Guide analytique – Série d'enquêtes sur les perspectives canadiennes 4 : Les sources d'information consultées pendant la pandémie

1.0 Description

La Série d'enquêtes sur les perspectives canadiennes (SEPC) est un ensemble de courtes enquêtes en ligne menées à compter de mars 2020 qui serviront à recueillir, au cours de l'année, des renseignements sur les connaissances et les comportements des résidents des 10 provinces canadiennes. Les questions de toutes les enquêtes de la série seront posées à un panel probabiliste de Statistique Canada. L'échantillon probabiliste comprend un sous-ensemble de répondants à l'Enquête sur la population active (EPA) qui ont accepté de participer à ce nouveau projet pilote entrepris en 2019. Un but important de la SEPC est de recueillir, en temps utile, des données directement auprès des Canadiens afin d'éclairer les décideurs politiques et de répondre rapidement aux nouveaux besoins en données. La SEPC vise à produire des données au niveau national (excluant les territoires).

Le programme de l'enquête est financé par Statistique Canada. Chacune des enquêtes de la SEPC est transversale. La participation au panel probabiliste et aux enquêtes subséquentes de la SEPC est volontaire.

La quatrième enquête de la SEPC est SEPC4 – Les sources d'information consultées pendant la pandémie. Elle a été menée du 20 au 26 juillet 2020.

Toutes les questions concernant l'enquête, la série d'enquêtes, les données ou leur utilisation devraient être adressées à :

Statistique Canada

Services à la clientèle
Centre de l'intégration et du développement des données sociales
Téléphone : 613-951-3321, ou numéro sans frais : 1-800-461-9050
Télécopieur : 613-951-4527
Courriel : statcan.csdidclientservice-ciddsservicealaclientele.statcan@canada.ca

2.0 Méthodologie d'enquête

Population cible et population observée

La population cible de la Série d'enquêtes sur les perspectives canadiennes (SEPC) est constituée des résidents des 10 provinces canadiennes âgés de 15 ans ou plus.

La base de sondage des enquêtes de la SEPC est le panel probabiliste du projet pilote de Statistique Canada. Le panel probabiliste a été créé par sélection aléatoire d'un sous-ensemble de répondants à l'Enquête sur la population active (EPA). Par conséquent, la population observée est celle de l'EPA, à l'exception près que les membres à temps plein des Forces armées canadiennes sont inclus. Sont exclus du champ de l'enquête les personnes vivant dans les réserves et d'autres peuplements autochtones des provinces, les pensionnaires d'établissements institutionnels et les ménages situés dans des régions extrêmement éloignées où la densité de population est très faible. Réunis, ces groupes exclus de l'enquête représentent moins de 2 % de la population canadienne âgée de 15 ans et plus.

L'échantillon de l'EPA est sélectionné à partir d'une base aréolaire et se fonde sur un plan stratifié à plusieurs degrés s'appuyant sur un échantillonnage probabiliste. Le plan de sondage de l'EPA en est un par renouvellement de panel. Dans les provinces, les logements sélectionnés demeurent dans l'échantillon de l'EPA pendant six mois consécutifs. Chaque mois, environ un sixième des logements sélectionnés dans le cadre de l'EPA en sont à leur premier mois d'enquête, un autre sixième des logements en sont à leur deuxième mois d'enquête, et ainsi de suite. Ces six échantillons indépendants sont appelés des groupes de renouvellement.

Pour le panel probabiliste utilisé dans le cadre de la SEPC, quatre groupes de renouvellement de l'EPA ont été utilisés à partir des provinces : les groupes de renouvellement qui ont répondu à l'EPA pour la dernière fois en avril, mai, juin et juillet 2019. Auprès de ces ménages, une personne âgée de 15 ans ou plus a été sélectionnée au hasard pour participer à la SEPC. Ces personnes ont été invitées à s'inscrire à la SEPC. On a demandé à celles qui ont accepté de participer à la SEPC de fournir une adresse courriel. Les participants issus de l'inscription à la SEPC ayant fourni des adresses courriel valides ont constitué le panel probabiliste. Le taux de participation au panel était d'environ 23 %. La population observée pour toutes les enquêtes de la SEPC est celle des participants au panel probabiliste. Les participants du panel avaient 15 ans ou plus au 31 juillet 2019.

