

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

### **Faire progresser la collecte de données équitables : aperçus du projet de recherche sur le Plan d'action sur les données désagrégées de la Division des méthodes d'intégration statistique de Statistique Canada**

par Andrew Pearce, Kenza Sallier et Christiane Laperrière

Date de diffusion : le 8 septembre 2025



Statistique  
Canada

Statistics  
Canada

Canada

## **Faire progresser la collecte de données équitables : aperçus du projet de recherche sur le Plan d'action sur les données désagrégées de la Division des méthodes d'intégration statistique de Statistique Canada**

Andrew Pearce, Kenza Sallier et Christiane Laperrière<sup>1</sup>

### **Résumé**

Pour répondre à l'appel à l'action concernant les 17 objectifs de développement durable des Nations Unies (ONU), et pour relever les défis sociaux, économiques et en matière d'équité au Canada, un financement a été accordé à la phase d'élaboration quinquennale de Statistique Canada pour le Plan d'action sur les données désagrégées (PADD) en 2021, afin de soutenir les décisions fondées sur les données concernant ces défis. Ainsi, le document s'intitulant « Principes directeurs : utiliser les données du Recensement de la population de 2021 pour produire des statistiques sur les groupes d'intérêt du Plan d'action sur les données désagrégées » a été rédigé. Ce document explique le cadre organisationnel du PADD au sein de l'organisme, décrit les sources de données existantes, tient compte des préoccupations d'ordre éthique et au chapitre de la protection des renseignements personnels, centralise les méthodes d'échantillonnage adaptées aux initiatives du PADD, en plus de tenir compte des caractéristiques qui peuvent compliquer les procédures d'échantillonnage et de collecte de données.

Mots clés : Désagrégation des données; échantillonnage; populations rares; Recensement de 2021.

### **1. Introduction**

#### **1.1 Contexte organisationnel**

Dans le cadre du Programme de développement durable à l'horizon 2030 de l'ONU, un appel à l'action mondial a été lancé pour tenir compte des 17 objectifs de développement durable. Parmi ceux-ci, mentionnons la réduction des inégalités, l'égalité des genres et l'élimination de la pauvreté (ONU, 2024). Dans le même ordre d'idées, au Canada, des mouvements sociaux à l'appui des droits autochtones, de la justice raciale et de l'équité économique sont à l'avant-plan de la société. Qu'il s'agisse de défis internationaux ou canadiens, dans les deux cas, il faut recourir à des statistiques officielles pour orienter les processus décisionnels les touchant. Pour tenir compte de ces besoins, le gouvernement du Canada a financé la phase d'élaboration quinquennale du Plan d'action sur les données désagrégées (PADD) de Statistique Canada en 2021.

En ce qui concerne les mesures concrètes, une stratégie de données désagrégées, comme le PADD de Statistique Canada, renvoie à une approche systématique axée sur la collecte, l'analyse et la diffusion de données sur un sujet particulier, en divisant les données en catégories plus petites et plus granulaires. En tant qu'organisme national de la statistique, Statistique Canada gouverne la mise en œuvre de sa stratégie de données désagrégées en ayant recours à des méthodes, à des normes et à des procédures de contrôle de la qualité rigoureuses, chacune visant à garantir l'exactitude et la fiabilité des statistiques produites, tout en tenant compte des facteurs relatifs à la protection des renseignements personnels, à la confidentialité et à l'éthique. Ainsi, le PADD de Statistique Canada porte sur tous les aspects du cycle de vie des données, ou le Modèle générique du processus de production statistique (CEE-ONU, 2019).

