

Recueil du Symposium de 2024 de Statistique Canada : Le futur des statistiques officielles

Recrutement et collecte de données auprès de panels en ligne à Statistique Canada

par Krista MacIsaac, Cilanne Boulet et Marnie Thomas

Date de diffusion : le 8 septembre 2025



Statistique
Canada

Statistics
Canada

Canada

Recrutement et collecte de données auprès de panels en ligne à Statistique Canada

Krista MacIsaac, Cilanne Boulet et Marnie Thomas¹

Résumé

En 2020, Statistique Canada a commencé à utiliser des panels en ligne probabilistes comme méthode alternative de collecte de données pour des statistiques officielles. Dans le cadre d'un panel en ligne, on demande à des répondants d'une autre enquête de fournir leurs coordonnées en vue de participer à de futures enquêtes courtes. Cette présentation mettra en lumière l'expérience de Statistique Canada en matière de panels après quatre années d'utilisation, y compris les leçons tirées sur le recrutement des participants au panel et sur la manière de collecter ultérieurement des données à l'aide d'enquêtes par panel. La manière dont les questions de recrutement sont présentées peut entraîner des taux de participation très différents. En outre, la richesse des informations auxiliaires disponibles sur l'enquête de recrutement peut être utilisée pour gérer activement les opérations de collecte par panel, en prédisant la probabilité de réponse et en utilisant ces probabilités pour cibler les efforts de suivi.

Mots clés : Panel en ligne probabiliste; recrutement de membres d'un panel; gestion de la collecte adaptative.

1. Introduction

1.1 Panels en ligne probabiliste à Statistique Canada

Statistique Canada a exploré l'utilisation de panels en ligne. Le premier panel pilote a été créé en 2020 (Baribeau, 2020). L'utilisation de panels en ligne vise à recueillir des données opportunes pour combler les lacunes émergentes en matière de données sociales en disposant d'un groupe de personnes facilement joignables qui peuvent être contactées rapidement lorsque le besoin s'en fait sentir.

Les panels en ligne réalisés par Statistique Canada ont tous été des panels en ligne fondés sur une approche probabiliste, ce qui signifie que la probabilité que chaque unité soit sélectionnée pour l'échantillon de recrutement initial est connue, facilitant ainsi l'inférence sur la population cible. Ces enquêtes par panel se déroulent comme suit. Le recrutement d'un panel en ligne se fait dans le cadre d'une enquête probabiliste classique. Après avoir rempli le questionnaire de cette enquête, les répondants sont invités à indiquer s'ils souhaitent participer à des enquêtes ultérieures et à fournir leurs coordonnées. La série de panels est ensuite menée auprès des participants volontaires. Cette méthode comporte plusieurs avantages pour l'enquête par panel. Plus particulièrement, toutes les questions de l'enquête de recrutement peuvent être utilisées comme covariables afin de corriger les données de la non-réponse au panel, car elles sont disponibles pour toutes les unités du panel. Le recrutement du panel est également effectué sans entraîner de coûts supplémentaires pour la série de panels.

Les enquêtes par panel en ligne consistent en des questionnaires très courts qui prennent de 5 à 10 minutes à remplir. Elles font généralement partie d'une série, où chaque vague porte sur un sujet spécifique dans le cadre du thème collectif de la série. On communique avec les participants volontaires à l'aide de l'adresse électronique fournie, ce qui constitue une méthode de collecte de données rapide et abordable. En outre, Statistique Canada a commencé à faire un nombre limité de suivis téléphoniques pour cibler certains sous-ensembles de la population.

¹Krista MacIsaac, Statistique Canada, 150 promenade Tunney's Pasture, Ottawa (Ontario), Canada, K1A 0T6, krista.macisaac@statcan.gc.ca; Cilanne Boulet, Statistique Canada, 150, promenade Tunney's Pasture, Ottawa, (Ontario), Canada, K1A 0T6, Cilanne.boulet@statcan.gc.ca; Marnie Thomas, Statistique Canada, 150, promenade Tunney's Pasture, Ottawa, (Ontario), Canada, K1A 0T6, marnie.thomas@statcan.gc.ca.

