

**Recueil du Symposium de 2022 de Statistique Canada :  
Désagrégation des données : dresser un portrait de données plus représentatif  
de la société**

**Série d'enquêtes sur les gens et leurs  
communautés**

par Agnes Waye et Cilanne Boulet

Date de diffusion : le 25 mars 2024



Statistique  
Canada

Statistics  
Canada

Canada

## Série d'enquêtes sur les gens et leurs communautés

Agnes Waye et Cilanne Boulet<sup>1</sup>

### Résumé

En 2021, Statistique Canada a lancé le Plan d'action sur les données désagrégées, une initiative pluriannuelle pour renforcer les méthodes de collecte de données en les rendant plus représentatives, améliorer les statistiques sur diverses populations afin de mener des analyses intersectionnelles et appuyer les efforts du gouvernement et de la société visant à atténuer les inégalités connues et à intégrer les facteurs d'égalité et d'inclusion dans la prise de décisions. Dans le cadre de cette initiative, nous montons la Série d'enquêtes sur les gens et leurs communautés, un nouveau panel probabiliste conçu précisément pour recueillir des données qui peuvent être désagrégées selon le groupe racisé. Ce nouvel outil nous permettra de combler des lacunes dans les données et de répondre aux nouvelles questions relatives à la diversité. L'article présentera un aperçu de la conception de la Série d'enquêtes sur les gens et leurs communautés.

Mots clés : données désagrégées; panel; diversité; analyses intersectionnelles.

## 1. Introduction

### 1.1 Contexte

Depuis quelques années, la demande de données désagrégées appuyant l'élaboration de politiques fondées sur des données probantes est croissante. En 2021, Statistique Canada a lancé le Plan d'action sur les données désagrégées (PADD). Le PADD a pour objectif d'appuyer la collecte de données afin de refléter la diversité de la population canadienne et de fournir des renseignements essentiels sur les groupes sous-représentés. Il vise également à donner des renseignements statistiques détaillés sur des groupes de population précis, comme les femmes ou les groupes racisés. Dans ce contexte, il est important de disposer de renseignements à un niveau très détaillé pour les analyses intersectionnelles.

### 1.2 Série d'enquêtes sur les gens et leurs communautés

La Série d'enquêtes sur les gens et leurs communautés (SEGC) est l'une des initiatives entreprises dans le cadre du PADD. La SEGC est une série de quatre enquêtes sociales portant sur divers sujets comme l'engagement communautaire, la confiance envers les gens et la confiance dans les institutions canadiennes. La SEGC a plusieurs objectifs clés. Premièrement, elle vise à mesurer l'expérience des groupes racisés et des immigrants au Canada. Son objectif secondaire est de produire des données désagrégées concernant des groupes ciblés par région géographique. Elle vise aussi à comparer les groupes racisés et les groupes non racisés. La SEGC vise à combler les lacunes statistiques et les besoins de données concernant divers groupes afin d'appuyer différents programmes et services gouvernementaux canadiens.

## 2. Plan de sondage

### 2.1 Aperçu de l'enquête

---

<sup>1</sup>Agnes Waye, Statistique Canada, 150, promenade Tunney's Pasture, Ottawa, Canada, K1A 0T6 (agnes.waye@statcan.gc.ca); Cilanne Boulet, Statistique Canada, 150, promenade Tunney's Pasture, Ottawa, Canada, K1A 0T6 (cilanne.boulet@statcan.gc.ca)

Le plan de la SEGC est un panel probabiliste trimestriel. La signification du terme « panel » varie selon le contexte. Dans le cadre de la SEGC, un panel consiste à recruter un groupe de personnes à partir d'un échantillon probabiliste initial qui fournissent leurs coordonnées en vue de participer aux enquêtes ultérieures de la série. Les personnes sélectionnées dans l'échantillon ne restent pas toutes dans le panel; seules les unités qui acceptent de participer et donnent leurs coordonnées demeurent dans le panel. Ces unités d'échantillonnage de la SEGC seront contactées jusqu'à trois autres fois au cours d'une année.

Le principal avantage de la création d'un panel est qu'on dispose d'un groupe de personnes facilement accessibles et joignables rapidement en cas de nouveaux besoins en données. Les panels sont extrêmement utiles pour joindre rapidement des populations. Statistique Canada en a fait l'expérience directe lorsqu'il a mené son premier panel en ligne expérimental au début de la pandémie de COVID-19. De plus amples renseignements sur cette première expérience des panels en ligne se trouvent dans Baribeau (2020).

