

Article

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

Recherche sur la collecte des données à l'aide de paradonnées à Statistique Canada

par François Laflamme

2009



Recherche sur la collecte des données à l'aide de paradonnées à Statistique Canada

François Laflamme¹

Résumé

Ces dernières années, Statistique Canada a mené plusieurs études analytiques en se servant de paradonnées afin de mieux saisir les diverses questions liées au processus et aux pratiques de collecte des données. En particulier, ces études avaient pour but de mieux comprendre le déroulement du cycle de collecte des données, et d'identifier les possibilités stratégiques, d'évaluer de nouvelles initiatives de collecte afin d'améliorer la façon dont l'organisme mène et gère ses enquêtes. L'objectif essentiel de l'exposé est de présenter les principaux résultats de ces études passées et courantes décrivant l'utilisation des paradonnées à Statistique Canada. Les futurs projets de recherche axés sur l'élaboration de stratégies opérationnelles fiables susceptibles d'améliorer l'efficacité de la collecte ou la qualité des données sont également discutés.

Mots clés : Paradonnées, historique des transactions d'appel, plan de sondage adaptatif, productivité.

1. Introduction

Le défi de tout organisme statistique est de recueillir des données de grande qualité de manière rentable en dépit du grand nombre de facteurs influents, tels que la baisse des taux de réponse, l'évolution du comportement de la population et l'accroissement du fardeau de réponse. La collecte des données est définitivement un élément clé du processus d'enquête, car elle a une incidence directe sur la qualité et est une composante importante du coût de nombreux produits statistiques. Ces dernières années, Statistique Canada a passé en revue ses processus de collecte des données afin d'évaluer et de surveiller les pratiques courantes et les nouvelles initiatives de collecte des données, et de dégager les possibilités stratégiques d'amélioration. Les paradonnées ont été la pierre angulaire de ces études et continueront d'être utilisées à grande échelle dans les futurs travaux de recherche.

La plupart des travaux de recherche dont il est question dans cet article s'appuient sur l'historique des transactions d'appel Blaise pour de nombreuses enquêtes sociales et agricoles menées par ITAO. L'exposé a pour principaux objectifs de faire part des faits saillants de ces travaux de recherche, y compris certaines grandes réalisations, et de présenter les plans et les priorités des activités de recherche en cours et futures sur la collecte des données. L'article débute par un aperçu de la collecte des données et du contexte des paradonnées à Statistique Canada, suivi d'une description des principaux objectifs de l'étude. À la section 4, nous présentons les faits saillants des études antérieures fondées sur des paradonnées qui ont permis de mieux comprendre nombre de questions liées au processus de collecte des données et d'identifier d'éventuels domaines propices aux améliorations qui pourraient requérir des études plus approfondies. À la section 5, nous décrivons la recherche en cours sur la productivité et le coût des enquêtes, tandis qu'à la section 6, nous discutons de l'évaluation de nouvelles initiatives de collecte des données. À la section 7, nous donnons un aperçu de certaines réalisations importantes résultant de la recherche à l'aide de paradonnées, y compris l'élaboration initiale d'une stratégie fondée sur des plans de sondage adaptatifs. La dernière section décrit les futurs plans de recherche, qui viseront à cerner les stratégies opérationnelles viables susceptibles d'accroître l'efficacité de la collecte des données ou la qualité des données.

¹ François Laflamme, Statistique Canada, Division des méthodes d'enquête auprès des entreprises, immeuble Jean-Talon, 6^e étage, Ottawa (Ontario), K1A 0T6, francois.laflamme@statcan.gc.ca

2. Collecte des données et paradonnées à Statistique Canada

La présente section donne un aperçu du contexte organisationnel de la collecte des données et des paradonnées à Statistique Canada, y compris une brève description des sources de données et du contenu des entrepôts de paradonnées.

2.1 Organisation de la collecte des données

Pour les enquêtes sociales, agricoles et auprès des entreprises mensuelles menées par ITAO, la collecte des données est réalisée et gérée par les centres d'appel des six bureaux régionaux (BR) de Statistique Canada répartis à travers le pays. Le portefeuille d'enquêtes varie selon l'emplacement et, en moyenne, chaque centre d'ITAO d'un BR réalise de cinq à dix enquêtes durant un mois donné. En 2007, environ 1 000 intervieweurs spécialisés en ITAO ont effectué, en moyenne, 700 000 appels par mois.

2.2 Contexte des paradonnées

Le Système d'intégration des données et planification de la production (IDPP), qui est l'entrepôt de paradonnées de Statistique Canada, a été créé en vue d'intégrer et de normaliser les données opérationnelles provenant de trois sources aux fins d'analyse et de production des rapports du Système d'information de la gestion (SIG). Ces sources sont 1) les fichiers de l'historique des transactions Blaise (BTH pour *Blaise Transaction History*) pour les enquêtes par ITAO, 2) les fichiers de gestion des cas (fichiers CaseMan) pour les enquêtes par IPAO et 3) l'information du Système de paye des opérations des enquêtes statistiques (SPOES) pour toutes les enquêtes, y compris celles qui sont non standard. Le système IDPP contient les paradonnées provenant de la plupart des enquêtes réalisées par Statistique Canada depuis 2003 et est mis à jour quotidiennement afin d'inclure les données provenant des enquêtes en cours actives. En pratique, cette information est accessible le jour après que les paradonnées ont été recueillies ou enregistrées.