Les deux premiers panels en ligne réalisés par Statistique Canada étaient la Série d'enquêtes sur les perspectives canadiennes (SEPC) et le Portrait de la société canadienne (PSC), qui visaient tous deux à produire des estimations pour la population canadienne en général. La SEPC comportait des participants recrutés à partir de l'Enquête sur la population active (EPA) et se déclinait en six vagues de questionnaires entre mars 2020 et janvier 2021. Les participants au PSC ont été recrutés dans le cadre de l'Enquête sociale générale (ESG) et le panel se déclinait en trois vagues de collecte de données entre mars 2021 et mai 2022. Le contenu de la SEPC et du PSC était lié à divers sujets concernant les répercussions sociales et économiques de la pandémie de COVID-19.

La série suivante de panels en ligne probabiliste de Statistique Canada ciblait certains sous-groupes de la population canadienne. L'objectif global de ces projets était de fournir des renseignements opportuns sur les populations d'intérêt. Le premier panel, la Série d'enquêtes sur les gens et leurs communautés (SEGC), ciblait les groupes racisés et les immigrants récents (Waye et Boulet, 2023). La partie panel de la SEGC a été recrutée au cours des première et deuxième vagues de la série d'enquêtes. La SEGC a été réalisée en cinq vagues entre octobre 2022 et avril 2024. Le contenu des cinq vagues était lié à la participation communautaire, à l'engagement politique et aux valeurs partagées, à la confiance envers les médias, à la participation aux sports et aux expériences de discrimination.

Les deux autres panels ciblés sont la Série d'enquêtes auprès des membres des Premières Nations, des Métis et des Inuit (SEMPNMI) et la Série d'enquêtes sur l'accessibilité (SEA). Ces panels ont été recrutés à partir de deux enquêtes postcensitaires : l'Enquête auprès des peuples autochtones (EAPA) et l'Enquête canadienne sur l'incapacité (ECI), respectivement. Chacun de ces panels comportait trois vagues de collecte commençant en 2024 et se terminant au début de 2025. La SEMPNMI a recueilli des renseignements sur l'accès aux soins de santé, les répercussions de la hausse des prix et la préparation aux situations d'urgence chez les Premières Nations, les Métis et les Inuit au Canada. La SEA a permis de recueillir des renseignements sur les obstacles à l'emploi, à la technologie et aux voyages rencontrés par les personnes souffrant de problèmes de santé de longue durée. Ces panels ont permis d'obtenir des renseignements supplémentaires sur de petits sous-groupes de la population dans un délai court de quelques mois seulement entre la collecte et la diffusion.

Le taux de réponse à une enquête par panel est mieux résumé par le taux de réponse cumulé qui combine la proportion de personnes ayant répondu à l'enquête de recrutement, la proportion de personnes ayant fourni des coordonnées et la réponse à la série d'enquêtes par panel elle-même (voir tableau 1.1-1). Par exemple, le taux de réponse cumulé de la SEPC était plus élevé que celui du PSC, principalement parce que le panel de la SEPC a été recruté à partir de l'EPA, une enquête obligatoire qui a donc un taux de réponse est plus élevé que celui de l'ESG, qui est volontaire. Le taux de réponse aux questionnaires du panel était également plus élevé dans le cadre de la SEPC. Cela s'explique probablement par son sujet et la période de collecte qui s'est déroulée au début de la pandémie de COVID-19. La proportion de répondants ayant fourni des coordonnées dans le cadre de la SEPC (97 %) était très élevée en raison de la conception de la question qui demandait des coordonnées. Cette question sera examinée plus en détail à la section 2. Étant donné que la pondération du panel peut intégrer des informations auxiliaires riches provenant de l'enquête de recrutement, les taux de réponse cumulés ne sont pas directement comparables aux taux de réponse des enquêtes réalisées en une seule phase. Toutefois, de faibles taux de réponse cumulés peuvent être associés à un risque de biais, ce qui est examiné par Mather, Boulet et Ra (2025).

