

Mise à l'essai des stratégies de collecte pour les enquêtes en ligne fondées sur l'autodéclaration

Margaret Wu, François Brisebois et Lecily Hunter¹

Résumé

En janvier et février 2014, Statistique Canada a réalisé un essai en vue de mesurer l'efficacité des diverses stratégies de collecte au moyen d'une enquête en ligne fondée sur l'autodéclaration. On a communiqué avec les unités échantillonnées en leur envoyant des lettres de présentation par la poste et on leur a demandé de répondre à l'enquête en ligne, sans contact avec un intervieweur. L'essai visait à mesurer les taux de participation à une enquête en ligne, ainsi que les profils des répondants et des non-répondants. On a mis à l'essai différents échantillons et lettres, afin de déterminer l'efficacité relative des diverses méthodes. Les résultats du projet serviront à renseigner les responsables de diverses enquêtes sociales qui se préparent à ajouter à leurs enquêtes la possibilité d'y répondre par Internet. Le document présentera la méthode générale de l'essai, ainsi que les résultats observés pendant la collecte et l'analyse des profils.

Mots clés : Questionnaire électronique; enquête par la poste; taux de réponse; profil de réponse.

1. Introduction

L'utilisation des enquêtes en ligne fondées sur l'autodéclaration est de plus en plus populaire ces dernières années à Statistique Canada et dans d'autres organismes statistiques nationaux partout dans le monde. De façon plus particulière, Statistique Canada a commencé à adopter le mode de collecte en ligne dans les enquêtes auprès des ménages en tant que complément d'autres modes de collecte, au moyen d'un questionnaire électronique (QE).

Au Canada, dans le cadre du Recensement de la population, on a offert pour la première fois l'option de réponse en ligne en 2006 (Tallon et Ramsay, 2010), et on a continué de le faire en 2011 (Hamel, 2012). Pour d'autres enquêtes auprès des ménages, l'intégration d'une option de réponse par Internet est permanente, et on s'attend à ce que la collecte en ligne devienne le principal mode de collecte. Dans le cadre de l'Enquête sur la population active au Canada, on a mené un essai pilote au moyen de QE en 2013 (Gambino, Laflamme et Wu, 2014). Pour l'Enquête sociale générale sur l'identité sociale, on a mené un essai pilote au moyen de QE en 2012, et on a commencé à offrir une option de QE pour la collecte régulière à partir de 2013 (Gambino, Laflamme et Wu, 2014).

On s'attend à ce que les avantages que présente le mode de réponse en ligne soient nombreux, y compris des économies de coût potentiellement importantes par rapport à d'autres modes de collecte. Toutefois, l'utilisation de QE dans les enquêtes auprès des ménages à Statistique Canada n'en est encore qu'à ses débuts, et des questions subsistent quant à la meilleure façon d'utiliser ce mode. Une étude a été menée en 2014 pour vérifier l'efficacité de deux stratégies de collecte pour le QE, les unités échantillonnées ayant été contactées uniquement au moyen d'invitations par la poste. En comparaison, dans le cadre de l'Enquête sur la population active et de l'Enquête sociale générale, on a établi le premier contact par téléphone, avant d'offrir une option de QE.

Le présent document comprend les détails et les résultats de cette étude. La section 2 explique les objectifs de l'essai, les stratégies de collecte mises à l'essai et la méthodologie. La section 3 présente les taux de participation de

¹Margaret Wu, Statistique Canada, 100, promenade Tunney's Pasture, Ottawa (Ontario) Canada, K1A 0T6, margaret.wu@statcan.gc.ca; François Brisebois, Statistique Canada, 100, promenade Tunney's Pasture, Ottawa (Ontario) Canada, K1A 0T6, francois.brisebois@statcan.gc.ca; Lecily Hunter, Statistique Canada, 100, promenade Tunney's Pasture, Ottawa (Ontario) Canada, K1A 0T6, lecily.hunter@statcan.gc.ca

ces stratégies. La section 4 présente les résultats d'une analyse de profils de répondants, et la section 5, un sommaire des résultats.