Plan d'échantillonnage et taille de l'échantillon

Le plan d'échantillonnage pour les enquêtes de la SEPC est fondé sur celui de la SEPC-Inscription, la méthode utilisée pour créer le panel probabiliste pilote. L'échantillon brut de la SEPC-Inscription comprenait 31 896 personnes âgées de 15 ans et plus sélectionnées de façon aléatoire dans les ménages ayant répondu à l'EPA qui terminaient leur dernière interview de l'EPA entre avril et juillet 2019. Parmi ces personnes, 31 626 étaient dans le champ de l'enquête au moment de la collecte pour la SEPC-Inscription réalisée de janvier à mars 2020. Parmi les personnes ayant accepté de participer à la SEPC, c'est-à-dire celles qui se sont jointes au panel, 7 242 possédaient une adresse courriel valide. Tous les participants du panel sont invités à répondre aux enquêtes de la SEPC.

Degrés d'échantillonnage	n
Échantillon brut pour la SEPC-Inscription	31 896
Unités dans le champ de l'enquête de la SEPC-Inscription	31 628
Panélistes pour la SEPC (ayant des adresses courriel valides)	7 242
Échantillon brut pour les enquêtes de la SEPC	7 242

3.0 Collecte des données

SEPC-Inscription

L'enquête SEPC-Inscription utilisée pour créer le panel probabiliste de Statistique Canada a été menée du 15 janvier 2020 au 15 mars 2020. Un contact initial a été fait par l'envoi postal d'une lettre à l'échantillon sélectionné. La lettre expliquait l'objectif de la SEPC et invitait les répondants à se rendre en ligne, à utiliser leur code d'accès sécurisé et à remplir le formulaire d'inscription. On demandait aux répondants ayant décidé de ne pas participer la raison principale de leur refus. On a demandé à ceux qui ont accepté de se joindre au panel de vérifier les renseignements démographiques de base et de fournir une adresse courriel valide. Le suivi des non-réponses pour la SEPC-Inscription comportait une approche multimodale. Des lettres de rappel et des messages textes supplémentaires (lorsqu'un numéro de téléphone cellulaire était fourni) ont été envoyés pour encourager les personnes faisant partie de l'échantillon à répondre. Un suivi des non-réponses a aussi été mené par interview téléphonique assistée par ordinateur (ITAO).

L'application comprenait un ensemble de codes de réponse normalisé pour indiquer tous les résultats possibles. L'application a été mise à l'essai avant son utilisation pour s'assurer que seules les réponses valides aux questions pourraient être saisies et que tous les enchaînements de questions seraient suivis correctement. Ces mesures ont permis de veiller à ce que les données collectées soient déjà « épurées » à la fin du processus de collecte.

Les intervieweurs ont suivi une approche normalisée utilisée dans le cadre de nombreuses enquêtes de Statistique Canada pour présenter l'agence. Les personnes sélectionnées ont été informées que leur participation à l'enquête était volontaire et que les renseignements fournis resteraient strictement confidentiels.

SEPC4 - Les sources d'information consultées pendant la pandémie

Tous les participants au panel pilote de la SEPC, sauf ceux qui ont demandé de ne plus participer au panel après la SEPC1, ont reçu une invitation par courriel contenant un lien vers l'enquête SEPC4 et un code d'accès sécurisé en vue de répondre à l'enquête en ligne. La collecte des données de l'enquête a commencé le 20 juillet 2020. Des courriels de rappel ont été envoyés le 21 juillet, le 23 juillet et le 25 juillet. L'application est restée ouverte jusqu'au 26 juillet 2020.

3.1 Contrôle de la divulgation

La loi interdit à Statistique Canada de rendre publique toute donnée susceptible de révéler de l'information obtenue en vertu de la Loi sur la statistique et se rapportant à toute personne, entreprise ou organisation reconnaissable sans que cette personne, entreprise ou organisation le sache ou y consente par écrit. Diverses règles de confidentialité s'appliquent à toutes les données diffusées ou publiées afin d'empêcher la publication ou la divulgation de toute information jugée confidentielle. Au besoin, des données sont supprimées pour empêcher la divulgation directe ou par recoupement de données reconnaissables.

4.0 Qualité des données

Diverses sources expliquent les erreurs d'enquête. Ces sources peuvent être classées en deux principales catégories, soit les erreurs non dues à l'échantillonnage et les erreurs d'échantillonnage.

4.1 Erreurs non dues à l'échantillonnage

Les erreurs non dues à l'échantillonnage peuvent être définies comme des erreurs possibles pendant à peu près toutes les activités d'enquête, mis à part l'échantillonnage. Ces erreurs se retrouvent dans les enquêtes-échantillons et les recensements (contrairement à l'erreur d'échantillonnage qui est présente seulement dans les enquêtes-échantillons). Les principales sources d'erreurs non dues à l'échantillonnage sont les suivantes : la non-réponse, la couverture, la mesure et le traitement.