---

<sup>1</sup> Andrew Pearce, Statistique Canada, 150, promenade Tunney's Pasture, Ottawa (Ontario) K1A 0T6 (andrew.pearce@statcan.gc.ca); Kenza Sallier, Statistique Canada, 150, promenade Tunney's Pasture, Ottawa (Ontario) K1A 0T6 (kenza.sallier@statcan.gc.ca); Christiane Laperrière, Statistique Canada, 150, promenade Tunney's Pasture, Ottawa (Ontario) K1A 0T6 (christiane.laperriere@statcan.gc.ca)

## **1.2 Projet de recherche de 2022-2023 de la Section des opérations du recensement**

Comme mentionné plus haut, on s'affaire à mettre en œuvre le PADD pour de nombreux programmes de Statistique Canada, du plan d'échantillonnage à l'étape de la collecte. Dans le cas des programmes et enquêtes d'ordre social, à l'étape de la conception du projet, il faut déterminer la taille de la population des groupes désagrégés dans le champ d'enquête, ainsi que les lignes directrices connexes pour choisir les méthodes appropriées requises pour atteindre les objectifs au chapitre des données désagrégées. Comme le présent article le souligne, les méthodes d'échantillonnage classiques ne sont pas toujours suffisantes pour permettre une estimation robuste, en raison des défis sur le plan méthodologique qui surgissent compte tenu des caractéristiques particulières des sous-populations visées par le PADD. La Section des opérations du recensement de la Division des méthodes d'intégration statistique (DMIS) a pris l'initiative de mener un projet de recherche, qui décrivait comment exploiter les données du Recensement de la population de 2021, qui, au moment de la rédaction du présent article, est la source des données sociodémographiques les plus détaillées de la société canadienne, afin de soutenir ces procédures. En particulier, les données du recensement de la population permettent de désigner les quatre groupes visés par l'équité en matière d'emploi (les femmes, les Autochtones, les personnes en situation de handicap et les membres des minorités visibles), en plus de répartir ces groupes et d'autres groupes à différents niveaux géographiques.

Finalement, ce projet de recherche avait pour objectif principal d'élaborer un document interne recommandant des principes directeurs qui se fondaient sur la taille et les caractéristiques de la population, afin de déterminer si des estimations directes sont possibles (par exemple, au moyen du suréchantillonnage) ou si d'autres méthodes indirectes et techniques de modélisation sont nécessaires. Plus précisément, on a défini deux produits livrables principaux, présentés dans le présent article, pour atteindre ces objectifs. Le premier produit livrable est un document interne définissant des principes directeurs pour les employés à tous les échelons de l'organisme (comme les gestionnaires d'enquête) qui s'intéressent à la production d'estimations de paramètres pour des sous-groupes d'intérêt particuliers du PADD. Le document sur les principes directeurs sert également d'emplacement central pour les renseignements sur les sources de données, les facteurs éthiques et pratiques, le fardeau de réponse, en plus d'inclure une analyse documentaire des procédures d'échantillonnage/de collecte de données appropriées dans le contexte du PADD. Le deuxième produit livrable prend la forme de tableaux de chiffres se fondant sur les données du Recensement de la population de 2021, qui ont été produits pour soutenir les applications pratiques de ces principes directeurs en permettant d'évaluer la prévalence des sous-populations des groupes d'intérêt du PADD tout en intégrant l'intersectionnalité, qui, selon l'Agence de la santé publique du Canada (2022), « fait référence à la manière dont les sources de discrimination se chevauchent et se renforcent mutuellement ».

## **2. Facteurs relatifs aux données et facteurs d'ordre éthique et pratique**

### **2.1 Éthique des données et fardeau de réponse**

Statistique Canada dispose d'un processus interne d'examen éthique pour l'acquisition de données d'enquête et d'autres données, afin d'évaluer la justification éthique de l'acquisition et de l'utilisation appropriée des données. Les examens éthiques sont une condition nécessaire pour que l'organisme puisse expliquer pourquoi il a besoin d'acquérir les données et comment il garantira leur utilisation correcte, ce qui contribuera à une plus grande acceptation sociale et au maintien de la confiance du public dans l'utilisation des données (Deblois et Côté, 2021). Les examens éthiques sont intégrés à de nombreuses parties du cycle de vie des données, les mêmes cycles de vie des données couverts par le PADD. Ainsi, les initiatives du PADD peuvent soulever des questions d'ordre éthique particulières qui ne correspondent pas à celles habituellement associées aux enquêtes typiques. Si certains défis, comme les héritages ou des sensibilités historiques complexes, les obstacles socioéconomiques et la méfiance générale à l'égard du gouvernement, ne sont pas gérés parallèlement au besoin en matière d'information, des effets négatifs durables peuvent apparaître, nuisant aux initiatives futures axées sur ces sous-populations.

