

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

**Initiatives de l'Enquête sur la population
active dans le cadre du Plan d'action sur les
données désagrégées de Statistique Canada**

par Martin Pantel et Yelly Camara

Date de diffusion : le 25 mars 2024



Statistique
Canada

Statistics
Canada

Canada

Initiatives de l'Enquête sur la population active dans le cadre du Plan d'action sur les données désagrégées de Statistique Canada

Martin Pantel et Yelly Camara¹

Résumé

Conformément au Plan d'action sur les données désagrégées (PADD) à long terme de Statistique Canada, plusieurs initiatives ont été mises en œuvre dans l'Enquête sur la population active (EPA). Une des initiatives les plus directes fut une augmentation ciblée dans la taille de l'échantillon mensuel de l'EPA. En outre, un programme d'enquête supplémentaire régulier a été introduit, où une série supplémentaire de questions a été posée à un sous-ensemble de répondants de l'EPA et analysée dans un cycle de production mensuel ou trimestriel. Enfin, les estimations fondées sur les méthodologies de l'estimation sur petits domaines (EPD) sont réintroduites dans l'EPA et comprendront une portée élargie avec davantage de valeur analytique qu'auparavant. Le présent article donne un aperçu de ces trois initiatives.

Mots clés : Enquête sur la population active; estimation sur petits domaines; désagrégation

1. Contexte

L'Enquête sur la population active (EPA) est une enquête mensuelle qui sert à évaluer l'état actuel du marché du travail canadien. Il s'agit d'une source clé de données sur l'emploi, le taux de chômage et d'autres indicateurs du marché du travail. Son échantillon, composé d'environ 80 000 logements, est divisé en six groupes de renouvellement, dont l'un est remplacé chaque mois par de nouveaux logements appelés unités de naissance. La répartition de l'échantillon de l'enquête est fondée sur des cibles de qualité utilisant le coefficient de variation ($CV = \text{Erreur} - \text{type} / \text{Estimation}$).

Le présent article décrit trois initiatives mises en œuvre dans l'EPA conformément au Plan d'action sur la désagrégation des données (PADD) à long terme de Statistique Canada, une initiative à l'échelle de l'organisme dont les objectifs consistent notamment à fournir aux utilisateurs des données statistiques détaillées qui mettent en lumière les expériences vécues par des groupes de population particuliers, dont les femmes, les peuples autochtones, les populations racisées et les personnes ayant une incapacité. La section 2 présente le programme d'enquête supplémentaire, un nouveau produit analytique qui peut être adapté pour tenir compte des nouvelles tendances du marché du travail. La section 3 décrit l'intégration d'un échantillon supplémentaire à un plan d'enquête par panel complexe et elle fournit des analyses préliminaires de l'incidence de cette intégration sur la qualité des données. La section 4 de l'article présente le programme de modélisation de l'estimation sur petits domaines associé à l'EPA. Enfin, la dernière section envisage les perspectives de travaux possibles sur les programmes de modélisation des estimations sur petits domaines et des enquêtes supplémentaires.

2. Programme d'enquêtes supplémentaires de l'EPA

Des données supplémentaires mensuelles ont commencé à être recueillies régulièrement en avril 2020. Régulier et dynamique, le programme d'enquêtes supplémentaires actuellement utilisé a commencé à recueillir des données mensuelles en janvier 2022 et des données trimestrielles en juillet 2022. Pour contrôler le fardeau global, les questions sont posées à un sous-ensemble de répondants de l'EPA. En effet, comme le montre le tableau 2-1., les données des

¹Martin Pantel, Statistique Canada, 100, promenade Tunney's Pasture, Ottawa (Ontario), Canada, K1A 0T6 (martin.pantel@statcan.gc.ca); Yelly Camara, Statistique Canada, 100, promenade Tunney's Pasture, Ottawa (Ontario), Canada, K1A 0T6 (yelly.camara@statcan.gc.ca).

enquêtes supplémentaires mensuelles sont recueillies auprès des répondants au cours de leur deuxième à cinquième mois dans l'échantillon, c.-à-d. au 4/6 de l'échantillon, et pour les enquêtes supplémentaires trimestrielles, auprès des répondants au cours de leur dernier mois dans l'échantillon. L'admissibilité dépend également de la situation d'activité sur le marché du travail : par exemple, les questions sur le travail à domicile ne seraient posées qu'aux répondants ayant un emploi. Les données collectées sont traitées au moyen de stratégies appropriées d'imputation et de repondération.