4.1.1 Non-réponse

La non-réponse découle de l'incapacité à recueillir des renseignements complets pour toutes les unités sélectionnées dans l'échantillon.

La non-réponse peut occasionner deux types d'erreurs dans les estimations de l'enquête. En premier lieu, les non-répondants ont souvent des caractéristiques différentes de celles des répondants, et les estimations de l'enquête seront biaisées si la non-réponse n'est pas entièrement corrigée par la pondération. En deuxième lieu, la non-réponse diminue la taille réelle de l'échantillon parce qu'il était prévu que davantage d'unités répondraient à l'enquête. La variance d'échantillonnage augmente donc au détriment de la précision des estimations. Le taux de réponse est calculé comme suit :

[Unités répondantes / (Unités sélectionnées – unités hors champ)] x 100 %

Le tableau suivant résume les taux de réponse obtenus pour la SEPC4 – Les sources d'information consultées pendant la pandémie. Les taux de réponse sont répartis en deux étapes. Le tableau 4.1.1a montre les taux de participation au panel de la SEPC-Inscription et le tableau 4.1.1b montre les taux de réponse à la SEPC4 – Les sources d'information consultées pendant la pandémie.

Tableau 4.1.1a Participation au panel probabiliste pilote pour la SEPC-Inscription

	Degrés d'échantillonnage pour la SEPC-Inscription			
	Échantillon brut pour la SEPC-Inscription	Unités dans le champ de l'enquête de la SEPC-Inscription	Panélistes pour la SEPC (ayant des adresses courriel valides)	Taux de participation au panel pour la SEPC
n	31 896	31 628	7 242	22,9 %

Tableau 4.1.1b Taux de réponse à la SEPC4 – Les sources d'information consultées pendant la pandémie.

	Degrés d'échantillonnage pour la SEPC4 – Les sources d'information consultées pendant la pandémie			
	Panélistes pour la SEPC (ayant des adresses courriel valides)	Répondants à la SEPC4 – Les sources d'information consultées pendant la pandémie	Taux de réponse à la SEPC4 – Les sources d'information consultées pendant la pandémie	Taux de réponse cumulatif
n	7 242	4 218	58,2 %	13,3 %

Comme le montre le tableau 4.1.1b, le taux de réponse à la SEPC4 – Les sources d'information consultées pendant la pandémie est de 58,2 %. Cependant, lorsque l'on tient compte de la non-participation au panel, le taux de réponse cumulatif à l'enquête est de 13,3 %. Ce taux de réponse cumulatif est inférieur aux taux de réponse habituellement observés dans le cadre des enquêtes sociales menées à Statistique Canada. Cela est attribuable aux deux étapes de non-réponse (ou participation) et à d'autres facteurs tels que le mode unique utilisé pour les enquêtes de la SEPC (courriels d'invitation à participer aux enquêtes contenant un lien vers le questionnaire d'enquête à remplir en ligne), la fatigue des répondants en raison de leur participation antérieure à l'EPA et l'incapacité de la population hors ligne à participer, entre autres.

Compte tenu de la non-réponse supplémentaire observée dans le cadre de la SEPC4 – Les sources d'information consultées pendant la pandémie, on note un risque accru de biais dû au fait que les répondants sont différents des non-répondants. C'est pour cette raison qu'une courte étude sur le biais a été menée. Se référer à la section 6.0 pour connaître les résultats de cette validation.

4.1.2 Erreurs de couverture

Les erreurs de couverture comprennent les omissions, les inclusions erronées, les doubles comptes et les erreurs de classification des unités dans la base de sondage. Elles ont des répercussions sur chaque estimation de l'enquête et sont donc l'un des plus importants types d'erreurs. Elles peuvent même être la principale source d'erreurs d'un recensement. Les erreurs de couverture peuvent causer des estimations biaisées et les répercussions peuvent varier pour différents sous-groupes de la population. Il s'agit d'une erreur très difficile à mesurer ou à quantifier de façon précise.

Dans le cadre de la SEPC, la population couverte est constituée des personnes âgées de 15 ans et plus au 31 juillet 2019. Puisque la collecte des données de la SEPC4 – Les sources d'information consultées pendant la pandémie a été réalisée du 20 juillet au 26 juillet 2020, on remarque un sous-dénombrement des résidents des 10 provinces qui ont eu 15 ans après le 31 juillet 2019. On note aussi un sous-dénombrement des personnes qui n'ont pas d'accès à Internet. Ce sous-dénombrement est plus marqué chez les personnes de 65 ans et plus.