Outre l'éthique des données, le fardeau de réponse est un autre aspect important à prendre en compte lors de la sélection d'une approche en matière de collecte de données dans le contexte du PADD. Le fardeau de réponse correspond à l'effort requis par les répondants pour fournir des renseignements personnels au moyen d'un questionnaire ou de mesures directes (courantes dans le cadre des enquêtes sur la santé) ou lors d'une entrevue.

Lorsqu'il n'est pas géré correctement, le fardeau de réponse plus élevé peut entraîner une augmentation de la non-réponse totale et de la non-réponse partielle et, selon le sujet, peut susciter des inquiétudes de la part du public. En outre, les effets du fardeau de réponse sont aggravés lorsque l'on s'intéresse aux groupes du PADD, car ces groupes sont souvent surreprésentés dans les échantillons pour produire des estimations fiables. Pour les mêmes raisons que celles évoquées ci-dessus (sensibilités, obstacles et méfiance), en plus de la difficulté à recueillir des données, lorsque le fardeau n'est pas pris en compte, il peut compliquer dans l'avenir l'échantillonnage d'une population déjà difficile à échantillonner.

Actuellement, Statistique Canada dispose de processus pour contrôler les différents aspects du fardeau de réponse, allant des consultations avec des experts pour veiller à ce que le questionnaire soit conçu correctement au respect du Cadre de nécessité et de proportionnalité pour s'assurer que seuls les renseignements pertinents sont demandés, proportionnellement au degré de délicatesse et d'acceptabilité sociale du sujet (Statistique Canada, 2024b). Parmi les façons de surveiller le fardeau de réponse, il y a le fait d'utiliser les fichiers historiques de sélection de l'échantillon, qui documentent les sélections antérieures d'un ménage donné lors d'enquêtes précédentes. Les fichiers historiques de sélection assurent une meilleure coordination entre les enquêtes, à l'intérieur et à l'extérieur de la même division, tout en aidant à minimiser la possibilité de choisir à nouveau les mêmes logements au cours d'une brève période, ce qui, en retour, réduit les taux de non-réponse totale et partielle.

### **3. Principaux produits livrables**

#### **3.1 Document sur les principes directeurs**

Le principal produit livrable du projet de recherche prend la forme d'un document interne de Statistique Canada, s'intitulant « Principes directeurs : utiliser les données du Recensement de la population de 2021 pour produire des statistiques sur les groupes d'intérêt du Plan d'action sur les données désagrégées ». Parmi les sous-groupes d'intérêt du PADD, mentionnons les groupes visés par l'équité en matière d'emploi énumérés précédemment (femmes, Autochtones, personnes en situation de handicap et membres des minorités visibles), ainsi que les immigrants au Canada, les personnes en situation de faible revenu, les personnes âgées et les anciens combattants. Les défis méthodologiques observés lors de la collecte de renseignements statistiques sur les sous-groupes d'intérêt pour les initiatives du PADD ont motivé la rédaction de ce document. En raison des caractéristiques couramment observées dans les sous-groupes ciblés par le PADD, les processus d'échantillonnage classiques peuvent ne pas convenir et d'autres plans d'enquête et méthodes d'estimation devraient être envisagés. Ainsi, on a déterminé que Statistique Canada trouverait utile un guide complet servant d'emplacement central pour les approches méthodologiques classiques ou non traditionnelle en matière d'échantillonnage et de collecte de renseignements sur les sous-populations correspondant aux caractéristiques du PADD. Pour chaque méthode d'échantillonnage et procédure de collecte de données décrite dans le guide, les facteurs théoriques correspondants, des facteurs importants et des avantages et inconvénients sont fournis.