Un enregistrement d'historique des transactions d'appel correspond essentiellement à de l'information de production concernant l'ITAO (fichiers BTH) ainsi que l'IPAO (fichiers CaseMan). Un enregistrement est créé chaque fois qu'un cas est ouvert, pour la collecte de données ou pour d'autres raisons. Il contient des renseignements détaillés sur chaque visite ou appel effectué en vue de prendre contact avec chaque unité échantillonnée durant le processus de collecte des données. Il renferme l'information sur l'identification de l'enquête et du cas, la date, la durée de l'ouverture du cas, l'intervieweur qui s'est occupé du cas, le résultat de l'appel, ainsi que d'autres renseignements pertinents au sujet de chaque visite ou appel. Par ailleurs, le Système de paye des opérations des enquêtes statistiques (SPOES) contient des données administratives et financières sur les demandes de paye des intervieweurs pour toutes les activités de collecte. Un enregistrement est créé dans le SPOES chaque fois qu'un intervieweur soumet une demande pour une enquête particulière et une tâche particulière un jour donné, pour des activités de collecte directes (p. ex. interview, dépistage, etc.) ou pour d'autres activités (p. ex. supervision, formation particulière, etc.). Chaque demande comprend l'identification de l'intervieweur, le type d'intervieweur (c.-à-d. ordinaire, principal), le nom de l'enquête, la date, le code de tâche (p. ex., interview, formation, dépistage, etc.), le nombre d'heures facturées et les frais pour les enquêtes par ITAO/IPAO, ainsi que des renseignements supplémentaires, tels que les dépenses et le nombre de kilomètres parcourus pour les enquêtes par IPAO. À la fin de 2008, l'IDPP contenait environ 60 millions d'enregistrements de transaction et 9 millions d'enregistrements administratifs au sujet de 722 cycles d'enquête (p. ex., une enquête mensuelle correspond à 12 cycles d'enquête) qui représentent 80 enquêtes uniques (19 agricoles, 9 auprès des entreprises et 52 sociales) accumulés depuis 2003.

En plus de l'information disponible grâce à l'IDPP, l'information sur le plan de sondage et les renseignements externes recueillis par les intervieweurs ou enregistrés pour chaque intervieweur peuvent également être considérés comme des sources importantes de paradonnées. Par exemple, pour chaque enquête, l'information sur le plan de sondage est très souvent utilisée afin d'accroître la valeur analytique des études ou pour de nouveaux projets de collecte, tels que celui de la gestion active. Par ailleurs, les caractéristiques de l'intervieweur (p. ex. son expérience) et les renseignements sur son comportement recueillis par surveillance de l'ITAO/IPAO (ou par piste de vérification), de même que l'information subjective fournie par les intervieweurs au sujet des caractéristiques du

quartier et de la coopération éventuelle des cas échantillonnés (p. ex. évaluation de la probabilité qu'un cas aboutisse à une interview complète après un premier refus) font aussi partie des parodonnées. La plupart de cette information n'a pas été utilisée dans les études dont il est question dans l'exposé, à l'exception de l'expérience de l'intervieweur.

3. Objectifs de recherche

La recherche qui s'appuie sur des parodonnées peut être menée avant, pendant et après la collecte des données afin de comprendre, d'évaluer, de surveiller et d'améliorer le processus de collecte des données. Les objectifs généraux de toutes ces études englobent un ou plusieurs des objectifs particuliers suivants :

- en apprendre davantage au sujet du processus de collecte des données pour des enquêtes individuelles et pour un ensemble d'enquêtes;
- cerner les possibilités d'accroître l'efficacité opérationnelle;
- évaluer le processus de collecte des données, y compris les nouvelles initiatives et les questions de l'heure;
- fournir rapidement une rétroaction et des renseignements personnalisés pour la gestion active des enquêtes;
- maintenir et accroître la qualité des données;
- améliorer la façon dont la collecte des données est menée et gérée.

Toutes ces études sont fondées sur des parodonnées empiriques recueillies de manière automatisée tout au long de la période de collecte des données et non sur des observations isolées ou ponctuelles. En outre, la plupart de ces études ont été réalisées pour de nombreuses enquêtes afin de comparer les résultats en fonction du type d'enquête (p. ex. longitudinales, transversales et enquêtes par composition aléatoire), de même que pour valider et généraliser les conclusions tirées des travaux de recherche.

4. Leçons apprises sur le processus de collecte des données

La présente section contient les faits saillants de la recherche antérieure à l'aide de parodonnées qui ont permis de beaucoup mieux comprendre nombre de problèmes liés au processus et aux pratiques de collecte des données, pour des enquêtes particulières ou un ensemble d'enquêtes. La grande majorité de ces études ont également mis en évidence des possibilités d'amélioration opérationnelle, dont certaines ont déjà été mises en œuvre, tandis que d'autres nécessitent une analyse plus approfondie. Dans l'avenir, on accordera la priorité aux travaux de recherche axés sur les options stratégiques qui pourraient être opérationnellement viables et aboutir à des réductions de coûts ou à un accroissement de la qualité. Comme il est mentionné plus haut, la plupart des travaux de recherche présentés ont trait à des enquêtes sociales et agricoles réalisées par ITAO.