4.1.3 Erreurs de mesure

Les erreurs de mesure (erreurs de réponse) désignent la différence entre la réponse inscrite à une question et la « vraie » valeur. Le répondant, le questionnaire, la méthode de collecte des données et l'outil de mesure peuvent susciter ce genre d'erreur. Ces erreurs peuvent entraîner un biais systématique si elles ne sont pas aléatoires. Il est très onéreux de mesurer avec précision les erreurs de réponse et très peu d'enquêtes mènent ce genre d'étude.

4.1.4 Erreurs de traitement

Les erreurs de traitements sont associées aux activités menées une fois que les réponses ont été reçues. Elles comprennent toutes les activités de traitement suivant la collecte et précédant l'estimation. Elles peuvent être aléatoires comme les autres erreurs et accroître ainsi la variance des estimations de l'enquête, ou elles peuvent être systématiques et introduire un biais. Il est difficile d'obtenir des mesures directes des erreurs de traitement, ainsi que de leur incidence sur la qualité des données, puisqu'elles sont souvent confondues avec d'autres types d'erreurs (non-réponse, mesure et couverture).

4.2 Erreurs d'échantillonnage

L'erreur d'échantillonnage est définie comme étant l'erreur découlant de l'estimation d'une caractéristique de la population fondée sur la mesure d'une partie de la population plutôt que de l'ensemble de la population. Des méthodes de calcul de l'erreur d'échantillonnage s'appliquent aux enquêtes-échantillon probabilistes. Ces méthodes découlent directement du plan d'échantillonnage et de la méthode d'estimation de l'enquête.

La mesure appliquée le plus souvent pour quantifier l'erreur d'échantillonnage est la variance d'échantillonnage. La variance d'échantillonnage détermine la portée avec laquelle l'estimation d'une caractéristique de divers échantillons possibles de même taille et de même conception diffère d'un échantillon à l'autre. Dans le cas des plans d'échantillonnage qui utilisent l'échantillonnage probabiliste, l'ampleur de la variance d'échantillonnage d'une estimation peut être déterminée.

Les facteurs qui ont des répercussions sur l'ampleur de la variance d'échantillonnage pour une taille d'échantillon donnée comprennent :

1. La variabilité de la caractéristique d'intérêt dans la population : plus la caractéristique dans la population est variable, plus la variance d'échantillonnage est grande.
2. La taille de la population : en général, la taille de la population a des répercussions sur la variance d'échantillonnage seulement pour les populations de petite taille ou de taille moyenne.
3. Le taux de réponse : la variance d'échantillonnage augmente avec la réduction de la taille de l'échantillon. Puisque les non-répondants diminuent en fait la taille de l'échantillon, la non-réponse accroît la variance d'échantillonnage.
4. Le plan d'échantillonnage et la méthode d'estimation : certains plans d'échantillonnage sont plus efficaces que d'autres parce que, pour la même taille d'échantillon et la même méthode d'estimation, un plan peut donner une variance d'échantillonnage moindre qu'un autre.

L'erreur type d'un estimateur est la racine carrée de sa variance d'échantillonnage. Cette mesure est plus facile à interpréter parce qu'elle donne une indication de l'erreur d'échantillonnage à

l'aide de la même échelle que l'estimation, tandis que la variance est fondée sur les différences quadratiques.

Le coefficient de variation (CV) est une mesure relative de l'erreur d'échantillonnage. Il est défini comme l'estimation de l'erreur type divisée par l'estimation elle-même, habituellement exprimée en pourcentage (10 % au lieu de 0,1). Il est très utile pour mesurer et comparer l'erreur d'échantillonnage de variables quantitatives avec de grandes valeurs positives. Cependant, il n'est pas recommandé pour des estimations telles que les proportions, les estimations des changements ou des différences, et les variables qui peuvent avoir des valeurs négatives.

Il est considéré comme une pratique exemplaire à Statistique Canada de faire état de l'erreur d'échantillonnage d'une estimation par l'entremise de son intervalle de confiance à 95 %. L'intervalle de confiance à 95 % d'une estimation signifie que si l'enquête était répétée à maintes reprises, 95 % du temps (ou 19 fois sur 20), l'intervalle de confiance couvrirait la véritable valeur de la population.

5.0 Pondération

Le principe de base qui sous-tend l'estimation dans un échantillon probabiliste comme ceux de la SEPC repose sur le fait que chaque personne de l'échantillon représente, en plus d'elle-même, plusieurs autres personnes qui ne font pas partie de l'échantillon. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne de l'échantillon en représente 50. Dans la terminologie en usage ici, nous dirons que nous avons attribué à chaque personne un poids de 50.