##### **3.1.1 Caractéristiques des sous-populations du PADD**

Comme mentionné précédemment, la taille et les caractéristiques d'un sous-groupe d'intérêt du PADD peuvent poser des problèmes méthodologiques aux stades de l'échantillonnage, de la collecte et de l'estimation. Par conséquent, les processus d'échantillonnage ou les procédures de collecte de données classiques peuvent ne pas convenir, et d'autres plans d'enquête et méthodes d'estimation non traditionnels doivent être envisagés.

Un système de classification générale pour les sous-groupes d'une population en fonction de leur taille au sein de la population en question a été lancé par Kish (1987). Le tableau 3.1.1-1 présente cette classification des domaines, et comprend un exemple de chaque classification tiré des tableaux des chiffres de population.

**Tableau 3.1.1-1****Système de classification de domaine proposé par Kish (1987)**

| Classification du domaine | Prévalence au sein de la population | Exemple se fondant sur le Recensement de la population de 2021 |
|---------------------------|-------------------------------------|----------------------------------------------------------------|
| Majeure                   | $\geq 10 \%$                        | Personnes qui ont immigré au Canada                            |
| Mineure                   | De 1 % à 10 %                       | Anciens combattants des Forces canadiennes                     |
| Mini                      | De 0,1 % à 1 %                      | Personnes transgenres au Canada                                |
| Rare                      | $< 0,1 \%$                          | Canadiens membres des Premières Nations ou Métis               |

Même si la taille du domaine fournit des données sur la faisabilité de l'échantillonnage de différentes sous-populations, ce n'est pas le seul facteur pris en considération lorsqu'on choisit une méthode d'échantillonnage appropriée dans le contexte des sous-groupes d'intérêt du PADD. Parmi les autres caractéristiques complexes d'une sous-population, il y a le fait qu'elle peut être difficile à atteindre. Smith (2014) décrit cette situation comme « des sous-groupes qui sont socialement cachés [traduction] » et que « dans de nombreux cas, les groupes se cachent activement en raison des attributs qui font d'eux l'objet de la recherche [traduction] ». Parmi les exemples courants associés aux populations cachées, il y a celui de la consommation de drogues. Le sujet étant considéré comme délicat, il n'est pas réaliste de penser qu'on peut obtenir des renseignements fiables sur les habitudes des toxicomanes au moyen d'un questionnaire habituel, car les répondants peuvent mentir sur leur appartenance au groupe.

### 3.1.2 Méthodes de suréchantillonnage

Parmi les difficultés auxquelles il a fallu faire face lors de l'estimation des paramètres d'intérêt pour une population mineure ou plus petite, il y a le fait que, si la sous-population n'est pas précisément ciblée dans le plan d'échantillonnage, il est probable qu'il y aura trop peu de répondants dans la sous-population d'intérêt pour produire des estimations fiables. Les méthodes de suréchantillonnage peuvent être utilisées pour cibler la sous-population d'intérêt, faisant en sorte que la sous-population est surreprésentée dans l'échantillon par rapport à sa prévalence observée dans la population générale. Les méthodes de suréchantillonnage, comme la stratification disproportionnée dans laquelle la taille de l'échantillon d'au moins une strate n'est pas proportionnelle à la taille des strates dans la population (Daniel, 2011), sont courantes au sein de l'organisme. Elles sont grandement utilisées dans le cadre des enquêtes sociales, ainsi que des initiatives du PADD, afin de veiller à ce qu'on fasse un échantillonnage d'un nombre suffisant de membres de la population cible. Récemment, dans le cadre du test du Recensement de la population de 2024 (Statistique Canada, 2024a), trois groupes d'intérêt du PADD (les personnes ayant une expérience militaire, les Autochtones, les personnes en situation d'itinérance cachée) ont été précisément échantillonnés de manière plus disproportionnée par rapport à leur taille dans la population canadienne. Cela a permis une évaluation plus exacte des répercussions qu'auraient les changements proposés en vue du prochain cycle du recensement sur eux. Même si ce n'est pas toujours une méthode de suréchantillonnage, Statistique Canada utilise souvent les procédures d'échantillonnage à deux phases ou à plusieurs phases. En fait, les enquêtes postcensitaires, comme l'Enquête canadienne sur l'incapacité, utilisent le recensement de la population comme première phase, avant d'utiliser les identificateurs de domaine obtenus pour procéder à l'échantillonnage des sous-populations d'intérêt au cours des phases suivantes. Enfin, il peut être utile d'utiliser de multiples bases de sondage lors de l'échantillonnage de petits groupes d'intérêt du PADD, car une base peut couvrir insuffisamment la sous-population d'intérêt, alors qu'une autre peut la couvrir suffisamment tout en couvrant de manière insuffisante le reste de la population.