Tableau 1.1-1
Taux de réponse aux panels en ligne

Panel	Enquête de recrutement	Taux de réponse à l'enquête de recrutement	Coordonnées fournies	Taux de réponse au panel ¹	Taux de réponse cumulé
Série d'enquêtes sur les perspectives canadiennes	Enquête sur la population active	87 %	23 %	De 54 % à 64 %	De 10,7 % à 12,6 %
Portrait de la société canadienne	Enquête sociale générale	40 %	22 %	De 41 % à 44 %	De 3,7 % à 3,9 %
Série d'enquêtes sur les gens et leurs communautés	Série d'enquêtes sur les gens et leurs communautés	53 %	97 %	De 24 % à 28 %	De 12,9 % à 15,4 %
Série d'enquêtes auprès des membres des Premières Nations, des Métis et des Inuit	Enquête auprès des peuples autochtones	52 %	21 %	À déterminer	À déterminer
Série d'enquêtes sur l'accessibilité	Enquête canadienne sur l'incapacité	61 %	25 %	À déterminer	À déterminer

1. Les taux de réponse aux panels sont présentés sous forme de fourchette, car les valeurs varient d'une vague à l'autre.

Le présent document se poursuit avec une description des apprentissages que l'on a tirés sur la manière d'obtenir efficacement les coordonnées des participants pour les futurs panels en ligne (section 2), puis de mener à bien la collecte par panel en ligne (section 3).

2. Recrutement des participants aux panels

2.1 Tests de questions de recrutement

Les participants aux panels en ligne sont recrutés à la fin du questionnaire d'une enquête probabiliste, juste avant la soumission du questionnaire. Les répondants sont invités à fournir leurs coordonnées (courriel et/ou numéro de téléphone) pour participer à de futures enquêtes. De juillet à août 2023, des essais ont été réalisés pour déterminer la formulation et la présentation optimales de la question afin de maximiser le nombre de personnes fournissant des coordonnées. Trois éléments différents ont été testés : le contenu du texte d'introduction et de sortie, l'inclusion d'un court paragraphe sur le caractère volontaire des futures enquêtes et la présence d'une question explicite d'acceptation. Chaque composante comportait deux options, soit un total de huit combinaisons. Les 20 000 unités sélectionnées de l'Enquête sociale canadienne (ESC) ont été utilisées pour le test et un huitième des personnes de l'échantillon a été assigné au hasard à chacune des huit options.

La formulation du texte d'introduction et de sortie explore différentes options afin de souligner l'importance de la participation et l'objectif des futures enquêtes. Le paragraphe facultatif expliquant la nature volontaire du panel indique aux répondants qu'ils pourront choisir de ne pas participer au panel après avoir fourni leurs coordonnées. Le troisième élément a permis d'évaluer l'effet d'un écran contenant une question de recrutement facultative. Dans la version à deux écrans, si le répondant accepte la question de recrutement, l'écran suivant lui demande ses coordonnées. Si le répondant refuse la question de recrutement, l'écran de collecte des coordonnées de suivi n'apparaît pas. L'autre option consiste à utiliser un seul écran et à demander des coordonnées après le texte d'introduction.

2.2 Résultats des tests de questions de recrutement

Les tests de questions de recrutement ont montré que le choix de la formulation de la description et de la question n'a produit qu'une différence marginale dans le taux de répondants fournissant des coordonnées (55,9 % comparativement à 56,9 %, tableau 2.2-1). Cependant, la présence de la question sur la volonté de participation a eu une incidence considérable, alors que 25,6 % des répondants fournissaient leurs coordonnées (courriel et/ou numéro de téléphone) lorsqu'ils étaient explicitement invités, sur un premier écran, à accepter de participer à des panels futurs avant même d'avoir eu la possibilité de fournir leurs coordonnées, comparativement à 87,8 % qui ont donné leurs coordonnées sans la question sur la volonté de participation. Ces résultats concordent avec les observations des panels précédents. Tel qu'il est illustré dans le tableau 1.1-1, le taux de répondants ayant fourni des coordonnées était de 21 à 25 % pour les panels qui demandaient d'abord si le répondant souhaitait participer à des panels futurs, et de 97 % pour la SEPC, où les répondants étaient simplement invités à fournir leur adresse électronique et leur numéro de téléphone. On ne sait pas encore si ces options entraîneront des taux différents de participation à un panel, car on n'a pas encore communiqué avec les personnes de ce test pour les inviter à participer à un panel.