2. Aperçu de l'étude

2.1 Deux stratégies de collecte pour le QE

Dans les deux stratégies, des invitations sur papier à une enquête en ligne ont été envoyées par la poste aux unités échantillonnées, et il s'agissait de la seule méthode de contact. Les objectifs de l'étude étaient de mesurer les taux de participation en vertu de ces stratégies et d'effectuer une analyse des profils des répondants. Les deux stratégies étaient les suivantes :

1. **Stratégie du ménage :** Des lettres ont été envoyées par la poste à un échantillon de logements, avec une invitation à participer à l'enquête au moyen d'un QE. L'invitation était adressée à « L'occupant ». Une fois en ligne, le répondant du ménage a dû dresser une liste des membres du ménage. À ce moment-là, une personne de 15 ans et plus du ménage a été sélectionnée de façon aléatoire pour répondre au questionnaire. Si le répondant sélectionné était la même personne que celle qui avait dressé la liste, l'application du QE est passée au contenu; autrement, l'application du QE a pris fin et une nouvelle invitation a été envoyée au répondant sélectionné (appelé le répondant secondaire). Si le répondant du ménage avait fourni une adresse de courriel, la nouvelle invitation a été envoyée par courriel; autrement, elle a été envoyée par la poste et était adressée au répondant sélectionné désigné.
2. **Stratégie du répondant cible :** Des lettres ont été envoyées à un échantillon de personnes sélectionnées avec une invitation à participer à l'enquête au moyen d'un QE. Chaque lettre d'invitation était adressée au nom de la personne sélectionnée (« cible »). Grâce à cette stratégie, il n'était pas nécessaire de dresser une liste, et seule la personne sélectionnée pouvait répondre au QE. Deux lettres d'invitation différentes ont été mises à l'essai en vertu de cette stratégie; la lettre P1 était une lettre type, la même que celle utilisée pour la stratégie des ménages; la lettre P2 comportait une phrase additionnelle indiquant que le répondant pouvait être contacté par un intervieweur s'il ne répondait pas².

Le principal inconvénient de la stratégie du ménage est que le déroulement était plus compliqué. Cette stratégie nécessitait une liste et l'envoi d'une invitation à un répondant secondaire sélectionné à partir de cette liste. Cela a entraîné un deuxième niveau de non-réponse, étant donné qu'il pouvait y avoir deux répondants par ménage. La stratégie du répondant cible a comme avantage un enchaînement plus simple, mais par ailleurs, exige une base de sondage au niveau de la personne avec des noms et des adresses à jour, ce qui est généralement plus difficile à obtenir à Statistique Canada.

2.2 Base de sondage et échantillon

La population cible des deux stratégies était composée de tous les ménages qui avaient répondu à l'Enquête nationale auprès des ménages (ENM) de 2011 de Statistique Canada et pour lesquels une adresse d'envoi par la poste était disponible dans le Fichier de l'univers des logements (FUL) de la Base de sondage pour les enquêtes auprès des ménages (BSEM) de Statistique Canada. La base de sondage représentait par conséquent une combinaison de l'ENM et du FUL et était identique pour les deux stratégies mises à l'essai. Les territoires canadiens, les réserves indiennes et les logements collectifs étaient exclus. À noter que les adresses postales provenaient du FUL, et que les noms utilisés pour envoyer les lettres d'invitation pour la stratégie du répondant cible provenaient de l'ENM. À noter aussi que l'ENM a eu lieu deux ans et demi avant cette étude, ce qui signifie que l'information concernant les noms des répondants pouvait être périmée.

² Il n'y avait pas de plan de suivi des non-répondants dans les faits. Cette phrase a été ajoutée simplement pour vérifier les répercussions sur les taux de participation d'une « menace » de suivi possible.

La taille de l'échantillon pour la stratégie du ménage était de 30 000 ménages. L'échantillon pour la stratégie du répondant cible était de 6 500, réparti en deux échantillons de 5 000 et 1 500, l'échantillon le plus petit ayant reçu la lettre d'invitation spéciale mentionnant un suivi possible. Un échantillon aléatoire simple de ménages a été tiré dans chaque région du Canada; pour l'échantillon du répondant cible, une personne âgée de 15 ans et plus a été sélectionnée de façon aléatoire à partir de chaque ménage sélectionné, au moyen de la liste de l'ENM.

2.3 Collecte

La collecte a duré environ six semaines et s'est déroulée en janvier et en février 2014. Compte tenu des ressources limitées et du temps nécessaire pour la mise à l'essai, on a réutilisé un module de QE existant : un module de l'Enquête sociale générale (ESG) lié à l'identité sociale et à l'utilisation d'Internet. Le titre de l'enquête officielle pour l'essai était l'Enquête sur l'utilisation d'Internet et des autres médias (EUIAM).

3. Taux de participation

L'objectif principal de l'essai était de mesurer les taux de participation à partir des stratégies de QE possibles pour les enquêtes auprès des ménages.