Au sein de l'organisme, il existe une idée étroitement liée au suréchantillonnage, un suréchantillon. Dans les deux cas, les répondants de la sous-population ciblée sont surreprésentés. Cependant, contrairement au suréchantillonnage d'une sous-population, qui est prévu dans le plan d'échantillonnage de l'enquête, un suréchantillon est un échantillon supplémentaire et secondaire par rapport à l'échantillon principal de l'enquête. Le suréchantillon est constitué pour atteindre des objectifs de données hors des exigences de base de l'enquête principale. Par exemple, un suréchantillon a été utilisé lors de l'Enquête sur la santé dans les collectivités canadiennes de 2021-2022, afin de mieux comprendre l'état de santé, l'utilisation des soins de santé et les déterminants de la santé parmi les quatre groupes visés par l'équité en matière d'emploi. Dans ce cas-ci, un suréchantillon a été utilisé, car la phase de plan d'enquête était trop avancée pour modifier le plan d'échantillonnage.

### 3.1.3 Méthodes pour les populations cachées

Lorsque l'on cible des populations cachées, la stigmatisation ou la participation à des activités illicites de la population peut avoir des répercussions sur les procédures de collecte (Heckathorn, 1997). Plus précisément, ces populations peuvent délibérément divulguer des renseignements inexacts sur l'appartenance à un domaine aux équipes d'enquête par crainte d'être jugées ou poursuivies. C'est pourquoi, en raison de la nature cachée de certaines populations, l'identification complète ou partielle est rarement possible et, même lorsqu'elle l'est, des inquiétudes peuvent surgir quant à son exactitude. Cependant, les populations cachées présentent souvent d'autres caractéristiques qui peuvent à leur tour être exploitées pour les échantillonner et recueillir des données auprès d'elles. Par exemple, on sait que certaines populations cachées fréquentent divers lieux connus, ce qui permet un échantillonnage spatio-temporel, tandis que d'autres peuvent avoir une structure sociale sous-jacente, ce qui permet des méthodes d'échantillonnage par références enchaînées.

L'échantillonnage spatio-temporel (EST) peut servir à procéder à l'échantillonnage des membres de sous-populations qui, selon ce que l'on sait, fréquentent différents lieux, comme des sous-populations sans adresse fixe et les sous-populations mobiles (qui se déplacent fréquemment d'un lieu à un autre). En bref, l'EST permet de sélectionner des personnes en effectuant l'échantillonnage dans des lieux qu'on sait qu'elles fréquentent (comme des banques alimentaires et des refuges pour itinérants). Il est possible d'établir une base de sondage des lieux connus et des moments en vue de l'échantillonnage, procurant des résultats probabilistes. Toutefois, il peut y avoir sous-couverture si des membres de la population se rendent dans d'autres lieux à d'autres moments, inconnus de l'équipe d'enquête, ou si des lieux sont exclus de la base de sondage (par exemple, si l'on juge ces endroits dangereux). En revanche, en faisant un échantillonnage à différents endroits et à différents moments, il est possible que le même membre de la population soit sélectionné plusieurs fois dans l'échantillon, ce qui pose le problème de la multiplicité (Kalton, 2009).