4.1 Nombre d'appels versus temps consacré à la collecte

Afin de mieux comprendre et évaluer la collecte des données, il convient de savoir comment le temps et les efforts consacrés à la collecte des données sont répartis tout au long du processus de collecte. D'habitude, l'accent est mis sur la réduction du nombre d'appels nécessaires pour obtenir un premier contact. Toutefois, en général, une part importante du temps de collecte est consacrée, après le premier contact, aux efforts en vue d'obtenir la coopération du répondant et d'achever l'interview ou de confirmer une non-réponse. Cela vaudrait la peine d'accorder plus d'attention aux appels effectués après le premier contact pour étudier la stratégie et les pratiques de collecte courantes.

4.2 Taux de contact

Aux fins de ces analyses, un contact s'entend de tout appel durant lequel un intervieweur arrive à parler à un membre du ménage, même s'il ne s'agit pas de la personne sélectionnée pour participer à l'enquête. En moyenne, il

faut de deux à quatre appels pour obtenir un premier contact avec un ménage. Comme l'ont montré de nombreuses autres études, les soirées durant la semaine et les fins de semaine sont les moments qui produisent systématiquement les meilleurs taux de contact pour les enquêtes sociales par ITAO, tandis que le matin, l'heure du déjeuner et le début de la soirée sont des moments plus appropriés pour les enquêtes agricoles. Ces conclusions ont été corroborées par les résultats de l'Enquête sociale générale (ESG) sur l'emploi du temps, qui fournit des renseignements permettant de déterminer la probabilité qu'une personne sera à son domicile à une heure particulière d'après l'information socioéconomique. Cette information générale pourrait être utilisée à la phase initiale du processus de collecte des données conjointement avec l'initiative des tranches de temps (voir la section 6.2) en vue d'améliorer les taux de contact durant les quelques premiers appels, surtout si l'on possède certains renseignements socioéconomiques sur la personne avant le début de la collecte des données, comme cela est le cas pour les enquêtes longitudinales.

4.3 Séquence des appels

Il est, certes, important de faire le premier appel à un moment approprié, mais la majorité des appels effectués durant une enquête sont des appels subséquents. L'historique particulier de ces appels subséquents devrait également être pris en considération pour chaque cas échantillonné, parce que la séquence des appels fournit des renseignements spécifiques et ciblés au sujet de ces cas durant la collecte des données et pourrait être utilisée de plusieurs manières.

Au départ, la séquence des appels pourrait servir à accroître la probabilité de prendre contact avec une personne particulière au cours des quelques appels suivants. Par exemple, le taux de contact après les deux premiers appels dépend de l'heure à laquelle chaque appel a été fait. Dans le cas des enquêtes sociales par ITAO, le taux de contact demeure relativement élevé quand au moins un des appels a été fait en soirée. La même conclusion s'applique aux séquences d'appel plus longues. À mesure que le nombre d'appels augmente, la séquence des appels devient un élément d'information plus important que l'information générale sur la population, parce qu'elle fournit des renseignements spécifiques au sujet du profil de chaque cas échantillonné. En outre, elle pourrait être utilisée conjointement avec le résultat des appels précédents. Des études antérieures ont montré que, quand un contact a eu lieu au moment du premier appel, l'heure du second appel n'est pas aussi critique que si aucun contact ne s'est produit. Cependant, il est avantageux, pour les cas où il n'y a pas eu de contact, de faire une nouvelle tentative en soirée. Contrairement aux données sociodémographiques qui fournissent des renseignements généraux au sujet du « meilleur moment » pour appeler un ménage particulier, la séquence des appels fournit des renseignements spécifiques pour chaque cas sur les moyens d'accroître la probabilité de rejoindre le ménage. Deuxièmement, la séquence des appels et les codes de résultat pourraient aussi être utilisés pour prendre des décisions. Ainsi, dans le cas d'enquêtes par composition aléatoire (CA), les études préliminaires ont indiqué que les cas pour lesquels les deux ou trois premiers appels consécutifs se soldent par un « signal rapide occupé » sont plus susceptibles que les autres d'être désignés et classés comme des cas hors du champ de l'enquête à la fin de la période de collecte des données. Par conséquent, cette information pourrait être utilisée pour prendre tôt des décisions au sujet de ces cas et pour économiser les ressources ou les rediriger vers d'autres types de cas en cours. Enfin, l'information au sujet de la séquence des appels peut être conjuguée aux renseignements sur le temps écoulé entre les appels et la durée de la période de collecte des données, afin de s'assurer que les tentatives d'appel sont réparties plus uniformément tout au long de la période de collecte de manière à maximiser les taux de réponse. Il n'est pas souhaitable d'atteindre le nombre limite d'appels d'un cas (limitation du nombre d'appels) dès le milieu de la période de collecte des données, puisqu'on n'aura alors plus la chance d'essayer de prendre contact ultérieurement. En bref, la séquence des appels pourrait être traitée, analysée et utilisée durant le processus de collecte des données pour rajuster et améliorer la stratégie de collecte en fonction de renseignements spécifiques au sujet de chaque cas échantillonné. Cette approche serait semblable à celle des fonctions de score adoptée pour les enquêtes auprès des entreprises. Il s'agit d'un exemple pratique de stratégie de collecte adaptative des données définie dans le cadre conceptuel de la collecte adaptative des données discuté à la section 7.