Les taux de participation pour l'échantillon des ménages sont présentés dans les tableaux 3-1, 3-2 et 3-3. Le tableau 3-1 présente les chiffres et les taux de renvoi par le bureau de poste, d'ouverture de session et de soumission (c.-à-d. les cas qui ont terminé l'enquête). Dans l'ensemble, une fois exclus les renvois par le bureau de poste, le taux d'ouverture de session était de 30 % et le taux de soumission de 22 % pour les 30 000 cas échantillonnés. Le tableau 3-2 ventile le nombre de cas soumis, selon que la même personne a été sélectionnée et a poursuivi immédiatement l'enquête après la liste, ou qu'un membre du ménage différent a été sélectionné et qu'une nouvelle invitation a été lancée : les cas dans lesquels la même personne était sélectionnée représentaient 75 % de tous les questionnaires soumis, et les cas où quelqu'un d'autre avait été sélectionné, les 25 % qui restent. Le tableau 3-3 fournit plus de renseignements sur les cas pour lesquels une personne différente a été sélectionnée : il y avait 3 465 cas où un membre différent du ménage a été sélectionné à partir de la liste; 2 830 de ces cas ont reçu des invitations à participer à l'enquête. À noter que ce ne sont pas tous les cas qui ont reçu des invitations, par exemple, si on considérait que la période de collecte était trop avancée; à noter aussi que la date de clôture pour les invitations par courriel était postérieure à celle des invitations par la poste parce qu'on n'avait pas à tenir compte du délai d'envoi par la poste. Parmi les invitations envoyées, les taux globaux d'ouverture de session et de soumission ont été de 58 % et de 57 %. Ces taux étaient assez différents selon la méthode de contact : les taux étaient supérieurs d'environ 20 % pour les personnes contactées par courriel par rapport à celles contactées par la poste (64 % d'ouverture de session et 63 % de soumission pour le courriel; 43 % d'ouverture de session et 42 % de soumission pour l'envoi par la poste). À noter toutefois que c'est le répondant qui fait le choix initial de fournir une adresse de courriel pour le membre sélectionné du ménage, ce qui détermine par la suite le mode de la deuxième invitation.

Les taux de participation pour l'échantillon des répondants cibles sont présentés dans le tableau 3-4. Globalement, le taux de renvoi par le bureau de poste était de 12 %. Si l'on exclut les renvois par le bureau de poste, les taux d'ouverture de session et de soumission étaient de 31 % et de 30 % au total. Toutefois, ces taux variaient selon la version de la lettre d'invitation : les taux d'ouverture de session et de soumission étaient de 30 % et de 29 % pour l'échantillon ayant reçu la lettre type, et de 34 % et 34 % pour l'échantillon ayant reçu la lettre mentionnant un suivi possible.

Le tableau 3-5 présente des comparaisons entre les deux échantillons et comprend les résultats des tests statistiques des différences. Les taux d'ouverture de session initiaux pour les échantillons des ménages et des répondants cibles sont comparables, soit 30 %. Toutefois, le taux de présentation pour les cas terminés est significativement différent (22 % pour les ménages, 29 % pour les répondants cibles avec la lettre type), probablement en raison du fait qu'il a fallu communiquer à nouveau avec l'échantillon des ménages lorsque quelqu'un d'autre était sélectionné. Lorsque l'on compare les deux lettres d'invitation de la stratégie des répondants cibles, les taux d'ouverture de session sont significativement différents (30 % pour la lettre type et 34 % pour la lettre avec un suivi), ce qui montre que la « menace » d'un suivi a un effet sur la réponse. En outre, même si on ne l'a pas vérifié officiellement, le taux de renvoi par le bureau de poste est beaucoup plus élevé pour l'échantillon des répondants cibles (12 % pour les

répondants cibles, 2 % pour les ménages). Cela est probablement dû au fait que les coordonnées étaient périmées; il faut se rappeler que les noms des répondants ont été tirés de l'ENM, qui s'est déroulée deux ans et demi avant l'étude, et il est presque certain que certaines personnes ont déménagé au cours de cette période.

Tableau 3-1
Échantillon des ménages : Résultats globaux

	Chiffre	Taux
Échantillon	30 000	
Renvois par le bureau de poste	630	2 %
Ouvertures de session	8 673	30 %
Soumissions (contenu complet)	6 324	22 %

Nota : Les taux d'ouverture de session et de présentation excluent les renvois par le bureau de poste.