Lorsqu'on procède à l'échantillonnage de populations ayant des liens sociaux, les méthodes d'échantillonnage par références enchaînées peuvent être appropriées. Il s'agit d'un ensemble de méthodes d'échantillonnage qui dépend d'un groupe initial de répondants (recruteurs). Lorsqu'ils participent à l'enquête, on leur demande, moyennant un incitatif, de fournir le nom d'autres personnes qu'ils connaissent au sein de la population pour qu'elles participent à l'enquête. La référence séquentielle d'autres personnes à l'enquête crée des chaînes de référence, qui progressent et prennent de l'expansion jusqu'à ce qu'on obtienne une taille d'échantillon souhaitée ou le degré de précision d'une estimation désiré. Dans le guide, on discute de deux de ces méthodes, l'échantillonnage en boule de neige et l'échantillonnage en fonction des répondants (EFR). L'échantillonnage en boule de neige peut fournir des résultats probabilistes si les recruteurs initiaux sont choisis aléatoirement depuis une base de sondage. Cependant, cela n'est pas possible dans le cas des populations cachées dans le contexte du PADD. L'EFR, qu'on peut considérer comme un prolongement de l'échantillonnage en boule de neige, peut fournir des résultats probabilistes indépendamment des recruteurs sélectionnés initialement (Heckathorn, 1997) au moyen d'un ensemble d'hypothèses rigoureuses qui sont difficiles à défendre en pratique.

### 3.1.4 Facteurs pratiques

Parmi les autres objectifs principaux du document sur les principes directeurs, il y avait le fait d'orienter, au moyen d'une série de facteurs et de questions pratiques, le lecteur vers une méthode d'enquête ou une stratégie de collecte de données adaptée à ses besoins particuliers. L'ensemble de facteurs et de questions pratiques commence par demander au lecteur d'indiquer s'il peut justifier de manière rigoureuse la collecte de données sur un sous-groupe particulier du PADD. Cette question souligne le fait que, si le PADD cherche à faire de la désagrégation des données une pratique courante au sein de l'organisme, la justification de la collecte des données doit aller au-delà du simple fait de combler une lacune actuelle dans les données, en particulier si l'on pose des questions délicates. On recommande aux lecteurs du guide de suivre le Cadre de nécessité et de proportionnalité pour mieux évaluer leur justification quand vient le temps de recueillir ces données. Il faudrait également tenir compte du fardeau de réponse avant de trouver des méthodes d'échantillonnage ou de collecte des données appropriées. Les lecteurs peuvent consulter la section sur le fardeau de réponse du document sur les principes directeurs. Elle décrit les outils accessibles mentionnés dans le document pour évaluer le fardeau qu'imposerait leur initiative au sous-groupe d'intérêt du PADD possiblement de petite taille.

Lorsque le lecteur a évalué le fardeau et les répercussions éthiques de son initiative, on lui demande ensuite d'indiquer les caractéristiques du sous-groupe d'intérêt du PADD, afin de déterminer la taille de la sous-population, si elle est difficile à atteindre et le degré d'identification. Le niveau d'identification, c'est-à-dire la proportion de la sous-population pour laquelle des identificateurs de domaine sont accessibles, peut varier pour de nombreuses raisons. Évidemment, la sous-population peut être cachée, et des identificateurs de domaine peuvent être manquants pour cette raison. Certaines questions d'identification ne peuvent être posées qu'au moyen du questionnaire détaillé du recensement. C'est pourquoi seulement environ 25 % de la population est accessible. En outre, il se peut que les questions d'identification n'aient pas été posées dans le cadre du recensement, mais plutôt dans le cadre d'autres enquêtes sociales. Le cas échéant, on conseille aux lecteurs d'utiliser les outils de recherche des bases de données internes pour évaluer le niveau d'identification actuel, en déterminant les sources de données qui existent pour la sous-population en question. Enfin, lorsque des questions d'identification ne sont pas posées dans le cadre du recensement ou d'enquêtes sociales, il est possible d'utiliser plutôt des identificateurs de domaine de substitution, en ayant recours à des variables accessibles qui sont corrélées avec la population d'intérêt, afin d'identifier les membres probables de la sous-population d'intérêt.