Tableau 2.2-1 : Tailles d'échantillon, nombre de répondants et pourcentage de répondants qui ont fourni des coordonnées

Élément de test	Option	Échantillon	Répondants	Pourcentage de personnes ayant fourni des coordonnées
Tous les cas	-	20 000	9 288	56,4 %
Texte d'introduction et de sortie	Original	10 000	4 680	55,9 %
	Solution de rechange	10 000	4 608	56,9 %
Texte de participation volontaire	Original (aucun texte supplémentaire)	10 000	4 653	55,9 %
	Solution de rechange	10 000	4 635	56,8 %
Présence de la question sur la volonté de participation	Original (question sur la volonté de participation)	10 000	4 687	25,6 %
	Solution de rechange (aucune question sur la volonté de participation)	10 000	4 601	87,8 %

Le taux de répondants ayant fourni des coordonnées variait également en fonction du mode de réponse utilisé, soit par questionnaire autoadministré en ligne, soit par intervieweur au téléphone (tableau 2.2-2). Lorsque la question sur la volonté de participation était présente, 35,6 % des répondants fournissaient leurs coordonnées par téléphone, comparativement à 22,9 % en ligne, probablement encouragés par l'intervieweur à le faire. Lorsqu'un seul écran demandait les coordonnées, 96,4 % des répondants fournissaient leurs coordonnées dans le questionnaire électronique, comparativement à seulement 60,1 % par téléphone. Cela est probablement dû à la vérification avec avertissement présente dans le questionnaire, qui incite les répondants à fournir leurs coordonnées s'ils tentent de soumettre le questionnaire lorsque les champs sont vides. Les intervieweurs savaient qu'en cliquant une deuxième fois sur le bouton « suivant », ils pouvaient poursuivre le questionnaire.

**Tableau 2.2-2
Taux de réponse aux panels en ligne**

Présence d'une question sur la volonté de participation	Coordonnées fournies	
	En ligne	Par téléphone
Oui	22,9 %	35,6 %
Non	96,4 %	60,1 %

En ce qui concerne les futures enquêtes de Statistique Canada, les questions de recrutement de participants aux panels utiliseront la solution à écran unique, car elle a été la plus efficace pour obtenir des coordonnées. Cependant, la vérification avec avertissement qui demande aux répondants de fournir les informations manquantes ne sera pas incluse, car la participation aux panels est volontaire et cette vérification pourrait donner l'impression que ce n'est pas le cas.

3. Collecte de données auprès des panels en ligne

3.1 Aperçu de la stratégie de collecte

Les premiers panels en ligne réalisés par Statistique Canada, soit la SEPC et le PSC, étaient entièrement effectués en ligne. Les participants au panel recevaient une invitation par courriel à participer à l'enquête le premier jour de la collecte, puis trois à quatre rappels leur étaient envoyés par courriel au cours des une à deux semaines suivantes. Pour ces enquêtes par panel et celles ciblant une population spécifique, tous les courriels envoyés aux répondants contiennent le code d'accès sécurisé (CAS) nécessaire pour répondre à l'enquête. Cela signifie que les répondants peuvent répondre à l'enquête de n'importe où, dès qu'ils reçoivent le courriel, tant qu'ils ont accès à Internet.

La figure 3.1-1A montre le nombre de réponses par jour lors de la 3^e vague du PSC. Chaque fois qu'un courriel est envoyé, on observe un pic dans les taux de réponse, indiquant ainsi que les répondants répondent immédiatement aux courriels. Cette période de réponse est également relativement courte, le pic de réponse ne durant que de deux à trois jours. Pour cette raison, il est possible d'envoyer les courriels de manière rapprochée. En général, le nombre de réponses diminue à chaque nouveau courriel de rappel.