4.4 Contact versus interview

Selon le type d'enquête, de 40 % à 50 % du nombre total de répondants à l'ITAO sont rejoints durant le premier appel de prise de contact. Une tranche supplémentaire de 4 % à 7 % de répondants sont rejoints le même jour que

celui de la première tentative de prise de contact, par exemple, dans le cas de la prise d'un rendez-vous le matin suivie de l'interview en soirée. Rejoindre la deuxième moitié des répondants requiert un effort considérable. Une étude plus approfondie du processus et des pratiques de collecte après le premier contact avec un ménage serait utile pour accroître l'efficacité.

4.5 Interaction entre les enquêtes

Durant une journée donnée, les bureaux régionaux réalisent de nombreuses enquêtes par ITAO qui se font concurrence pour les ressources de collecte des données disponibles. Certaines enquêtes pourraient en être au début de leur période de collecte, tandis que d'autres en sont au milieu ou à la fin de la leur. En outre, les enquêtes n'ont pas toujours la même priorité. Par exemple, l'Enquête sur la population active (EPA), qui est l'une des enquêtes les plus importantes réalisées par Statistique Canada, recueille des données auprès d'environ 55 000 ménages sur dix jours chaque mois, ce qui mobilise la presque totalité des ressources de collecte pendant les premières journées de collecte.

4.6 Dotation en personnel versus charge de travail en cours

Certaines études ont été réalisées afin de mieux saisir la relation entre les efforts déployés pour obtenir les interviews et la charge de travail en cours attendue durant le cycle de collecte des données. Des travaux de recherche antérieurs (Laflamme, 2008a), ainsi que de nombreuses observations sur le terrain émanant de l'initiative de gestion active (Laflamme et coll., 2008b) ont laissé entendre que le nombre d'intervieweurs n'est pas toujours en harmonie avec l'échantillon de cas en progrès et la productivité attendue. Par exemple le fait que les cas seront vraisemblablement appelés fréquemment durant une même journée pendant la deuxième moitié de la période de collecte suggère qu'à ce stade-là, le niveau de dotation en intervieweurs est plus élevé que la charge de travail relative à l'échantillon. La proportion de cas en cours faisant l'objet de tentatives de prise de contact pendant des journées consécutives, ainsi que la diminution rapide de la productivité, sont également de bons indicateurs. Les gestionnaires de la collecte des données ont, par conséquent, besoin d'outils de gestion et de planification de la dotation en intervieweurs afin de réduire certaines pressions exercées sur la productivité et les coûts de collecte (Couper et coll., 1998). Ces types de projets feront définitivement partie des priorités de recherche au cours des quelques prochaines années. Il convient de souligner que les règles concernant les avis de changement de quart s'adressant à des intervieweurs syndiqués doivent aussi être prises en compte dans tous les plans d'action qui résulteront de ces travaux de recherche.

4.7 Recherche ponctuelle

En plus de ces projets de recherche particuliers, nombre d'études ponctuelles ont été réalisées au sujet de questions de l'heure ou de problèmes opérationnels qui nécessitaient une intervention immédiate, ainsi que pour répondre à des questions au sujet des progrès et des résultats de la collecte des données soulevées par la haute direction. Ainsi, la relation entre l'expérience de l'intervieweur et la productivité, ainsi que l'étude de l'effet d'un plan d'échantillonnage adaptatif dans un environnement de collecte à plusieurs modes et à plusieurs emplacements ne sont que deux exemples. La plupart de ces études ponctuelles ont également contribué à une meilleure compréhension du processus et des pratiques de collecte des données dans le cas d'enquêtes particulières et d'un ensemble d'enquêtes.

5. Recherche en cours sur la productivité et le coût des enquêtes

Le principal objectif de ce genre de recherche est d'étudier la relation qui existe entre les données sur la production et les coûts durant la période de collecte. Le premier volet du projet de recherche consiste à fusionner et à consolider les données sur la production (BTH) et sur les coûts (paye) en créant un enregistrement unique qui contient un résumé de l'information de production ainsi que l'information financière par intervieweur, enquête et journée. Il comporte aussi l'évaluation de la cohérence de cette information consolidée entre les deux sources de données, dont nous ne présentons pas les résultats ici. Les concepts relatifs à cette information résumée sont décrits aux sous-sections 5.1 et 5.2. Les sous-sections qui suivent présentent les faits saillants de cette relation, ainsi qu'un

aperçu des premiers travaux de recherche portant sur les indicateurs de productivité et des études d'analyse des coûts.