La phase de la pondération donne lieu au calcul du poids d'échantillonnage de chaque personne échantillonnée. Ce poids, qui figure dans le fichier de microdonnées, doit servir à obtenir des estimations représentatives de la population cible à partir de l'enquête. Par exemple, si l'on doit évaluer le nombre de personnes qui fument quotidiennement, on le fait en choisissant dans l'échantillon les enregistrements des personnes qui présentent cette caractéristique et en faisant la somme des poids attribués à ces enregistrements. La phase de la pondération est une étape où l'on calcule ce nombre (ou poids) pour chaque enregistrement. La présente section contient les détails sur la méthode utilisée pour calculer les poids d'échantillonnage de la SEPC4 – Les sources d'information consultées pendant la pandémie.

La pondération de l'échantillon pour la SEPC4 – Les sources d'information consultées pendant la pandémie comporte de multiples étapes pour refléter les étapes de l'échantillonnage, de la participation et de la réponse pour obtenir l'ensemble final de répondants. Les sections suivantes couvrent les étapes de pondération pour d'abord créer les poids du panel, puis les étapes de pondération pour créer les poids de l'enquête pour la SEPC4 – Les sources d'information consultées pendant la pandémie.

5.1 Création des poids du panel

Quatre échantillons consécutifs de ménages sortants de l'EPA étaient le point de départ pour former l'échantillon du panel de la SEPC. Comme les ménages sélectionnés à partir des échantillons de l'EPA constituent le point de départ, les poids des ménages de l'EPA sont la première étape pour calculer les poids du panel.

5.1.1 Poids des ménages

Calcul des poids de sondage initiaux des ménages – HHLD_W0, HHLD_W1

Les poids initiaux du panel sont les sous-poids de l'EPA (SUBWT). Ce sont les poids de sondage de l'EPA après correction pour tenir compte de la non-réponse, mais pas encore étalonnés selon les totaux de contrôle de la population. Ces poids constituent le poids de sondage initial du ménage pour l'enquête par panel (HHLD_W0).

Puisque seulement quatre échantillons sortants ont été utilisés au lieu des six utilisés dans un échantillon complet de l'EPA, ces poids ont été corrigés par un facteur de 6/4 pour être représentatifs. Après cette correction, les poids ont été appelés HHLD_W1.

Étalonnage des poids des ménages – HHLD_W2

L'étalonnage est une étape qui permet de garantir que la somme des poids à l'intérieur d'un domaine donné correspond aux totaux démographiques projetés. Les SUBWT de l'EPA ne sont pas étalonnés; les HHLD_W1 ne sont donc pas étalonnés non plus. L'étape suivante consiste à s'assurer que la somme des poids des ménages correspond aux totaux de contrôle par taille du ménage. Les HHLD_W1 ont été étalonnés pour correspondre aux totaux de contrôle par province et taille du ménage en utilisant les groupes de taille 1, 2 ou 3+.

5.1.2 Poids du panel au niveau de la personne

Calcul des poids de sondage initiaux au niveau de la personne – PERS_W0

Une personne âgée de 15 ans ou plus par ménage a été sélectionnée pour la SEPC-Inscription, l'enquête qui a servi à créer le panel probabiliste. Le poids de sondage initial affecté à la personne est obtenu en multipliant HHLD_W2 par le nombre de personnes admissibles dans le logement (c.-à-d. nombre de personnes âgées de 15 ans et plus).

Retrait des unités hors champ – PERS_W1

Certaines unités ont été identifiées comme étant hors du champ de l'enquête lors de la SEPC-Inscription. Ces unités ont obtenu un poids de PERS_W1 = 0. Pour toutes les autres unités, PERS_W1 = PERS_W0. Les personnes ayant un poids de 0 sont ensuite retirées des futurs ajustements de poids.

Ajustement pour la non-réponse/non-participation – PERS_W2

Lors de la collecte des données de la SEPC-Inscription, une certaine proportion des unités échantillonnées a inévitablement terminé comme une non-réponse ou une non-participation au panel. Les poids des unités de non-réponse/non-participation ont été redistribués aux unités participantes. Les poids des unités n'ayant pas participé au panel ont été redistribués aux unités participantes ayant des caractéristiques semblables dans les groupes de réponses homogènes (GRH).

De nombreuses variables de l'EPA pouvaient permettre de bâtir les GRH (comme la situation d'activité, le niveau de scolarité, la composition du ménage), en plus des renseignements du processus de collecte des données de l'EPA. Le modèle était précisé par province, car les variables choisies dans le modèle pouvaient être différentes d'une province à l'autre.