Avec la SEGC, Statistique Canada utilise pour la première fois un panel probabiliste qui vise à atteindre un segment précis de la population canadienne. Afin de représenter les expériences des groupes racisés et des immigrants au Canada, l'échantillon de la SEGC est stratifié par groupe racisé, statut d'immigration et région, et la taille des échantillons est déterminée de manière à ce que les objectifs de désagrégation puissent être atteints pour des groupes suffisamment grands. Une attention particulière a été accordée à éviter un suréchantillonnage excessif dans les petites strates pour que d'autres initiatives du PADD puissent également être mises en œuvre sans imposer un fardeau de réponse excessif. L'échantillon comprenait un groupe témoin composé de personnes qui ne font pas partie de groupes racisés. La présence du groupe témoin permet d'établir des comparaisons entre les membres des groupes racisés et ceux des groupes non racisés. Le groupe témoin a également été stratifié selon le statut d'immigration, avec un suréchantillonnage d'immigrants. Cette mesure était nécessaire pour permettre des comparaisons avec l'expérience des immigrants qui constituent une partie importante des groupes racisés ciblés.

## **2.2 Population cible et base de sondage**

La population cible de la SEGC est la population canadienne de 15 ans et plus qui vit dans l'une des 10 provinces canadiennes. En pratique, la population sondée est légèrement plus petite que la population cible. Les personnes qui vivent dans les établissements institutionnels et les réserves sont exclues de l'échantillon.

La base de sondage de cette enquête est composée des répondants au questionnaire détaillé du Recensement de la population canadienne de 2021. (Le questionnaire abrégé du Recensement de la population canadienne a été mené pour l'ensemble de la population canadienne, tandis que le questionnaire détaillé du recensement a été réalisé sur 25 % de la population canadienne environ.) Le questionnaire détaillé du recensement posait des questions précises sur de nombreux sujets, y compris l'identité de groupe racisé et culturel, la scolarité, l'état matrimonial, etc. De plus, le statut d'immigrant est disponible pour les répondants au questionnaire détaillé, car il était intégré par couplage à des données administratives dans le cadre du traitement des données du recensement. Le questionnaire détaillé du recensement fournit donc une base naturelle à la SEGC, car il permet un suréchantillonnage ciblé qui permet de désagréger les données par groupes racisés, par région géographique, par statut d'immigrant (non-immigrant, immigrant récent et non-immigrant).

## **2.3 Vagues de la SEGC**

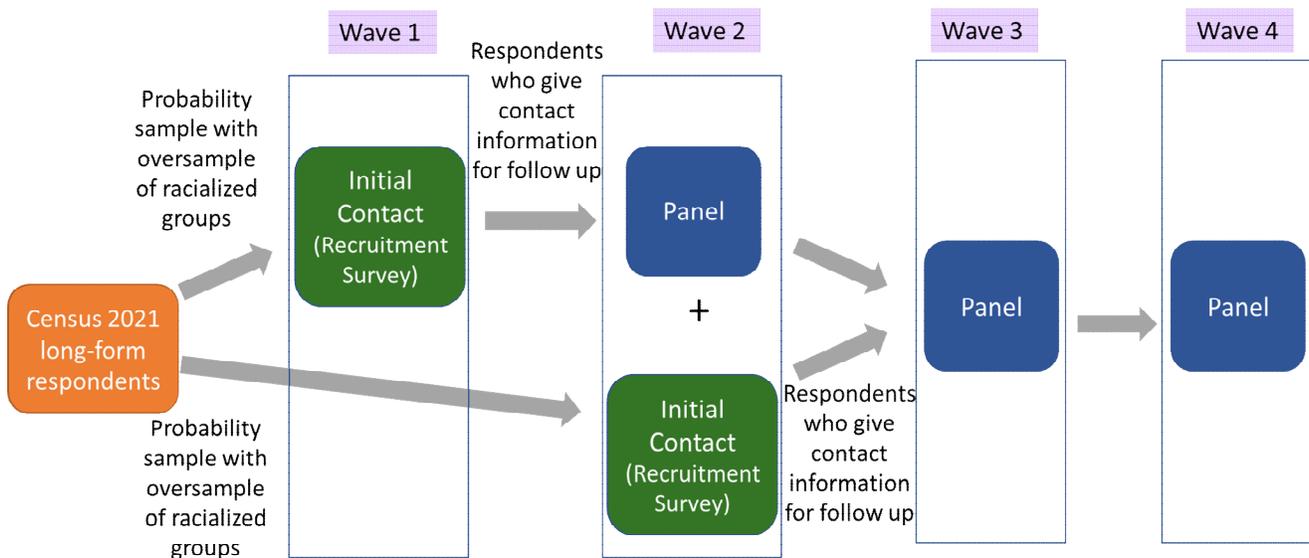
La SEGC compte quatre vagues, chacune ayant un questionnaire différent. Le premier contact a lieu à la vague 1 ou 2 et sert à recruter les panélistes avec lesquels il faudra communiquer aux vagues ultérieures. En raison de contraintes opérationnelles, l'enquête de recrutement de la SEGC a été divisée en deux vagues, au lieu d'avoir une seule vague de recrutement. Cette division du recrutement en deux vagues permettra également des ajustements au plan d'échantillonnage à la vague 2, au besoin.