5.1 Paradonnées sur la production

Le temps système de non-interview représente le temps consacré à un cas pour prendre contact, essayer d'obtenir la coopération du répondant, confirmer une non-réponse ou déterminer s'il s'agit d'un cas hors du champ de l'enquête (c.-à-d. les appels effectués avant l'interview proprement dite, y compris le temps de dépistage enregistré dans le système). Par ailleurs, le temps système d'interview représente uniquement le temps consacré à l'exécution des interviews. La somme de ces deux variables est égale au temps système total qui représente la durée de la connexion au système se rapportant à un cas ouvert. Il ne tient pas compte d'activités telles que le dépistage effectué en dehors du système, des pauses et du temps écoulé entre deux appels. Ces activités représentent une partie de la différence entre le temps système et le temps rémunéré.

5.2 Paradonnées sur les coûts

Les heures de collecte directe rémunérées représentent le temps facturé pour des activités directes de collecte de données (p. ex. interviews, appels avec contact, appels sans contact, dépistage, etc.) qui est conceptuellement comparable au temps système total. Les autres heures de collecte rémunérées représentent le temps facturé pour les activités indirectes de collecte de données, telles que la formation spéciale, la supervision ou d'autres activités ou tâches de collecte connexes. Le nombre total d'heures rémunérées, qui est égal à la somme du nombre d'heures de collecte directe rémunérées et du nombre d'autres heures de collecte rémunérées, fournit une très bonne approximation du coût des enquêtes réalisées par ITAO.

5.3 Relation entre la production et le coût d'enquête

La figure 5.3-1 illustre l'évolution de la production et des coûts tout au long du cycle complet d'enquête pour l'Enquête sur la dynamique du travail et du revenu (EDTR 2007) réalisée de janvier à avril 2007. Toutes les séries ont un même profil, ce qui donne à penser que la relation entre les données de production et les données administratives est très bonne. En particulier, il existe une forte corrélation entre le temps système total et les heures de collecte directe rémunérées ($\rho > 0,95$). Il convient également de souligner que la proportion du temps de production est un très bon prédicteur de la proportion d'heures rémunérées facturées, comme l'illustre le tableau 5.3-1. Pour les besoins du SIG et du suivi des coûts d'enquête, le temps système total peut être utilisé comme approximation fiable des coûts d'enquête directs. Ces premiers résultats ont généré de nombreuses idées de travaux de recherche, en particulier, en ce qui concerne l'analyse de la productivité et des coûts d'enquête, comme nous le décrivons brièvement dans les deux sous-sections qui suivent.

Figure 5.3-1
Distribution du temps système et des heures rémunérées par jour de collecte : Enquête sur la dynamique du travail et du revenu (EDTR 2007)

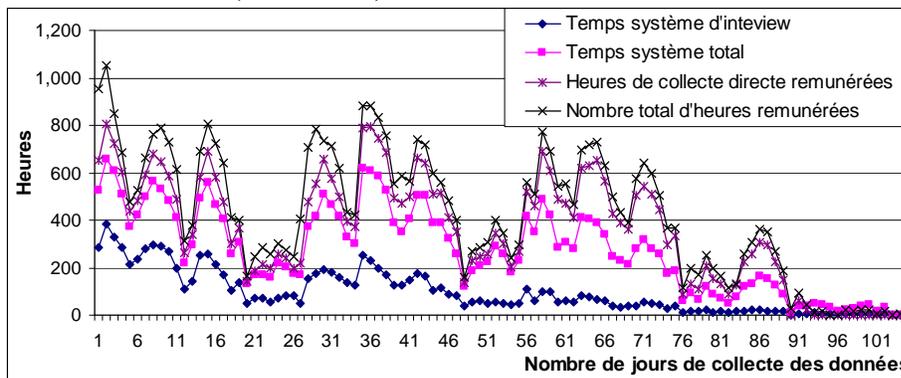


Tableau 5.3-1**Distribution du temps système cumulatif et des heures rémunérées par période de 10 jours – EDTR 2007**

Période de 10 jours	Taux de réponse cumulatif %	% cumulatif		
		Temps système total	Heures de collecte directe rémunérées	Nombre total d'heures rémunérées
1-10 jours	21.7	17.9	16.5	16.8
11-20 jours	33.8	30.3	27.8	28.4
21-30 jours	41.2	39.3	36.3	37.9
31-40 jours	54.1	55.3	51.8	52.9
41-50 jours	61.6	67.0	63.0	63.6
51-60 jours	67.3	78.2	74.0	74.1
61-70 jours	72.6	89.2	87.4	86.8
71-80 jours	75.2	95.0	94.9	94.5
81-90 jours	76.7	98.5	99.7	99.4
91 jours et plus	77.1	100.0	100.0	100.0