Une fois que ces trois caractéristiques sont évaluées, on dirige le lecteur vers une section appropriée du document qui décrit les facteurs importants applicables dans sa situation, les méthodes d'échantillonnage correspondant à ses besoins, ainsi que les avantages et inconvénients de ces méthodes. On oriente le lecteur au moyen d'un ensemble de questions, comme « votre population d'intérêt est-elle cachée? » ou « puisque votre population d'intérêt n'est pas cachée, de quel niveau d'identification disposez-vous pour celle-ci? ». Par exemple, si la sous-population d'intérêt du lecteur est très petite (p. ex. mini ou rare, comme le montre le tableau 3.1.1-1), on le dirigera vers des renseignements portant sur l'estimation sur petits domaines ou la combinaison d'échantillons transversaux. Si la sous-population n'a aucun renseignement identificatoire, ou en a très peu avant la collecte, on présenterait au lecteur des renseignements sur les approches de sélection.

### **3.2 Tableaux des chiffres de population**

Afin de relever les défis méthodologiques liés à la taille de l'échantillon des sous-populations d'intérêt et de formuler des recommandations plus concrètes, le projet de recherche a assuré la production interne de tableaux de chiffres pour de nombreux sous-groupes d'intérêt du PADD en fonction du Recensement de 2021. Cependant, avant de produire ces tableaux, il fallait choisir la source des données parmi trois possibilités accessibles au sein de Statistique Canada : la base de données des réponses (BDR) du Recensement de 2021, qui comprend les réponses brutes fournies par les répondants, la base de données du contrôle et de l'imputation (C et I) du Recensement de 2021, qui comprend des variables nettoyées, agrégées, imputées et extraites en fonction des réponses des répondants, et la base de données de la diffusion, qui comprend les variables dont la diffusion publique est prévue et qui respectent les règles de confidentialité. On a décidé d'utiliser la base de données de C et I de 2021 pour quelques raisons, notamment en raison de la propreté des données, ce qui est particulièrement utile lorsque des variables ont déjà été extraites par d'autres équipes disposant d'un degré supérieur d'expertise (p. ex. la production de chiffres pour la population non binaire). Ensuite, la base de données de la diffusion ne dispose peut-être pas de tous les renseignements nécessaires en raison des règles de confidentialité et de suppression mises en application, faisant en sorte qu'on ne peut pas accepter que ces chiffres soient utilisés par le public (p. ex. les chiffres de la population non binaire sont restreints).

À l'aide de la base de données de C et I, on a produit des chiffres de population au niveau de la personne pour les sous-groupes du PADD, répartis en fonction de facteurs d'intersectionnalité. Cela a permis de souligner la faisabilité de l'échantillonnage, et de déterminer les stratégies d'échantillonnage à envisager en fonction de la taille estimative de la population. Au départ, dans le cadre des discussions avec un groupe de travail du PADD au sein de l'organisme et au moyen de définitions normalisées allant de pair avec les données du Recensement de 2021, on a déterminé que les sous-populations suivantes représentaient les groupes d'intérêt : minorités de genre (transgenre, non binaire), Autochtones (Premières Nations, Métis, Inuit) et identité autochtone, groupes des minorités visibles et membres actifs ou anciens combattants des Forces canadiennes. Cependant, après avoir consulté les gestionnaires d'enquête de l'organisme, afin de mieux comprendre leurs besoins en matière d'enquête lors de la mise en œuvre d'initiatives du PADD, les groupes pris en compte ont été élargis, afin d'y ajouter les personnes ayant une limitation des activités, les immigrants au Canada et différents groupes religieux. Dans le cas de toutes les sous-populations d'intérêt, on a assuré une répartition selon le genre (homme cisgenre, femme cisgenre, homme transgenre, femme transgenre et diverses identités de genre), la province ou le territoire, ainsi que le groupe d'âge. Dans le cas de certains sous-groupes d'intérêt,

il y a eu répartition selon la catégorie de revenu et la taille de l'agglomération. Dans le cas des Autochtones, les chiffres ont été désagrégés de manière plus approfondie selon le fait que la personne vivait dans une réserve ou hors réserve.