Les variables suivantes ont été conservées dans le modèle de régression logistique final : education_lvl (variable de niveau de scolarité comprenant 10 catégories), nameissueflag (indicateur créé pour identifier les répondants n'ayant pas fourni un nom valide), elg_hhldsize (nombre de personnes pouvant être sélectionnées dans le ménage), age_grp (groupe d'âge de la personne choisie), sexe, kidsinhhld (indicateur de la présence d'enfants dans le ménage), marstat (état matrimonial – 6 catégories), cntrybth (indique si le répondant est né au Canada ou non), lfsstat (situation d'activité du répondant – 3 catégories), nocs1 (premier chiffre du code de la Classification nationale des professions si le répondant est employé – 10 catégories) et dwelrent (indique si le répondant est propriétaire ou locataire de son logement). Des GRH ont été formés à l'intérieur des provinces. Un facteur d'ajustement a été calculé à l'intérieur de chaque groupe de réponse de la façon suivante :

$$\frac{\text{Somme des poids des répondants et des non} - \text{répondants}}{\text{Somme des poids des répondants}}$$

Les poids des répondants ont été multipliés par ce facteur pour produire les poids PERS_W2, ajustés pour tenir compte de la non-participation au panel. Les unités non participantes ont été retirées du panel.

5.2 Création des poids de la SEPC4

Les enquêtes de la SEPC commencent par la création de l'échantillon à partir des participants du panel. Au total, 7 242 personnes forment le panel, chacune ayant le poids corrigé pour tenir compte de la non-réponse, PERS_W2.

Calcul des poids de sondage initiaux – WT_DSGN

Le poids de sondage initial est le poids de la personne corrigé pour tenir compte de la non-réponse calculée pour les participants au panel (PERS_W2). Aucune unité hors du champ de l'enquête n'a été observée lors de la collecte des données d'enquête de la SEPC4 **–** Les sources d'information consultées pendant la pandémie. Puisque toutes les unités faisaient partie du champ de l'enquête, WT_DSGN=PERS_W2 et aucune unité n'a été retirée.

Ajustement pour la non-réponse – WT_NRA

Compte tenu du fait que l'échantillon pour la SEPC était formé par les personnes ayant accepté de participer à un panel Web, les taux de réponse à l'enquête étaient relativement élevés. De plus, le panel a été conçu pour produire des estimations au niveau national; les tailles d'échantillon par province n'étaient donc pas très grandes. Ainsi, la non-réponse était plutôt uniforme dans de nombreuses provinces. Les GRH correspondaient à une combinaison quelconque de facteurs (groupe d'âge, sexe, niveau de scolarité, situation de locataire, situation selon l'EPA, présence d'enfants dans le ménage, taille du ménage admissible et premier chiffre du code de la Classification nationale des professions pour les répondants qui ont un emploi). Un facteur d'ajustement a été calculé à l'intérieur de chaque groupe de réponse de la façon suivante :

$$\frac{\text{Somme des poids des répondants et des non – répondants}}{\text{Somme des poids des répondants}}$$

Les poids des répondants ont été multipliés par ce facteur pour produire les poids WT_NRA, ajustés pour tenir compte de la réponse à l'enquête. Les unités non répondantes ont été retirées de l'enquête.

Étalonnage des poids-personne – WT_FINL

Les totaux de contrôle ont été calculés en utilisant les données de projection démographique de l'EPA. Durant l'étalonnage, un facteur d'ajustement est calculé et appliqué aux poids d'enquête. Cet ajustement est calculé afin que les sommes pondérées reflètent les totaux de contrôle. La plupart des enquêtes sociales étalonnent les poids au niveau personne aux totaux de contrôle par sexe, groupe d'âge et province. Pour SEPC4, l'étalonnage par province n'a pas été possible, car il y avait très peu de répondants dans certaines catégories dans les provinces de l'Atlantique et des Prairies. En outre, on a dénombré très peu de répondants de sexe masculin âgés de 15 à 24 ans dans les provinces de l'Atlantique, de sexe masculin âgés de 15 à 24 ans en Colombie-Britannique et de sexe féminin âgées de 15 à 24 ans en Colombie-Britannique. Pour cette raison, les totaux de contrôle utilisés pour SEPC4 – Les sources d'information consultées pendant la pandémie ont été établis par groupe d'âge, sexe et région géographique. Chez les hommes de la région de l'Atlantique, et les hommes et femmes en Colombie-Britannique, les deux groupes d'âge les plus jeunes ont été combinés. La prochaine section comprendra des recommandations pour l'analyse selon la région géographique et le groupe d'âge.