5.4 Productivité de l'enquête

L'élaboration d'indicateurs de productivité qui pourraient être utilisés pour gérer le rendement et les coûts d'enquête durant le processus de collecte des données a été considérée comme étant une composante importante du projet global de recherche axé sur les paradonnées. Ainsi, le ratio du temps système d'interview (c.-à-d. le temps consacré uniquement à l'interview proprement dite) au temps système total (qui comprend tous les appels fructueux et non fructueux) est un bon indicateur de l'évolution de la productivité de l'enquête au cours du processus de collecte des données, car il relie l'effort en vue d'obtenir l'interview au résultat (c.-à-d. l'interview) pour une enquête et une période particulière, généralement une journée. Cet indicateur de productivité diminue au cours du temps pour toutes les enquêtes par ITAO, ce qui est normal puisque la proportion de cas en cours plus difficiles augmente vers la fin de la collecte. Cependant, il est important de souligner que les indicateurs de ce genre dépendent directement de la durée de l'interview et du taux de réponse. Par exemple, pour un taux de réponse donné, les enquêtes dont l'interview est longue auront un ratio de productivité plus élevé que celles dont l'interview est plus courte. Inversement, l'indicateur de production définit comme étant le ratio du temps système de non-interview cumulatif à la taille d'échantillon est moins affecté que le précédent par la durée de l'interview (mais il continue de dépendre du taux de réponse). En pratique, il donne une mesure du temps moyen requis pour obtenir la coopération des répondants ou pour confirmer une non-réponse ou un cas hors du champ de l'enquête pour tous les cas échantillonnés. Ce type d'indicateur convient mieux que le précédent pour évaluer et comparer la productivité de diverses enquêtes. Enfin, il faut aussi noter que le ratio du temps système total au nombre de cas achevés (c.-à-d. le nombre d'interviews) pourrait également être utile pour estimer le temps requis (et ensuite les ressources de collecte) en vue d'atteindre les cibles de réponse dans certaines enquêtes (Lepkowski et coll., 2007). Cependant, comme ce ratio augmente généralement au cours de la période de collecte, il sous-estime souvent la quantité de ressources requises.

5.5 Analyse des coûts d'enquête

Récemment, l'accès rapide à des renseignements détaillés sur les coûts, ainsi que la constatation de l'existence d'une relation entre les données de production et les données sur la paye ont initié une analyse des coûts d'enquête. Nous avons notamment entrepris une étude de l'effet de la limitation du nombre d'appels sur les coûts d'enquête pour certaines des enquêtes longitudinales les plus importantes (Enquête sur la dynamique du travail et du revenu (EDTR) et Enquête auprès des jeunes en transition (EJET)) et transversales (p. ex. Enquête sur la santé dans les collectivités canadiennes (ESCC)) réalisées par ITAO. Cette étude a tiré parti de la relation entre les données sur la production et celles sur les coûts, parce qu'il n'existe aucun lien direct entre un appel et son coût (p. ex. le temps rémunéré n'est relié à aucun cas en particulier). Selon les résultats provisoires, théoriquement, l'économie possible maximale aurait pu être de l'ordre de 3 % à 4 % sous une limitation à 40 appels quand il est tenu compte de tout le temps de production consacré à des appels au-delà de la limite imposée. En pratique, le temps consacré à des appels au-delà de la limite pour une enquête particulière n'est pas automatiquement « économisé », puisque les

intervieweurs continuent de travailler sur d'autres cas disponibles. Autrement dit, il reste toujours des cas requérant du travail. L'effet de la limitation du nombre d'appels sur les coûts d'enquête est encore à l'étude et la portée de la recherche sera étendue à d'autres types d'enquêtes.

6. Évaluation de nouvelles initiatives de collecte des données

La présente section donne un aperçu de deux nouvelles initiatives de collecte des données, à savoir la limitation du nombre d'appels et les tranches de temps, y compris les résultats provisoires pour ce qui est de leur effet sur la progression et les coûts de l'enquête. Ces travaux de recherche concernent essentiellement la réalisation du troisième objectif de l'ensemble des objectifs généraux de recherche sur la collecte des données (section 3).

6.1 Limitation du nombre d'appels

Avant janvier 2006, il n'existait aucune limite du nombre d'appels pouvant être effectués pour compléter un cas, en tant que réponse, non-réponse ou cas hors du champ de l'enquête. Il n'était pas inhabituel que plus de 25 appels soient faits avant de compléter une interview, voire même plus dans le cas des enquêtes longitudinales. Par exemple, en 2006, environ 11 % et 14 % des répondants à l'Enquête sur la dynamique du travail et du revenu (EDTR) et à l'Enquête auprès des jeunes en transition (EJET) ont été rejoints après 25 appels. À partir de 2007, une politique de limitation du nombre d'appels est entrée en vigueur progressivement pour toutes les enquêtes par ITAO. Le but de cette politique est de limiter le nombre d'appels permis par cas en vue de contrôler le fardeau de réponse et d'améliorer la rentabilité des appels effectués. En pratique, le nombre maximal d'appels a été fixé à 20 et à 5, respectivement, pour les numéros de téléphone publiés et non publiés dans le cas des enquêtes par CA, à 25 pour les enquêtes auprès de répondants ciblés (sauf pour l'Enquête sur la santé dans les collectivités canadiennes (ESCC) pour laquelle le maximum est fixé à 40), à 40 pour les enquêtes longitudinales et à 15, 25 ou aucune limite pour les enquêtes agricoles, selon la priorité du cas.