Tableau 2-1 : Sujets du programme d'enquêtes supplémentaires

		Mois dans l'échantillon		
		1	2 – 5	6
Janv.	x	Travail à domicile (TàD), Intention de quitter son emploi		Enquête canadienne sur le revenu
Fév.	x	TàD, Intention de démissionner		
...				
Juin	x	TàD, Avantages sociaux		Entrepreneurs dépendants, Travailleurs temporaires
Juil.	x	TàD, Formation, Maintien en poste du personnel		
Août	x	TàD, Éthique de travail		
Sept.	x	TàD, Garde d'enfants		
Oct.	x	TàD, Coût de la vie		(Enquêtes trimestrielles)
...				

Pendant les premiers mois de la pandémie, l'un des objectifs de l'EPA était de mieux mesurer les répercussions de la COVID-19 sur le marché du travail, notamment les absences liées à la COVID-19, le travail à domicile, la réception des prestations (p. ex. la Prestation canadienne d'urgence ou PCU), et les conséquences inégales sur les sous-populations (p. ex. les groupes racisés). À l'heure actuelle, ce programme permet aussi de mettre à l'essai du contenu qui pourrait éventuellement être transféré à l'EPA principale. Par exemple, l'identification des membres de groupes racisés a été intégrée au questionnaire principal en janvier 2022, de même que des méthodologies de traitement appropriées (imputation et calage des poids). La figure 2-1 montre l'incidence sur Terre-Neuve-et-Labrador (T.-N.-L.) du transfert de la question à l'EPA principale en janvier 2022. Avant cette date (avant la mise en œuvre du programme d'enquêtes supplémentaires actuel), l'identification des groupes racisés était effectuée pour cinq des renouvellements. Comme on pouvait s'y attendre, le passage de 5 à 6 renouvellements a amélioré la qualité des estimations à Terre-Neuve-et-Labrador. Le coefficient de variation (c.v.) du nombre d'emplois est passé de 22 % en décembre 2021 à 8 % en janvier 2022. Comme l'illustre le tableau 2-2, la variance plus petite signifie un intervalle de confiance (IC) plus étroit, ce qui facilite la détection des différences d'un mois à l'autre et des variations entre les groupes.

Figure 2-1 : Qualité de l'estimation des personnes occupées (c.v.), pour les groupes racisés, à Terre-Neuve-et-Labrador

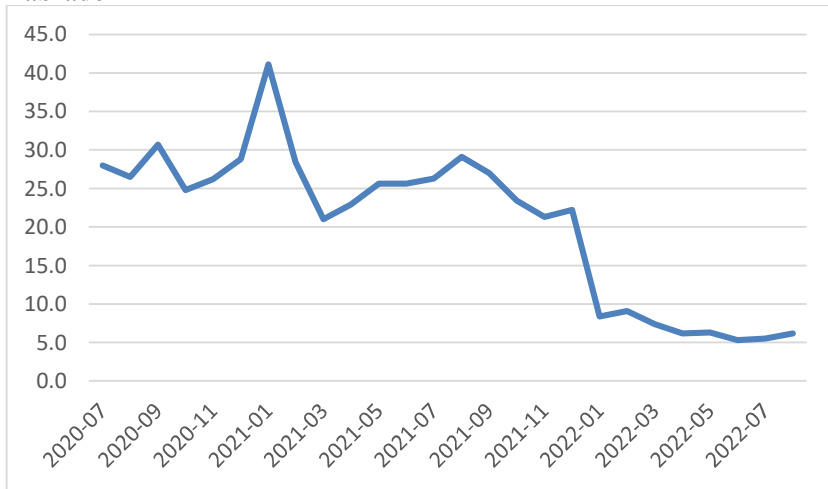


Tableau 2-2 : Qualité de l'estimation des personnes occupées (IC), pour les groupes racisés, à Terre-Neuve-et-Labrador

	Taille de l'échantillon	Estimation	95 % LB (borne inférieure)	95 % UB (borne supérieure)
Décembre 2021	55	8 600	4 600	12 000
Janvier 2022	92	8 200	6 800	9 500

3. Augmentation de l'échantillon

3.1 Objectifs

L'augmentation de l'échantillon visait une plus grande désagrégation démographique et géographique. L'objectif était d'améliorer la qualité des estimations pour les groupes racisés et celle des estimations au niveau des villes. Cependant, la faible taille de la population de la plupart des groupes racisés, leur concentration régionale et l'absence de registre au niveau de la personne ne permettaient pas de le faire efficacement dans le cadre actuel de l'EPA. En effet, pour obtenir un c.v. de 25 % pour un taux de chômage moyen mobile de trois mois de 5 % pour les groupes philippin, arabe, latino-américain et d'Asie du Sud-Est, il aurait fallu doubler la taille de l'échantillon de l'Ontario.