5.3 Poids bootstrap

Des poids bootstrap ont été créés pour le panel et les répondants à l'enquête SEPC4 – Les sources d'information consultées pendant la pandémie. Les poids bootstrap de l'EPA étaient les poids initiaux et tous les ajustements appliqués aux poids de l'enquête ont aussi été appliqués aux poids bootstrap.

6.0 Qualité de la SEPC et vérifications de l'enquête

Le panel probabiliste créé pour la SEPC est un projet pilote entrepris par Statistique Canada en 2019. Alors que le panel permet de recueillir des données rapidement, en profitant d'un ensemble de répondants qui ont déjà accepté de participer à de multiples courtes enquêtes en ligne, et qui possèdent des adresses courriel permettant d'accélérer la collecte des données d'enquête, certains aspects du concept de la SEPC font que les données obtenues risquent davantage d'être biaisées. Le taux de participation au panel est inférieur au taux de participation habituel aux enquêtes sociales menées par Statistique Canada, ce qui augmente le potentiel de biais de non-réponse. Qui plus est, puisque les enquêtes de la SEPC sont toutes des enquêtes en ligne dont le questionnaire est à remplir soi-même, les personnes qui n'ont pas d'accès Internet n'ont pas les moyens de participer à la SEPC et ne sont donc pas couvertes.

Lorsque le panel non pondéré a été comparé à l'échantillon d'origine à qui on a demandé de se joindre au panel, en particulier, on a observé une sous-représentation des personnes âgées de 15 à 24 ans, des personnes âgées de 65 ans et plus et de celles qui étaient sans diplôme d'études secondaires. Ces différences étaient attendues en raison de la nature du panel et de l'expérience des exemples internationaux de panels probabilistes. L'utilisation des ménages ayant répondu à l'EPA comme base de sondage pour le panel visait en premier lieu à tirer profit des renseignements disponibles de l'EPA pour corriger la sous-représentation et la surreprésentation observées au sein du panel. Les ajustements pour

tenir compte de la non-réponse effectués dans le cadre des ajustements de pondération du panel et les répondants à l'enquête ont utilisé les renseignements disponibles pour s'assurer que les poids des unités non répondantes/non participantes étaient attribués à des unités répondantes similaires. En outre, l'étalonnage selon les totaux par âge et par sexe a contribué à tenir compte de la sous-représentation par groupe d'âge.

Le tableau 6.1 montre les taux de glissement par certains domaines après étalonnage de l'enquête SEPC4 – Les sources d'information consultées pendant la pandémie. Le taux de glissement se calcule en comparant la somme des poids dans le domaine à celle du total de contrôle en fonction des projections démographiques. Un taux de glissement positif signifie que l'échantillon comporte un surdénombrement pour ce domaine. Un taux de glissement négatif signifie que l'enquête comporte un sous-dénombrement pour ce domaine. Selon les résultats indiqués aux tableaux 6.1 et 6.2, il est recommandé d'utiliser uniquement les données aux niveaux géographiques et dans les groupes d'âge ayant un taux de glissement de 0. À l'échelle nationale, selon la région géographique (provinces de l'Atlantique, Québec, Ontario, provinces des Prairies et Colombie-Britannique) et les quatre groupes d'âge les plus âgés.

Tableau 6.1 Taux de glissement par région géographique

Région	Domaine	n	Taux de glissement
Région géographique	Canada*	4 218	0 %
	Île-du-Prince-Édouard	101	12,0 %
	Terre-Neuve-et-Labrador	114	-10,1 %
	Nouvelle-Écosse	244	2,5 %
	Nouveau-Brunswick	192	1.3 %
	Québec	701	0 %
	Ontario	1 246	0 %
	Manitoba	338	-3,0 %
	Saskatchewan	279	3,4 %
	Alberta	445	0 %
	Colombie-Britannique	558	0 %

**Selon les 10 provinces; les territoires sont exclus*

Tableau 6.2 Taux de glissement par groupe d'âge

Région	Domaine	n	Taux de glissement
Groupe d'âge	15 à 24	270	18.7 %
	25 à 34	417	-15.7 %
	35 à 44	681	0 %
	45 à 54	680	0 %
	55 à 64	968	0 %
	65 et plus	1 202	0 %

Après la collecte des données de la SEPC4 – Les sources d'information consultées pendant la pandémie, une brève étude sur le biais a été menée pour évaluer le biais potentiel attribuable aux plus

faibles taux de réponse et au sous-dénombrement de la population hors ligne. Les données de l'EPA ont été utilisées pour produire des estimations pondérées pour l'échantillon dans le champ de l'enquête visé pour se joindre au panel probabiliste (en utilisant les poids et l'échantillon de PERS_W1). Les mêmes données ont été utilisées pour produire des estimations pondérées fondées sur l'ensemble de répondants de l'enquête SEPC4 et les poids de WT_FINL. Les deux ensembles d'estimations ont été comparés et sont présentés dans le tableau 6.3. Les différences significatives sont mises en évidence.