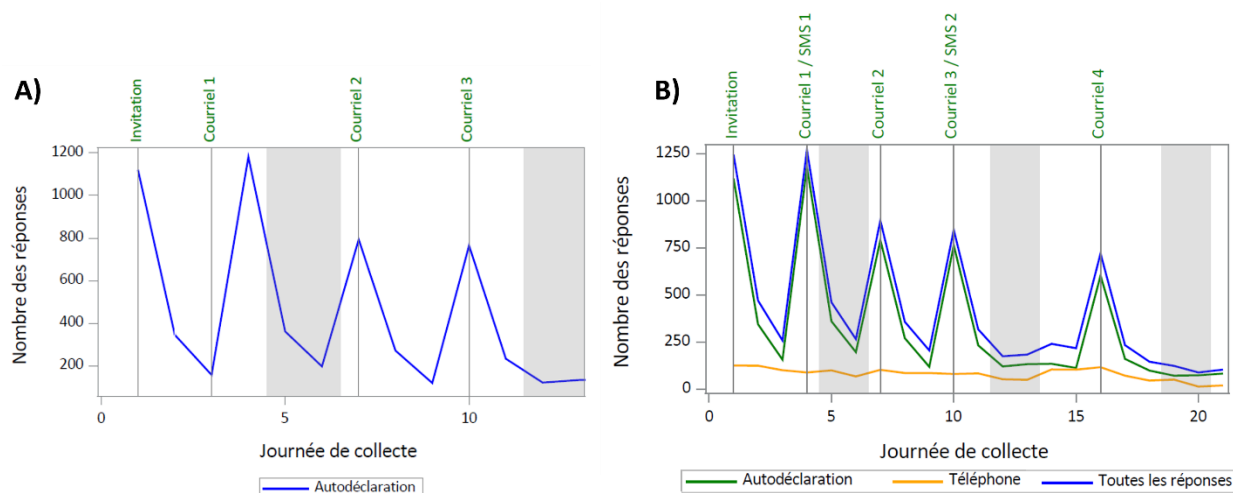


Figure 3.1-1 : Nombre de réponses par jour pour A) le Portrait de la société canadienne – vague 3 et B) la Série d'enquêtes sur les gens et leurs communautés – vague 5. Les barres grises indiquent les fins de semaine.

Les panels en ligne ciblés, à partir de la SEPC, ont également utilisé les numéros de téléphone des personnes recrutées pour le panel. Cela visait à assurer une meilleure couverture et une meilleure réponse des sous-groupes ciblés par les panels. Étant donné le coût plus élevé des interviews téléphoniques, seules les unités qui n'ont pas fourni d'adresse électronique ont commencé la collecte de données par téléphone. Toute personne ayant fourni une adresse électronique a reçu le courriel d'invitation au début de la collecte, et un nombre limité de ces cas a fait l'objet d'un suivi téléphonique pendant la deuxième moitié de la collecte (voir la section 3.2 pour plus de détails). Dans l'ensemble, les taux de réponse les plus élevés ont été obtenus auprès des personnes ayant fourni à la fois une adresse électronique et un numéro de téléphone (tableau 3.1-1). Le fait qu'une personne ait fourni les deux types de coordonnées pourrait indiquer qu'elle est plus encline à répondre aux enquêtes par panel. La proportion de répondants ayant répondu par téléphone est relativement faible. Cependant, avoir reçu un appel de suivi a pu encourager certains répondants à répondre au questionnaire en ligne.

Tableau 3.1-1 : Taux de réponse des vagues 3 et 5 de la SEPC selon le mode de réponse et la disponibilité des coordonnées

Coordonnées disponibles	Taux de réponse					
	SEPC – Vague 3			SEPC – Vague 5		
	En ligne	Par téléphone	Total	En ligne	Par téléphone	Total
Par téléphone seulement (20 %)	0,4 %	27,8 %	28,2 %	0,1 %	20,9 %	21,0 %
Par courriel seulement (14 %)	23,2 %	0,0 %	23,3 %	17,4 %	0,0 %	17,4 %
Par téléphone et par courriel (66 %)	27,9 %	1,3 %	29,2 %	25,0 %	2,8 %	27,7 %