C'est pourquoi l'accent a été mis sur le suréchantillonnage des plus grandes subdivisions de recensement (SDR), qui sont à peu près équivalentes aux municipalités. Cela devait permettre non seulement une désagrégation géographique, mais aussi un suréchantillonnage indirect des membres des groupes racisés, car ils sont, pour un grand nombre, surreprésentés dans les SDR les plus grandes.

3.2 Mise en œuvre

L'EPA est conçue avec soin pour effectuer un renouvellement dans sa base de sondage en 10 ans, tout en permettant à d'autres enquêtes de sélectionner des échantillons qui ne se chevauchent pas. De manière à s'adapter à ce plan, la méthode à deux cibles suivante a servi à calculer la taille supplémentaire de l'échantillon :

- un c.v. cible de 16,5 %, pour un taux de chômage moyen mobile sur trois mois de 5 %, pour les 20 plus grandes SDR;
- un c.v. cible de 25 %, pour un taux de chômage moyen mobile sur trois mois à un niveau de 5 %, pour les SDR 21 à 23 et les deux plus grandes dans chaque province (si elles ne font pas déjà partie des 23 premières).

Au moyen de simulations utilisant les données d'août 2020 à janvier 2021, on a déterminé que 13 300 logements devaient être ajoutés à l'échantillon. Ils ont tous été attribués aux provinces et intégrés graduellement à notre échantillon existant, un groupe de renouvellement à la fois, de novembre 2021 à avril 2022, en augmentant le nombre d'unités de naissance (unités dans leur premier mois de collecte).

Comme le plan de sondage n'a pas été modifié, les objectifs de qualité originaux de l'EPA sont automatiquement atteints.

3.3 Résultats

3.3.1 Incidence sur la qualité du nombre d'emplois

L'augmentation de l'échantillon a amélioré la qualité des estimations pour les SDR ciblées. Les figures 3-1 et 3-2 montrent la qualité des estimations pour deux des SDR ciblées, à mesure que l'échantillon supplémentaire était mis en place progressivement au cours de la période de novembre 2021 à avril 2022. Le c.v. du nombre d'emplois à Burnaby est passé de 14 % à 8 %, une diminution de 6 points de pourcentage. À Toronto, où il y avait déjà un grand échantillon (et un petit c.v.) avant l'initiative du PADD, le c.v. du nombre d'emplois a diminué d'un point de pourcentage pour s'établir à 3 %.