Bien que la limitation du nombre d'appels soit maintenant appliquée à la plupart des enquêtes, ce qui a éliminé un très grand nombre d'appels, des réserves ont été formulées quant à son effet sur les taux de réponse, les estimations d'après les données d'enquête et les coûts (section 5.5). Ainsi, de nombreuses études ont été réalisées pour évaluer l'incidence de la limitation du nombre d'appels sur les taux de réponse en se servant des parodonnées disponibles avant son introduction en 2006. Il est estimé que 1,6 %, 2,1 % et 2,5 % des répondants auraient été perdus pour l'ESCC, l'EDTR et l'EJET, respectivement, sous une limitation à 40 appels. Dans le cas des enquêtes longitudinales, telles que l'EDTR et l'EJET, la limitation du nombre d'appels pourrait avoir un effet négatif sur la qualité des estimations, étant donné l'effet cumulatif d'une vague à l'autre. À l'heure actuelle, le nombre de cas soumis à une limitation dans la plupart des enquêtes est évalué dans le contexte de l'initiative de la gestion active afin de surveiller l'effet de la limitation des appels sur les taux de réponse. Par ailleurs, l'étude de son effet sur les estimations et les coûts se poursuit.

6.2 Tranche de temps

Étant donné la limitation des appels, il est important de s'assurer que tous les appels sont traités de la meilleure manière possible. L'option Tranche de temps de l'ordonnanceur d'appels du système de gestion de l'ITAO a été utilisée pour faciliter la gestion de la nouvelle politique de limitation du nombre d'appels. En pratique, les tranches de temps permettent de s'assurer qu'un nombre déterminé de tentatives d'appel est effectué à différents moments de la journée et divers jours de la semaine avant qu'un cas soit finalisé. L'introduction des tranches de temps a produit des signes encourageants d'une légère réduction du nombre moyen d'appels pour obtenir un premier contact dans le cas des enquêtes par CA. Si l'exécution de deux à quatre appels paraît raisonnable, il est important de souligner que de 40 % à 50 % des répondants sont rejoints au moment du premier contact et que, par conséquent, de nombreux autres requièrent un nombre beaucoup plus élevé d'appels. Enfin, il faut mentionner qu'évaluer l'effet de la tranche de temps tout au long de la période de collecte des données n'est pas une tâche simple, parce que de nombreux facteurs opérationnels interviennent.

7. Sommaire et réalisations

La présente section offre un résumé des études portant sur les paradonnées en fonction des objectifs de recherche, y compris une brève description des modifications apportées au processus de collecte des données et des faits nouveaux résultant des résultats de ces études.

Les analyses passées, courantes et ponctuelles des paradonnées ont définitivement permis de mieux comprendre le processus et les pratiques de collecte de données dans le contexte d'enquêtes par ITAO individuelles ou d'un ensemble d'enquêtes. Des modifications ont déjà été apportées à la collecte des données afin de tirer parti des résultats de ces analyses. Par exemple, la proportion de quarts de soirée par opposition aux quarts de jour augmente maintenant progressivement au cours de la période de collecte afin d'accroître le taux de contact et la productivité. En outre, les tranches de temps ont été adaptées pour certains types d'enquête, afin d'utiliser l'information disponible avant et pendant la collecte, par exemple l'information sur le plan de sondage et celle sur les caractéristiques socioéconomiques des ménages recueillie durant l'établissement de la liste des membres du ménage. En outre, l'effet des nouvelles initiatives de collecte des données a été évalué et continue d'être surveillé, ce qui s'est traduit par une modification de la définition de la limitation du nombre d'appels pour certaines grandes enquêtes.

Les résultats des travaux de recherche ont également montré qu'il est nécessaire d'élaborer une stratégie de collecte des données plus souple et efficace, non seulement en vue de réduire les coûts de collecte, mais aussi pour mieux tirer parti des appels permis en vertu de la nouvelle politique de limitation du nombre des appels. L'approche devrait évoluer au cours de la période de collecte afin de tenir compte de la progression de l'enquête, de la productivité et des coûts. Cette stratégie de collecte fait essentiellement référence à l'approche du plan de sondage adaptatif défini initialement par Groves et Heeringa (2006). Récemment, certains travaux ont été entrepris à Statistique Canada en vue d'élaborer un cadre de conception adaptative (Laflamme et Mohl, 2007) qui comprend deux grandes composantes : la gestion active (Hunter et Carbonneau, 2005) et la collecte adaptative. L'idée essentielle est d'évaluer constamment la progression de la collecte de données (gestion active) en utilisant l'information la plus récente disponible et d'ajuster les stratégies de collecte de données afin d'utiliser le plus efficacement possible les ressources disponibles restantes (collecte adaptative). Plus précisément, un outil de gestion active est nécessaire pour évaluer et surveiller de près la progression, l'effort et les coûts d'enquête au cours du processus de collecte des données afin de prédire, de cerner et de corriger les problèmes opérationnels au besoin. Des données à jour et des indicateurs de production sont également nécessaires pour déterminer quelles sont les étapes critiques de la collecte de données qui requièrent l'apport de modifications plus importantes au processus de collecte. Autrement dit, une approche de collecte adaptative est requise pour qu'il soit possible d'apporter des modifications plus importantes à la stratégie de collecte des données à divers moments durant la période de collecte. L'approche tiendrait compte de la progression de l'enquête, du temps et de l'effort déjà consacrés aux cas en cours restants et de caractéristiques telles que la proportion de l'échantillon dans les groupes de cas de refus ou nécessitant un dépistage, ou le nombre moyen d'appels pour les cas en cours. Naturellement, l'approche tiendrait également compte des effets sur la qualité des données et sur les coûts. L'initiative de gestion active a été mise en œuvre pour la plupart des enquêtes sociales et agricoles par ITAO et a été élaborée pour les enquêtes par IPAO (Laflamme, Maydan et Miller 2008).