La SEPC a également testé l'ajout de rappels par messagerie texte. Le message texte a été envoyé le même jour que le courriel aux personnes ayant fourni un numéro de téléphone cellulaire et une adresse électronique. Le message texte indiquait simplement aux répondants de vérifier leurs courriels où ils trouveraient le CAS et le lien pour accéder au questionnaire. La figure 3.1-1B illustre le nombre de réponses par jour pour la vague 5 de la SEPC, dans le cadre de laquelle des rappels par messagerie texte ont été envoyés avec le premier et le troisième rappel par courriel. Contrairement au modèle de réponse de la figure 3.1-1A, le nombre de réponses lors de l'envoi du message texte (figure 3.1-1B) n'a pas diminué par rapport au rappel précédent, mais est plutôt demeuré à peu près le même. En tenant compte des taux de réponse observés parmi les unités avec et sans numéro de téléphone cellulaire, on a estimé que l'augmentation globale des taux de réponse attribuable aux rappels par messagerie texte s'élevait à environ 2,8 %.

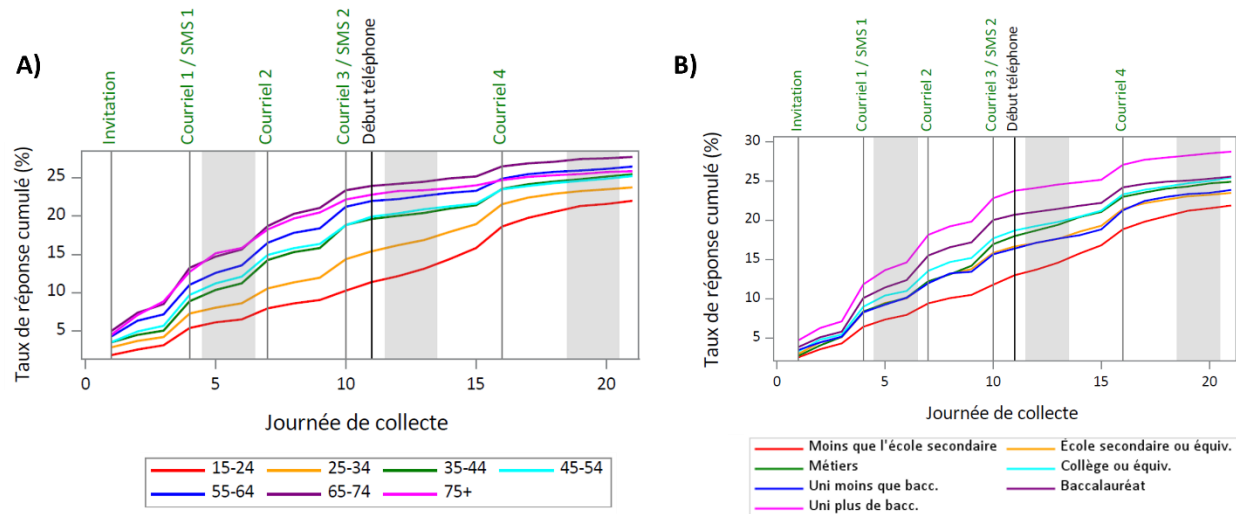
3.2 Gestion de la collecte adaptative

Comme mentionné, la collecte de données est lancée par l'envoi d'une invitation par courriel pour tous les cas disposant d'une adresse électronique. Cela vise à inciter les répondants à répondre par le mode le moins coûteux. Étant donné que le taux de réponse varie selon les groupes démographiques, notamment en fonction de l'âge et du niveau de scolarité, un suivi téléphonique est utilisé pour cibler les groupes ayant les taux de réponse les plus faibles. Cela est réalisé en utilisant un modèle de régression logistique pour estimer la probabilité de réponse en fonction des renseignements sociodémographiques disponibles dans la base de sondage. Puisque les personnes recrutées aux fins des enquêtes par panel sont des répondants à l'enquête de recrutement, ces renseignements sont connus pour toutes les unités de l'échantillon. La disponibilité de ces informations auxiliaires pour toutes les unités de l'échantillon constitue l'un des avantages des enquêtes par panel.