Figure 3-1 : c.v. du nombre d'emplois à Burnaby

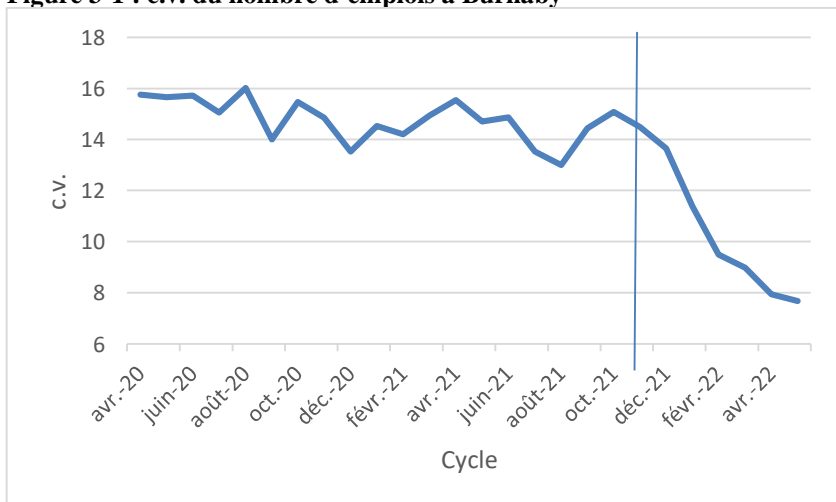
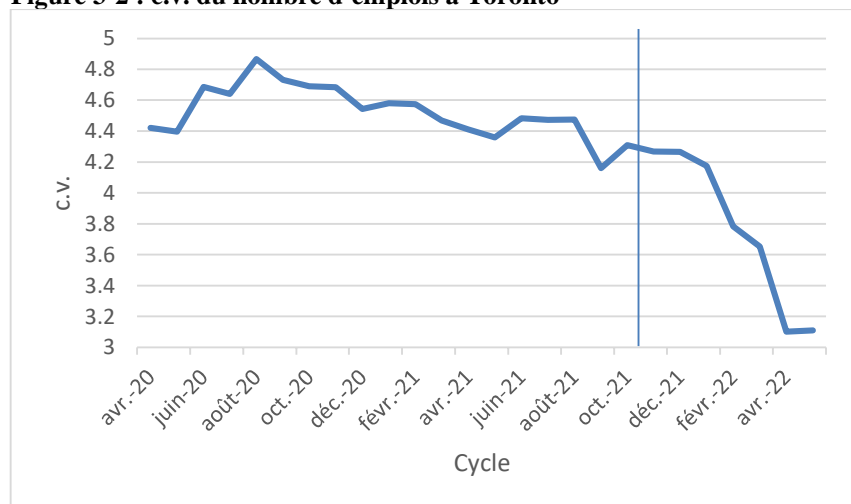


Figure 3-2 : c.v. du nombre d'emplois à Toronto



Étant donné que bon nombre des SDR ciblées comptaient une forte proportion de membres de groupes racisés, par exemple 51 % à Toronto et 63 % à Burnaby comparativement à 22 % pour l'ensemble du Canada (estimations du Recensement de 2016), l'augmentation de l'échantillon a contribué à l'amélioration de la qualité des estimations pour ces groupes.

3.3.2 Incidence sur le nombre de répondants immigrants

Avec plus de 15 000 répondants supplémentaires chaque mois, l'augmentation de l'échantillon a également eu un effet sur la proportion d'immigrants répondants dans l'échantillon recueilli (sans pondération). Avant l'augmentation de l'échantillon, environ 18 % de l'échantillon recueilli était constitué de personnes nées hors du Canada. Or ce même groupe représente environ 37 % des 15 000 répondants supplémentaires. La méthodologie de pondération tient compte des répondants supplémentaires, de sorte que les estimations pondérées ne sont pas touchées par l'augmentation de l'échantillon. Cependant, étant donné qu'un plus grand nombre d'immigrants contribuent aux estimations, il y a une augmentation de la qualité statistique des estimations concernant la population immigrante et les groupes racisés.

Le c.v. du nombre d'emplois chez les immigrants s'est amélioré, passant de 1,5 % en octobre 2021 à 1,2 % en mai 2022 à l'échelle nationale. À Markham (l'une des SDR ciblées), le c.v. du nombre d'emplois chez les immigrants s'est amélioré, passant de 24 % en octobre 2021 à 11 % en mai 2022.

4. Programme de modélisation des estimations sur petits domaines

L'EPA est conçue pour produire des estimations de grande qualité pour des agrégations géographiques clés. L'estimation sur petits domaines ou EPD utilise des données auxiliaires et des modèles statistiques (modèles au niveau du domaine de Fay-Herriot) pour produire des estimations fiables concernant les petites régions géographiques susceptibles de ne pas être ciblées par le plan de sondage de l'EPA.

Jusqu'en février 2020, ces modèles servaient à produire :