L'échantillon de la vague 1 est tiré de la base de sondage avec un suréchantillonnage des groupes racisés. Un premier contact est établi avec les unités de l'échantillon; on leur demande de répondre à une enquête et, à la fin de l'enquête, on leur demande leurs coordonnées (courriel et numéro de téléphone) pour qu'ils puissent participer aux prochaines enquêtes de la série. Les répondants de la vague 1 qui donnent leurs coordonnées pour participer aux questionnaires de suivi sont contactés de nouveau aux vagues 2, 3 et 4.

À la vague 2, un nouvel échantillon sans chevauchement avec celui de la vague 1 est sélectionné à partir de la base de sondage. On demande à cet échantillon de répondre à une enquête comportant des questions des vagues 1 et 2 et, à la fin de l'enquête, on leur demande également leurs coordonnées (courriel et numéro de téléphone) pour qu'ils puissent participer aux futures enquêtes de la série. Les répondants de la vague 2 qui donnent leurs coordonnées pour participer aux questionnaires de suivi seront contactés de nouveau aux vagues 3 et 4. À la vague 2, en plus de ce nouvel échantillon, nous communiquerons également avec les unités de la vague 1 qui nous ont donné leur numéro de téléphone ou leur adresse courriel et qui ont accepté de participer aux questionnaires de suivi; ce groupe ne reçoit que les questions de la vague 2.

Pour les vagues 3 et 4, nous communiquerons avec toutes les personnes qui ont donné leurs coordonnées pour participer au suivi des deux vagues. La figure 2.3-1 ci-dessous résume les diverses vagues de la SEGC.

**Figure 2.3-1**  
**Vagues de la série d'enquêtes sur les gens et leurs communautés**



## 2.4 Mode de collecte

La collecte de la vague 1 a lieu d'octobre 2022 à janvier 2023. Le mode de collecte des vagues 1 et 2 est principalement un questionnaire à remplir soi-même en ligne, ainsi qu'un suivi de non-réponse par interview téléphonique assistée par ordinateur (ITAO). La collecte aux vagues 3 et 4, auprès des participants au panel ayant fourni des coordonnées personnelles lors de l'enquête de recrutement, devrait être beaucoup plus rapide. Elle est également réalisée à la fois par questionnaire en ligne et par ITAO.

Le soutien disponible aux fins de collecte dans des langues non officielles, comme l'espagnol et l'arabe, est important, car il permet de joindre les répondants qui ne sont peut-être pas en mesure de communiquer en anglais ou en français.

## 2.5 Contenu de l'enquête

Pour la vague 1 de la SEGC, nous avons posé plusieurs questions sociodémographiques (comme le niveau de scolarité et l'état matrimonial) ainsi que des questions concernant certains sujets, par exemple la participation à des groupes (comme des clubs et des associations), le sentiment d'appartenance à la collectivité locale et la confiance dans les institutions.

Il est important de noter que certaines questions posées dans le questionnaire détaillé du recensement se chevauchent avec celles de la SEGC. Dans le cadre de la SEGC, nous avons posé des questions pour des variables dont la réponse a pu changer depuis le recensement. Par exemple, la scolarité ou l'état matrimonial a pu changer au cours de la période d'un an et demi écoulée entre le recensement et les périodes de collecte de la SEGC. La question sur l'identité de groupe racisé a été posée dans la SEGC, même si elle a déjà été posée dans le questionnaire détaillé du recensement, parce que le concept d'identité de groupes racisés peut être fluide et changer au fil du temps ou dans différentes situations.

