 ans | 35 km | 14 km |
| Déplacements effectués entre minuit et 6 h | 33 km | 95 km |

L'étude a donc confirmé que les données imputées pouvaient différer considérablement des données déclarées.

4.3. Étude de l'approximation fournie par le répondant

À l'étape de l'imputation, on utilise beaucoup l'approximation fournie par le répondant en ce qui concerne la distance parcourue au cours de la dernière semaine. Il est donc nécessaire de vérifier l'hypothèse selon laquelle la distribution de ces approximations peut remplacer celle des lectures d'odomètre. Dans le cas des répondants qui fournissaient les deux renseignements, on a comparé ces distributions par type de véhicule (d'après les données de 2000). Pour chaque type de véhicule, la corrélation entre les deux distributions était supérieure à 0,989. La figure 2 montre les distributions comparées en ce qui concerne les véhicules d'un poids brut de moins de 4 500 kg en Ontario. Comme, en moyenne, l'approximation et la lecture d'odomètre peuvent différer dans une proportion pouvant atteindre 20 %, on applique le ratio des moyennes pour rajuster l'approximation avant de l'utiliser à l'étape de l'imputation (voir la section 3). Lors d'essais portant sur d'autres types de véhicules et sur les provinces, les distributions ont présenté un comportement semblable. À l'étape de l'imputation, on calcule le rajustement et on l'applique au niveau de la catégorie d'imputation; il faut donc approfondir l'étude à ce niveau.

Figure 2. Distributions comparées des lectures d'odomètre et des approximations fournies par les répondants – véhicules légers en Ontario.



4.4 Étude « Air Care »

Les responsables du programme « Air Care » mené dans une région de la Colombie-Britannique nous ont fourni leur base de données des lectures d'odomètre annuelles. Chaque année, le programme procède à un essai portant sur les véhicules qui sont « faiblement utilisés » depuis plus de deux ans. Nous avons donc pu comparer une estimation fondée sur les données « Air Care » à l'estimation fondée sur l'EVC de 1999 pour la même région (en fonction de 388 réponses). La région visée par le programme « Air Care » englobe environ 6 % de la population de l'EVC, soit environ 1,04 million de véhicules.

On a constaté que la base de données « Air Care » présentait un taux d'erreur élevé dans les lectures d'odomètre. Il y avait des écarts négatifs nombreux et des écarts positifs extrêmes entre deux lectures annuelles consécutives. Les figures 3 et 4 montrent la distribution des écarts dans les lectures d'odomètre « Air Care » et un détail de cette distribution qui présente un nombre inhabituel d'observations aux alentours de 100 000 kilomètres par année. Ainsi, dans la figure 4, le nombre de véhicules observés qui sont noirs peut être considéré comme inexact en raison, par exemple, de l'ajout ou de l'omission d'un chiffre au moment de la lecture d'odomètre, d'où une source potentielle de ce genre d'erreur.

Il a donc fallu décider quelles observations du programme « Air Care » étaient erronées. Sur les conseils de spécialistes, on a calculé les estimations avec des seuils d'exclusion qui se situaient entre 80 000 km et 100 000 km. On a également calculé les estimations selon la disponibilité des lectures d'odomètre de 1998, 1999 et 2000.

Selon le choix des observations « Air Care » considérées comme erronées, les estimations des distances parcourues se situent entre 15,9 milliards et 17,7 milliards de kilomètres. Le nombre d'observations sur

lequel elles sont fondées varie entre 531 000 et 895 000 véhicules. Les c.v. sont inférieurs à 0,3 %. L'estimation de l'EVC est de 15,9 milliards de kilomètres, avec un c.v. de 6,3 %. Les estimations « Air Care » diffèrent donc des estimations de l'EVC dans une proportion de 0 % à 12 %

Figure 3. Distribution des écarts dans les lectures d'odomètre « Air Care » rajustés en fonction de 365 jours