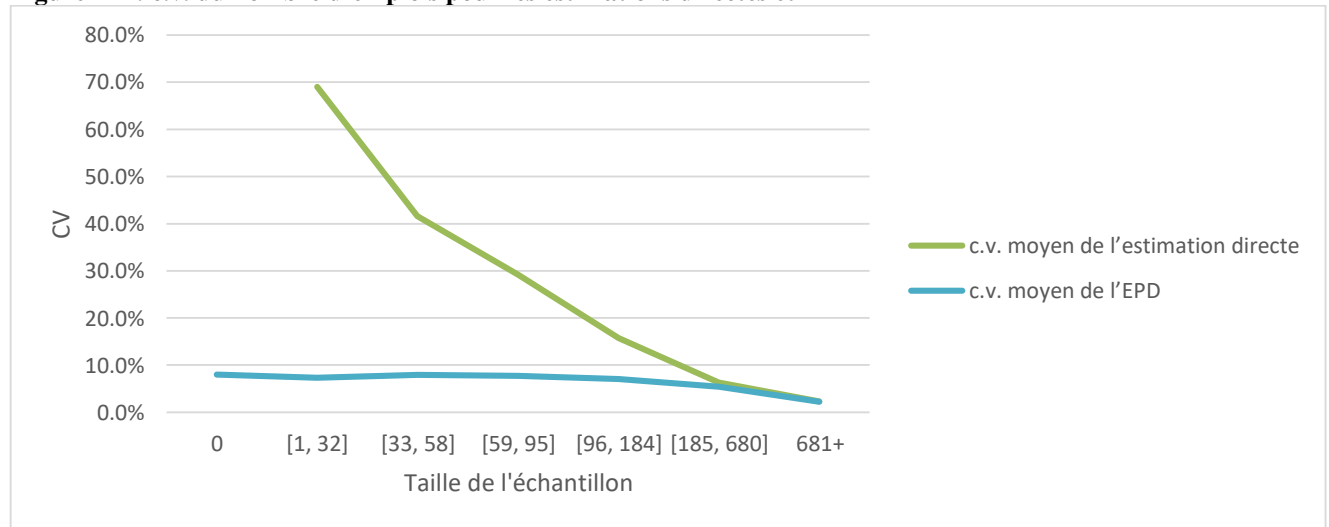
- des estimations du taux de chômage et du nombre d'emplois pour chacune des 149 régions métropolitaines de recensement et agglomérations de recensement;
- des estimations des salaires horaires pour les combinaisons de régions économiques et de professions.

La variable auxiliaire utilisée dans les modèles était le nombre de prestataires d'assurance-emploi (AE). Pendant la pandémie, de nouvelles prestations ont été instaurées par le gouvernement, et certaines personnes n'étaient pas certaines de leur propre statut sur le marché du travail. La question se posait alors de savoir si les données de l'assurance-emploi représentaient le chômage de la même façon qu'avant la pandémie. Le programme d'EPD a été

suspendu jusqu'à ce que des données auxiliaires fiables puissent de nouveau être obtenues. La disponibilité récente de données sur l'assurance-emploi ainsi que les objectifs établis par l'initiative du PADD ont permis de reprendre ce programme de modélisation.

La valeur ajoutée des EPD se constate quand on examine la qualité des estimations du nombre d'emplois, comme l'illustre la figure 4-1 ci-dessous. Nous observons des résultats semblables pour les estimations du taux de chômage et du salaire horaire. Quand la taille de l'échantillon recueilli est petite (de 1 à 32 personnes), le c.v. des estimations directes $\left(\frac{\sqrt{\text{estimation de la variance directe}}}{\text{estimation directe}}\right)$ est parfois si grand – près de 70 % – que l'utilisation ou la diffusion des estimations peut ne pas être recommandée. Grâce à la modélisation de l'EPD, il est possible de produire des estimations de bonne qualité pour de petits échantillons collectés, y compris dans des domaines dont la taille de l'échantillon est de zéro. Comme pour toute application de l'EPD au moyen de modèles de Fay-Herriot, les gains diminuent à mesure que la taille de l'échantillon recueilli augmente. Avec un échantillon important, le c.v. de l'EPD $\left(\frac{\sqrt{\text{erreur quadratique moyenne d'EPD}}}{\text{estimation directe}}\right)$ est proche du c.v. direct, et l'estimation de l'EPD s'approche de l'estimation directe.

Figure 4-1 : c.v. du nombre d'emplois pour les estimations directes et l'EPD



5. Travaux futurs

Les travaux futurs pour les enquêtes supplémentaires devraient porter sur de nouveaux sujets, notamment le stress financier, la formation, le travail sur les plateformes numériques, le travail à la demande (informel), les paiements d'aides et les dépenses de garde d'enfants. Pour le programme de modélisation des estimations sur petits domaines, de nouvelles géographies, de nouvelles sources de données administratives et des nouveaux indicateurs du marché du travail seront examinés.

Remerciements

Nous tenons à remercier Andrew Brennan, Dany Faucher, Tom Haymes, Wisner Jocelyn, Martin Lemire, Charles Mitchell, Isabelle Thony et François Verret pour leurs précieux commentaires.