Tableau 3-2
Échantillon des ménages : Ventilation des soumissions

Total des cas soumis (contenu complet)	6 324	
↳ Même personne sélectionnée	4 720	75 %
↳ Personne différente sélectionnée	1 604	25 %

Tableau 3-3
Échantillon des ménages : Résultats pour le répondant secondaire

	Liste soumise	Invitations envoyées	Ouverture de session		Soumission	
Total	3 465	2 830	1 632	58 %	1 604	57 %
Contact par courriel	s.o.	1 980	1 267	64 %	1 244	63 %
Contact par lettre	s.o.	850	365	43 %	360	42 %

Notes : Les chiffres pour la liste soumise comprennent quelques (12) cas supprimés comportant des erreurs dans les coordonnées. Les chiffres pour les invitations envoyées excluent 59 invitations par courriel qui sont revenues.

Tableau 3-4
Échantillon des répondants cibles : Résultats

	Lettre P1 (type)		Lettre P2 (suivi)		Total	
Échantillon	5 000		1 500		6 500	
Renvois par le bureau de poste	597	12 %	167	11 %	764	12 %
Ouvertures de session	1 321	30 %	454	34 %	1 775	31 %
Soumissions	1 295	29 %	447	34 %	1 742	30 %

Nota : Les taux d'ouverture de session et de soumission excluent les renvois par le bureau de poste.

Tableau 3-5
Comparaison des stratégies

Comparaison	Groupe 1		Groupe 2		valeur p
Taux d'ouverture de session	Ménage	30 %	Répondant cible, lettre type	30 %	0,60
Taux de soumission global	Ménage	22 %	Répondant cible, lettre type	29 %	<0,001
Taux d'ouverture de session pour les lettres (stratégie des répondants cibles)	Lettre type	30 %	Lettre mentionnant un suivi	34 %	0,02
Renvois par le bureau de poste	Ménage	2 %	Répondant cible	12 %	Non testé

4. Profils des répondants

Afin de déterminer quels types de ménages et de personnes seraient les plus susceptibles de répondre selon ces stratégies, des modèles de régression logistique ont été appliqués afin de trouver les caractéristiques comportant un lien significatif avec l'ouverture de session. Les caractéristiques provenaient de la base de sondage qui comprenait des données démographiques (indirectes) de l'ENM et des caractéristiques de type de logement et de ménage de la BSEM. Les rapports de cote d'ouverture de session pour les caractéristiques importantes sont présentés dans les figures 4-1 et 4-2.

Tant pour l'échantillon des ménages que pour celui des répondants cibles, les deux caractéristiques les plus importantes pour expliquer le taux d'ouverture de session étaient le mode de réponse à l'ENM (Internet ou une autre méthode), et le niveau de scolarité le plus élevé du ménage. Les cotes d'ouverture de session pour un ménage ou une personne étaient de 2 à 2,5 fois plus élevées si le ménage avait répondu à l'ENM par Internet que s'il avait répondu au moyen d'une autre méthode. Les cotes d'ouverture de session augmentaient en même temps que le niveau de scolarité, davantage pour l'échantillon des ménages que pour l'échantillon des répondants cibles. Dans l'ensemble, il y a des caractéristiques au niveau du logement, du ménage et de la personne qui sont importantes pour expliquer l'ouverture de session : certaines sont importantes pour expliquer l'ouverture de session, peu importe si cela vient du fait que n'importe qui du ménage peut ouvrir une session, ou du fait que seulement une personne sélectionnée en particulier peut ouvrir une session.

Figure 4-1
Rapports de cotes d'ouverture de session pour l'échantillon des ménages