Tableau 6.3 Variations des estimations attribuables à la non-participation à la SEPC et à l'enquête sur la COVID-19

Sujet	Variables recodées de l'EPA de 2019	Estimation de la population visée par l'enquête (n = 31 628)	Estimation pour la vague 3 de la SEPC (n = 4 218)	Différence en points de pourcentage
Niveau de scolarité	Sans diplôme d'études secondaires*	15,5 %	12,5 %	3,0 %
	Diplôme d'études secondaires	25,9 %	26,5 %	-0,6 %
	Diplôme d'études postsecondaires*	58,6 %	61,0 %	-2,4 %
Situation d'activité	Personne employée	61,2 %	62,3 %	-1,1 %
	Personne en chômage	3,4 %	3,5 %	-0,1 %
	Personne inactive	35,3 %	34,1 %	1,1 %
Pays de naissance	Canada*	71,7 %	75,7 %	-4,0 %
État matrimonial	Personne mariée ou vivant en union libre*	60,4 %	60,9 %	-0.5 %
	Personne divorcée, séparée ou veuve*	12,8 %	11,4 %	1,4 %
	Personne célibataire, jamais mariée	26,9 %	27,7 %	-0,9 %
Enfants	Présence d'enfants	31,7 %	33,8 %	-2,1 %
Taille du ménage	Personne seule	14,4 %	13,9 %	0,5 %
	Ménage de deux personnes*	34,8 %	37,2 %	-2,5 %
	Ménage de trois personnes ou plus*	18,4 %	18,4 %	-0,1 %

Sujet	Variables recodées de l'EPA de 2019	Estimation de la population visée par l'enquête (n = 31 628)	Estimation pour la vague 3 de la SEPC (n = 4 218)	Différence en points de pourcentage
Personnes admissibles au panel	Une personne admissible âgée de 15 ans ou plus	15,9 %	15,6 %	0,3 %
	Deux personnes admissibles*	49,3 %	52,8 %	-3,5 %
	Trois personnes admissibles ou plus*	34,8 %	31,6 %	3,2 %
Logement	Appartement	12,1 %	11,4 %	0,7 %
	En location	24,8 %	23,4 %	1,4 %
Code de profession	Gestion (CNP0)	6,0 %	6,3 %	-0,2 %
	Affaires, finance et administration (CNP1)	10,7 %	11,8 %	-1,1 %
	Sciences naturelles et appliquées et domaines apparentés (CNP2)	5,2 %	6,2 %	-1,0 %
	Secteur de la santé (CNP3)	4,7 %	4,3 %	0,4 %
	Enseignement, droit et services sociaux, communautaires et gouvernementaux (CNP4)	7,6 %	8,5 %	-0,9 %
	Arts, culture, sports et loisirs (CNP5)	2,5 %	3,0 %	-0,6 %
	Vente et services (CNP6)	16,6 %	16,3 %	0,3 %
	Métiers, transport, machinerie et domaines apparentés (CNP7)	9,6 %	9,6 %	-0,0 %
	Ressources naturelles, agriculture et production connexe (CNP8)*	1,6 %	1,1 %	0,5 %
	Fabrication et services d'utilité publique (CNP9)	2,9 %	2,5 %	0,4 %

**Estimations ayant une valeur significativement différente au niveau α= 5 %.*

Bien que de nombreuses estimations ne présentent pas de variation significative, les différences significatives montrent qu'il reste encore certains biais dans la SEPC4 – Les sources d'information consultées pendant la pandémie. Il y a une sous-représentation de ceux qui comptaient trois participants admissibles ou plus pour le panel, de personnes sans diplôme d'études secondaires, et de personnes qui travaillent en CNP8. En outre, il y a surreprésentation des personnes ayant un diplôme d'études postsecondaires, des personnes nées au Canada, des ménages comptant deux personnes au total et des ménages comptant deux participants admissibles pour le panel. Il serait important de garder ces petites différences à l'esprit lors de l'utilisation des données d'enquête de la SEPC4 – Les sources d'information consultées pendant la pandémie. Une étude sur les différences entre les estimations est en cours, et au fur et à mesure que des différences sont identifiées, des stratégies sont testées afin d'améliorer la méthodologie d'une vague de l'enquête à une autre.