Outre le temps qui s'est écoulé entre les périodes de collecte du questionnaire détaillé du recensement et celles de la SEGC, il a été décidé de reposer certaines questions dans le questionnaire de la SEGC pour obtenir des réponses plus exactes. En effet, dans le questionnaire détaillé du recensement, une personne peut répondre aux questions au nom de tous les membres du ménage, de sorte qu'il est possible que la personne échantillonnée aux fins de la SEGC donne une réponse différente de celle qui figure dans le questionnaire détaillé du recensement.

Le tableau 2.5-1 ci-dessous présente les variables provenant du questionnaire détaillé du recensement et de la vague 1 de la SEGC. Le contenu de l'enquête pour les vagues suivantes n'est pas encore définitif, mais des sujets comme la discrimination, l'engagement politique et la participation à des sports sont envisagés.

**Tableau 2.5-1**  
**Comparaison des variables du recensement et de la SEGC**

Variable	Recensement	SEGC Vague 1
Âge, sexe, genre	x	x
État civil	x	x
Revenu	x	
Langue	x	
Groupes racisés	x	x
Immigration	x	
Scolarité	x	x
Étudiants	x	x
Orientation sexuelle		x
Incapacité	x	x

## 2.6 Plan de sondage de la vague 1

À la vague 1 de la SEGC, 40 000 unités ont été échantillonnées. Il fallait considérer diverses contraintes lors de l'échantillonnage, qui s'expliquent par le fait que d'autres enquêtes de Statistique Canada souhaitent aussi sonder la même population cible (groupes racisés). Il était important de garder cela à l'esprit au moment de déterminer la taille de l'échantillon de la SEGC.

Parce que l'objectif de la SEGC est d'obtenir des données pour des groupes de population très précis, il a fallu créer un grand nombre de strates. Certaines de ces strates sont relativement petites. La taille limitée de l'échantillon global et la petite taille des strates signifiaient que, dans certaines de ces strates, nous avons très peu d'unités à échantillonner. Il se peut donc qu'il soit impossible de produire des estimations fiables pour toutes les strates. Par exemple, certains groupes racisés comptent très peu de membres dans certaines régions, comme l'Atlantique. De nombreuses estimations au niveau régional peuvent être produites pour les vagues de recrutement. En revanche, pour les vagues par panel, dans lesquelles le nombre de participants est plus faible, le plan de sondage vise seulement des estimations au niveau national. Ces contraintes ont rendu l'échantillonnage aux fins de la SEGC assez difficile, car il fallait adopter un plan d'échantillonnage qui garantissait un certain nombre d'unités d'échantillonnage pour chaque domaine d'intérêt dans lequel des données désagrégées doivent être produites.

Une partie de l'échantillon de la SEGC a été sélectionnée au moyen d'une autre méthode, décrite à la section 3.2.

## 2.7 Taux de réponse attendu

Nous avons pu calculer un taux de réponse attendu pour la SEGC en nous fondant sur l'expérience acquise lors d'autres enquêtes similaires réalisées par Statistique Canada. Quand nous communiquons avec les unités de l'échantillon pour la première fois, nous nous attendons à ce que 50 % d'entre elles répondent à l'enquête de recrutement. Parmi celles qui répondent, nous nous attendons à ce que seulement 50 % donnent leurs coordonnées pour participer aux prochaines vagues. Pour les vagues par panel, nous nous attendons à un taux de réponse de 40 %. Par conséquent, le taux de réponse cumulatif devrait être d'environ  $50\% * 50\% * 40\% = 10\%$ . Ce taux de réponse cumulatif est attendu pour les unités recrutées à la vague 1 et à la vague 2. Ce taux de réponse attendu est assez faible et peut être préoccupant pour ce qui est du biais de non-réponse. Cependant, nous disposons d'un très large éventail de variables tirées du questionnaire détaillé du recensement qui peuvent nous aider considérablement dans notre modèle d'ajustement pour la non-réponse à l'étape de la pondération. Malgré ce faible taux de réponse, nous maintenons l'utilisation de panels, car comme nous l'avons dit plus haut, ils présentent un avantage crucial, la rapidité d'obtention des renseignements ciblés. Concernant les répondants de la vague 1 de la SEGC, nous constatons que le pourcentage de personnes qui ont donné leurs coordonnées jusqu'à maintenant est beaucoup plus élevé que 50 %, ce qui signifie que le taux de réponse cumulatif estimé ci-dessus est probablement conservateur.