8. Futurs travaux

Parmi les possibilités d'amélioration déjà mentionnées et discutées, certaines requièrent des études plus approfondies afin de mieux évaluer leur faisabilité, les avantages opérationnels ainsi que les améliorations du processus de collecte des données dans un contexte de conception adaptative. En particulier, la priorité sera accordée aux futurs travaux de recherche sur les paradonnées relatifs à des projets qui pourraient être opérationnellement viables et donner lieu à des économies ou à l'accroissement de l'actualité ou de la qualité des données. Par exemple, une analyse plus détaillée de la séquence des appels effectués afin d'obtenir un premier contact (y compris le temps écoulé entre les appels), ainsi que le temps consacré à l'obtention de la coopération du répondant après un premier contact est nécessaire afin de maximiser la probabilité d'effectuer un moins grand nombre d'appels pour prendre contact avec le répondant et obtenir sa coopération. Un autre exemple est celui d'une étude sur la relation entre les

indicateurs de productivité et de progression de l'enquête en vue de définir des phases distinctes de collecte des données pour lesquelles des modifications importantes (collecte adaptative) sont requises dans le contexte d'un plan d'enquête adaptatif. Enfin, il faudra étudier des outils permettant de mieux planifier la bonne utilisation des ressources de collecte des données (intervieweurs) durant la période de collecte en fonction des progrès observés, parce qu'elle a une incidence directe sur la productivité et les coûts d'enquête. L'un des défis les plus importants consisterait à introduire progressivement, à intégrer et à consolider cette série d'initiatives dans une stratégie de collecte des données fondée sur un plan d'enquête adaptatif destiné à améliorer la rentabilité du processus de collecte des données à long terme.

9. Conclusion

Les parodonnées ont été la pierre angulaire de la recherche sur la collecte des données à Statistique Canada et continuent d'être utilisées à grande échelle. En fait, l'accès rapide et facile à une base de parodonnées complètes offre de nombreux avantages. En premier lieu, les résultats de la recherche sont fondés sur des mesures objectives et empiriques au sujet du processus de collecte. En deuxième lieu, cela permet de comparer divers types d'enquêtes, de même que valider les résultats de recherche, puisque les mêmes études peuvent être reproduites sur de nombreux types d'enquêtes ou cycles d'enquête. Troisièmement, l'accès à des données historiques donne l'occasion d'effectuer une analyse de tendance au cours du temps et d'évaluer l'effet de nouvelles initiatives. Quatrièmement, les parodonnées sont obtenues automatiquement durant le cycle de collecte, si bien que le coût est presque nul et que le fardeau imposé aux intervieweurs est très faible. Le coût principal de l'analyse des parodonnées tient à la création et à la tenue à jour de la base de parodonnées et aux rapports de SIG. L'analyse et l'interprétation de ces données comportent aussi un coût humain, qui inclut l'apprentissage et la formation. Même si la recherche à l'aide de parodonnées nécessite un investissement initial, les avantages de ce genre d'analyse sont d'une grande importance, vu que la collecte de données représente une part très importante du coût global d'enquête pour de nombreux programmes statistiques.

Bibliographie

- Couper, M. P., Baker, R.P., Bethlehem, J., Clark, C.Z.F., Martin, J. Nicholls W.L. et O'Reilly, J.M. (Éds.) (1998). *Computer Assisted Survey Information Collection*, Wiley, 301-306.
- Groves, R.M. et Heeringa, S.G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society, Series A*. 169.
- Hunter, L. et Carbonneau, J.-F. (2005). Une méthode de collecte de données d'enquête axée sur la gestion active, *Recueil: Symposium 2005, Défis méthodologiques reliés aux besoins futures d'information*. Statistique Canada.
- Laflamme, F. et Mohl, C. (2007). Research and Responsive Design Options for Survey Data Collection at Statistics Canada. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Laflamme, F., Maydan, M. et Miller, A. (2008). Using Paradata to Actively Manage Data Collection. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Laflamme, F., (2008), Understanding Survey Data Collection through the Analysis of Paradata at Statistics Canada. *Proceedings of the American Association for Public Opinion Research Section on Survey Research Methods*, American Statistical Association.
- Lepkowski, J.M. et coll. (Éds.) (2007). *Advances in Telephone Survey Methodology*, Wiley, 363-367.