Au cours d'une période de collecte de données de trois semaines, la première semaine est généralement consacrée à l'auto-déclaration et le modèle de régression logistique est ajusté aux données de la vague en cours au début de la deuxième semaine de collecte. L'ensemble exact des variables incluses dans le modèle dépend des données actuelles et varie d'une vague à l'autre. Les variables les plus couramment incluses dans le modèle sont le groupe d'âge, le plus haut niveau de scolarité atteint, le sous-groupe de population d'intérêt (groupe racisé pour la SEPC ou identité autochtone pour la SEMPNI), et l'emplacement géographique (province ou région). Le modèle est ajusté en fonction de toutes les unités de l'échantillon, indépendamment des coordonnées disponibles. Les unités ayant la probabilité de réponse la plus faible, qui ont fourni à la fois un courriel et un numéro de téléphone, sont ensuite admissibles à faire l'objet d'appels de suivi téléphonique.

Pour illustrer comment cela fonctionne, nous prendrons la 5^e vague de la SEPC, qui s'est déroulée sur une période de trois semaines, du 2 au 22 avril 2024 (voir la figure 3.2-1). L'échantillon contenait 35 468 unités, dont 23 328 ont fourni à la fois un courriel et un numéro de téléphone. Le modèle logistique a été ajusté aux données le 10^e jour de la collecte, en travaillant avec les 35 468 unités. À ce stade, les taux de réponse variaient de 10,3 % pour le groupe d'âge de 15 à 24 ans à 23,4 % pour le groupe d'âge de 65 à 74 ans. Étant donné que le groupe d'âge est fortement lié au mécanisme de réponse, il est utilisé comme prédicteur dans le modèle de régression. Les 6 669 cas dont la propension estimée à répondre était la plus faible parmi les 23 328 cas ayant fourni à la fois un courriel et un numéro de téléphone ont été sélectionnés pour faire l'objet d'un suivi téléphonique. Les groupes d'âge de 15 à 24 ans et de 25 à 34 ans représentaient respectivement 14,1 % et 18,4 % de l'échantillon, mais représentaient 43 % et 30,4 % des cas retenus

pour le suivi téléphonique. À la dernière journée de la collecte, la différence dans les taux de réponse par groupe d'âge avait diminué de plus de la moitié, pour atteindre 5,7 % (22 % pour le groupe d'âge de 15 à 24 ans et 27,7 % pour le groupe d'âge de 65 à 74 ans). On a observé une tendance similaire pour le plus haut niveau de scolarité atteint, où l'écart des taux de réponse a été réduit de 11 points de pourcentage à 6,9.



4. Conclusion

Les panels en ligne probabiliste sont un moyen pour Statistique Canada de recueillir des données rapidement et d'en diffuser les résultats de manière opportune. Depuis leur première introduction comme option de collecte de données en 2020, des essais ont été menés pour mieux comprendre les facteurs liés à la non-réponse aux panels en ligne et élaborer des méthodes pouvant être utilisées pour augmenter le taux de participation. Bien que les panels restent un outil qui affiche des taux de réponse cumulés faibles par rapport aux enquêtes plus traditionnelles, ces méthodes peuvent aider à augmenter le taux de participation et à égaliser les taux de réponse entre les sous-groupes.

5. Remerciements

Ces travaux sont le fruit des efforts de nombreuses personnes sans qui la mise en œuvre de ces panels en ligne n'aurait pas été possible. Plus particulièrement, Youngchang Ra, Marc Beauparlant et Jacquelyn Sun ont participé à la gestion de la collecte et aux premières versions des évaluations présentées dans ce document. Les auteurs souhaitent par ailleurs remercier Cindy Ubartas et John Stardom qui ont fourni des commentaires sur ce document et ont contribué à l'améliorer.

Bibliographie

- Baribeau, B. (2020), "Rompre avec la tradition: pilote d'un panel web probabiliste (anglais seulement). Document de travail de la Direction de la méthodologie DMSS-2020-002. Statistique Canada.
- Mather, A., Boulet, C., et Ra, Y. (2025), "Évaluation du biais des panels en ligne probabilistes de Statistique Canada", *Recueil du Symposium de 2024 de Statistique Canada*.

Waye, A. et Boulet, C. (2023), “Série d’enquêtes sur les gens et leurs communautés”, *Recueil du Symposium de 2022 de Statistique Canada*.