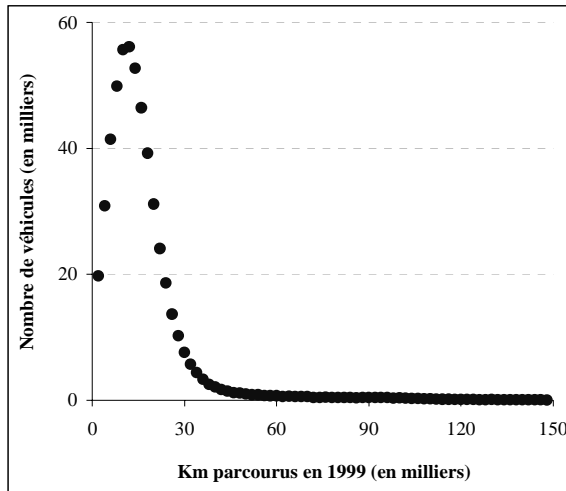
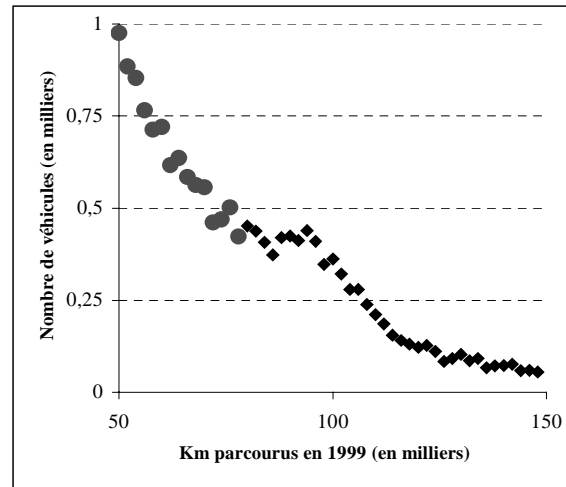


Figure 4. Détail de la distribution « Air Care »



5. CONCLUSIONS

L'EVC est une nouvelle enquête dont la méthodologie est unique. Les études d'évaluation n'ont fait ressortir aucun problème important, mais il faut approfondir la recherche pour améliorer les taux de réponse et d'imputation en tenant compte des contraintes de l'enquête au chapitre des ressources humaines et financières. De plus, pour pouvoir établir des comparaisons plus complètes, il faut prévoir de nouvelles initiatives en matière de données administratives (surtout de la part des provinces).

D'autres aspects doivent retenir l'attention et faire l'objet d'une étude approfondie :

- l'incidence de la non-réponse sur la qualité des estimations; par exemple, la corrélation entre l'utilisation intensive et la non-réponse;
- l'incidence de l'imputation sur la qualité des estimations;
- l'utilisation des approximations fournies par les répondants à l'étape de l'imputation.

Enfin, même si les estimations de l'enquête semblent très convaincantes, il faut aussi prendre soin d'avertir les utilisateurs que les taux élevés de non-réponse et d'imputation peuvent avoir une incidence sur leur qualité.

BIBLIOGRAPHIE

Briant, N. (2001), « Estimation de la distance parcourue par les véhicules du Programme administratif Air Care », rapport inédit, Statistique Canada.

Christoff, W. et A. Wronski, (2000), « Enquête sur les véhicules au Canada », document électronique, Statistique Canada, http://www.statcan.ca/francais/freepub/53F0004XIF/free_hm.

Matthews, S. (2000), « Canadian Vehicle Survey – Summary of Response and Imputation Distribution Study Findings », rapport inédit, Statistique Canada.

Xiao, P. (2000), « Canadian Vehicle Survey – Summary Q2 2000 Study Estimations and Response Rates Comparison », rapport inédit, Statistique Canada.

Wronski A. (2000), « Canadian Vehicle Survey (CVS) – Quarter 2, 2000 Collection Study », rapport inédit, Statistique Canada.