En ce qui concerne la production des tableaux de chiffres, il fallait prendre en compte d'autres facteurs en plus de la source de données, des sous-groupes d'intérêt et des variables d'intersectionnalité. Dans le cas de certains groupes de population, les renseignements utilisés pour les identifier proviennent exclusivement du questionnaire détaillé du recensement (par exemple, la situation militaire) qui n'est envoyé qu'à environ 25 % des logements. Ainsi, il a fallu utiliser les poids du questionnaire détaillé pour produire des tableaux de chiffres pondérés pour ces populations.

Dans le même ordre d'idées, les valeurs imputées ont également été prises en compte. Puisque l'un des objectifs du PADD est d'échantillonner des sous-groupes d'intérêt particuliers, il faut avoir des renseignements précis pour garantir la validité de l'appartenance d'une personne à un domaine. Étant donné que la valeur imputée indiquant leur appartenance à la sous-population peut être différente de la valeur réelle, tous les enregistrements dans lesquels la variable indiquant l'appartenance a été imputée ont été exclus des chiffres de la sous-population en question. Ces tableaux ne sont accessibles qu'au sein de l'organisme. L'accès est donc réservé exclusivement aux méthodologistes d'enquêtes sociales qui demandent un accès et qui justifient adéquatement leur utilisation.

## 4. Conclusion

Motivée par les 17 objectifs de développement durable de l'ONU et les enjeux sociodémographiques au Canada, la Section des opérations du recensement a réalisé avec succès un projet de recherche novateur axé sur l'utilisation des données du Recensement de la population de 2021, afin de compiler et de documenter des facteurs importants lorsque des initiatives du PADD sont menées. Dans le document sur les principes directeurs, on trouve des discussions et des ressources résumées et centralisées pour obtenir des renseignements sur le contexte organisationnel du PADD, l'éthique des données, le fardeau de réponse, les tableaux de chiffres des sous-populations répartis en fonction de facteurs d'intersectionnalité, des défis méthodologiques (taille de la population, population cachée, niveau d'identification), ainsi que les méthodes d'échantillonnage et les procédures de collecte de données appropriées qui peuvent servir à relever ces défis.

## Bibliographie

- Daniel, J. (2012), "Choosing the type of probability sampling", *Sampling essentials: Practical guidelines for making sampling choices*, p. 125-174.
- DeBlois, S., et D. Côté (2021), "Perceptions of data sensitivity: Study with Questionnaire Testing Participants", rapport non publié, Ottawa, Canada, Statistique Canada.
- Heckathorn, D. D. (1997), "Respondent-driven sampling: a new approach to the study of hidden populations", *Social problems*, Vol. 44, No. 2, p. 174-199.
- Kish, L. (1987), *Statistical Design for Research*. New York: Wiley & Sons, Inc.
- Kalton, G. (2009), "Methods for oversampling rare subpopulations in social surveys", *Survey Methodology*, p. 125-141.
- Statistique Canada (2024a), "Test du Recensement 2024" [En ligne] Article accessible à l'adresse : [Test du recensement de 2024](#) . Consulté : 2025-01-09.
- Statistique Canada (2024b), "Principes de nécessité et de proportionnalité", un article publié sur le site web de Statistique Canada : [Principes de nécessité et de proportionnalité](#) .

Agence de la santé publique du Canada (2022), “Comment intégrer la théorie de l’intersectionnalité dans les analyses quantitatives d’équité en santé ? Une revue rapide et liste de vérification de pratiques prometteuses”, un article publié sur le site web du gouvernement du Canada. [Comment intégrer la théorie de l’intersectionnalité dans les analyses quantitatives d’équité en santé - Canada.ca](#).

Smith, T. W. (2014), “Choosing the type of probability sampling”, *Sampling essentials: Practical guidelines for making sampling choices*, p. 125-174.

United Nations Economic Commission for Europe Statistics Division (2019), “Generic Statistical Business Process Model”, un article publié sur le site web de l’UNECE, [GSBPM v5.1 | UNECE](#).

Département des Affaires Économiques et Sociales des Nations Unies, (2024), “Développement durable” [En ligne] Accessible à l’adresse : [THE 17 GOALS | Sustainable Development](#), Consulté : 2024-12-13.