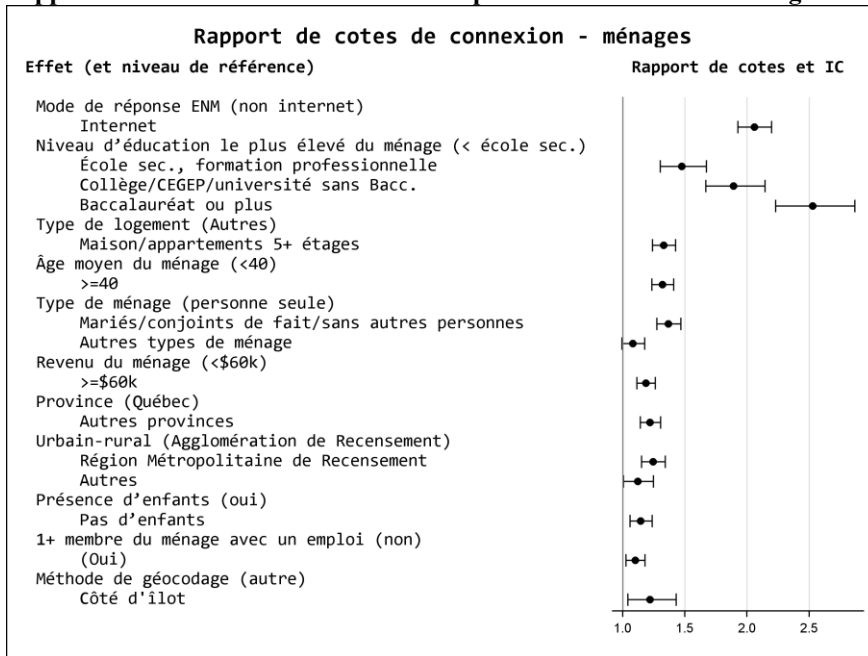
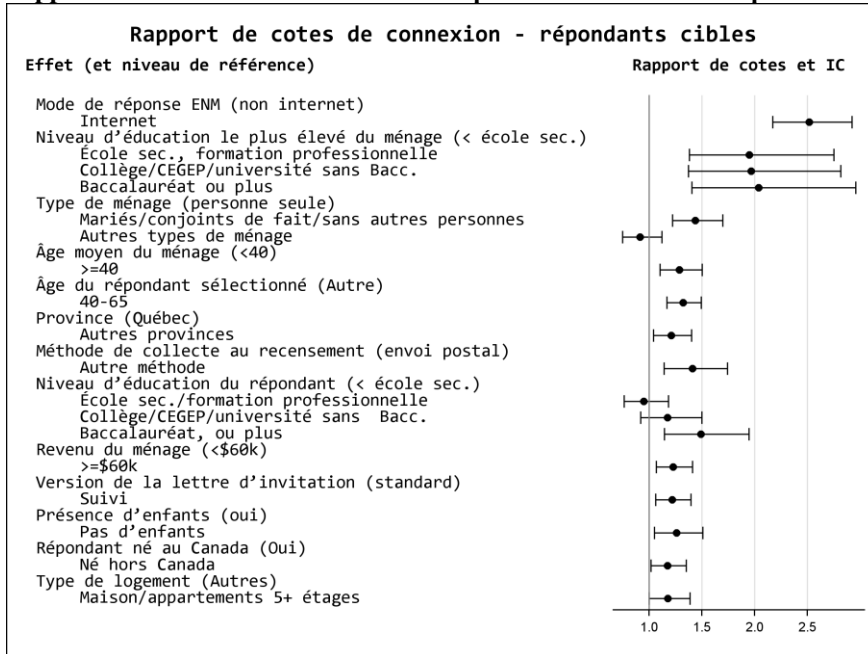


Figure 4-2
Rapports de cotes d'ouverture de session pour l'échantillon des répondants cibles



5. Sommaire

La présente étude explore l'idée de communiquer avec les unités échantillonnées pour les enquêtes auprès des ménages menées au moyen d'un QE à l'aide d'invitations envoyées par la poste. Deux stratégies de collecte ont été mises à l'essai et comparées. Les taux initiaux d'ouverture de session pour les stratégies étaient comparables, soit environ 30 %, mais le taux de soumission global était significativement plus élevé pour la stratégie des répondants cibles, en raison de l'absence de nouveau contact dans la stratégie des ménages. Toutefois, on ne sait pas clairement

si une stratégie est « meilleure » que l'autre, étant donné qu'il y a eu davantage de renvois par le bureau de poste pour l'échantillon des répondants cibles, ce qui indique que les noms des répondants figurant dans la base de sondage utilisée pour le contact étaient parfois périmés.

En outre, certains segments de la population étaient plus susceptibles de répondre selon ces stratégies dans cette étude. De façon plus particulière, ceux qui avaient répondu à l'ENM par Internet (comparativement à une autre méthode) et ceux qui avaient des niveaux de scolarité plus élevés étaient plus susceptibles de répondre.

Bibliographie

Gambino, J., G. Laflamme et M. Wu (2014). « Electronic questionnaires (EQ) for household surveys at Statistics Canada ». Rapport non publié, Ottawa, Canada : Statistique Canada.

Hamel, M. (2012). « Internet data collection in the Canadian census of population », Commission économique de l'Europe, Quatorzième réunion sur la collecte en ligne de données, mai 2012. Disponible en ligne à : www.unece.org.

Karaganis, M. et G. Laflamme (2012), « Use of E-questionnaires in Household Surveys ». Rapport non publié, Ottawa, Canada : Statistique Canada.

Tallon, P. et L. Ramsay (2010). « Une approche intégrée pour la collecte de données du Recensement au Canada », *Recueil du Symposium 2010 de Statistique Canada* : 275 à 281.