## 3. Défis et leçons tirées

### 3.1 Coordonnées

Pour la vague 1 de la SEGC, l'ajout des coordonnées récentes des unités de l'échantillon a constitué un défi de taille. Les coordonnées sont disponibles dans le recensement, mais comme la collecte de la SEGC a eu lieu plus d'un an après la date de référence du recensement, nous avons essayé de mettre à jour les coordonnées en intégrant de nouvelles informations provenant de différentes sources.

Pour les adresses, nous avons classé plusieurs sources par ordre de priorité en fonction de la qualité des données, de la date de la dernière mise à jour de l'information et de la cohérence entre les différentes sources. Nous avons accordé la priorité aux adresses qui se trouvaient dans plusieurs sources. Parce que les immigrants récents peuvent avoir tendance à déménager plus souvent, il est parfois plus difficile d'obtenir leurs coordonnées de manière fiable. Pour les prochaines vagues, nous pourrions étudier la possibilité de consulter des sources d'adresses provenant de programmes spéciaux proposant un soutien aux nouveaux immigrants.

Nous avons nettoyé et normalisé les adresses, les numéros de téléphone, les noms et les adresses de courriel, ce qui a nécessité des efforts considérables puisque nous avons combiné de nombreuses sources de données différentes. Nous avons dû tenir compte de considérations particulières quand les noms étaient d'origine autre qu'anglaise ou française. Nous avons également utilisé ces sources pour déterminer les unités de l'échantillon qui sont décédées ou ont émigré depuis la date du recensement.

### 3.2 Indicateur de substitution de l'origine ethnique

Une partie des unités de l'échantillon de la SEGC était fondée sur l'indicateur de substitution de l'origine ethnique (ISOE). L'ISOE est une initiative de recherche et développement expérimentale dans le cadre du Programme d'élaboration de données sociales longitudinales de Statistique Canada. L'ISOE est un algorithme utilisé pour attribuer directement, indirectement ou par modélisation l'identité de groupe racisé au niveau individuel pour l'ensemble de la population canadienne. Cet indicateur est fondé sur des données administratives et des données historiques du recensement.

Il faut garder à l'esprit que l'échantillon principal de la SEGC est tiré du questionnaire détaillé du recensement, qui couvre 25 % de la population canadienne. L'un des avantages de l'utilisation de l'ISOE est qu'il peut être défini pour l'ensemble de la population canadienne. De plus, il peut être produit chaque année, tandis que le recensement est réalisé tous les cinq ans. Par conséquent, sur le plan de la couverture et de l'actualité, il présente des avantages évidents aux fins de création d'une base de sondage qui comprend des renseignements sur l'appartenance probable à un groupe racisé.

La SEGC a intégré un échantillon fondé sur l'ISOE dans le cadre d'une expérience à la vague 1. À la vague 1, 12,5 % de l'échantillon était constitué d'unités sélectionnées à partir du questionnaire abrégé du recensement et auxquelles on avait attribué une identité de groupe racisé fondée sur l'ISOE. Étant donné que la question sur l'identité de groupe racisé est posée dans la SEGC, les réponses contribueront à valider l'ISOE et à évaluer la possibilité d'utiliser cet indicateur aux fins d'échantillonnage ciblé dans de futures enquêtes. Si nous constatons que l'ISOE attribue efficacement une identité de groupe racisé, nous serons en mesure de tirer des échantillons à partir de l'ensemble de la population canadienne plutôt que seulement 25 % des Canadiennes et Canadiens (ce qui est actuellement le cas pour la base de sondage de la SEGC à la vague 1). Le fait de couvrir l'ensemble de la population canadienne permettra de réduire les problèmes mentionnés à l'étape de l'échantillonnage, particulièrement quand les tailles de strate sont petites.

### **3.3 Conclusion**

Les enquêtes par panel peuvent être des outils très puissants pour recueillir des données en temps opportun, et la SEGC adapte cette méthode à la collecte de renseignements sur les expériences des Canadiennes et Canadiens appartenant à des groupes racisés. La mise en œuvre de ces enquêtes s'accompagne de défis et de limites particuliers. Ce constat nous incite à chercher les nombreuses possibilités d'amélioration et de changement en vue des futures vagues de la SEGC. La vague 1 servira de phase exploratoire à partir de laquelle des leçons seront tirées pour les itérations futures.

### **Bibliographie**

Baribeau, B. (2020), « Breaking with Tradition: Piloting a Probability Web Panel », document interne, Statistique Canada.