

## Techniques d'enquête

### Commentaires à propos de l'article « Progrès de la science et de la pratique des enquêtes : hier, aujourd'hui et demain »

par Constance F. Citro

Date de diffusion : le 30 juin 2025



Statistique  
Canada

Statistics  
Canada

Canada

---

## Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à [www.statcan.gc.ca](http://www.statcan.gc.ca).

Vous pouvez également communiquer avec nous par :

**Courriel** à [infostats@statcan.gc.ca](mailto:infostats@statcan.gc.ca)

**Téléphone** entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

- |   |                |
|---|----------------|
| • Service de renseignements statistiques                                    | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur   | 1-514-283-9350 |

## Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site [www.statcan.gc.ca](http://www.statcan.gc.ca) sous « Contactez-nous » > « [Normes de service à la clientèle](#) ».

## Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Publication autorisée par la ministre responsable de Statistique Canada

© Sa Majesté le Roi du chef du Canada, représenté par la ministre de l'Industrie, 2025

L'utilisation de la présente publication est assujettie aux modalités de l'[entente de licence ouverte](#) de Statistique Canada.

**Une [version HTML](#) est aussi disponible.**

*This publication is also available in English.*

---

## Commentaires à propos de l'article « Progrès de la science et de la pratique des enquêtes : hier, aujourd'hui et demain »

Constance F. Citro<sup>1</sup>

### Résumé

Dans son essai sur les défis que pose le paradigme des enquêtes par échantillonnage probabiliste (ou la tradition de recherche), Carl-Erik Särndal cite l'article que j'ai rédigé en 2014 et qui a été publié dans la présente revue, et dans lequel j'insiste sur la nécessité d'adopter un paradigme de sources de données multiples (ou de données mélangées). Mon intention n'est pas de déclasser les enquêtes probabilistes, mais de les combiner avec les dossiers administratifs et d'autres sources afin d'améliorer la qualité et la pertinence des données. Depuis la rédaction de l'article cité, les États-Unis ont réalisé des progrès en matière de données mélangées.

**Mots-clés :** Analyses de sensibilité; données mélangées; dossiers administratifs.

J'ai l'honneur de commenter les réflexions éclairantes et inspirantes de Carl-Erik Särndal sur l'histoire des enquêtes par échantillonnage probabiliste dans la production de statistiques officielles, entre autres utilisations. Son essai « conceptuel et exploratoire » recadre la science des enquêtes probabilistes dans le contexte plus large de l'évolution scientifique. L'essai rassure les théoriciens et les praticiens des enquêtes en montrant que les défis que pose le paradigme des enquêtes probabilistes, ou la tradition de recherche, pour utiliser le terme privilégié par Särndal, ne sont pas uniques dans l'histoire des sciences. L'essai peut néanmoins soulever des inquiétudes, car il illustre la réalité de l'évolution scientifique : des paradigmes universellement acceptés doivent évoluer ou disparaître quand leurs fondements sont minés, comme cela s'est sans doute produit dans le cas des enquêtes probabilistes, compte tenu des faibles taux de réponse et des coûts élevés.

M. Särndal fait mention de mon article, lequel a été publié dans la présente revue (Citro, 2014). Dans cet article, je demandais aux organismes nationaux de statistique de changer de paradigme, c'est-à-dire de passer de celui des enquêtes par échantillonnage probabiliste à celui des sources de données multiples. M. Särndal estime que mon appel laisse au lecteur le soin de l'interpréter et suggère « une frustration ou une impatience à l'égard de la façon de penser habituelle ». Je plaide coupable aux deux chefs d'accusation et je suis heureuse de disposer de la présente occasion pour expliquer davantage ma pensée.

En préconisant l'adoption d'un paradigme de sources de données multiples, plus communément appelé un paradigme de « données mélangées », je n'avais pas l'intention d'exclure ni même de déclasser les enquêtes par échantillonnage probabiliste. En effet, ces dernières permettent de recueillir des données d'une importance cruciale afin de produire bon nombre des estimations que les organismes nationaux de statistique fournissent aux utilisateurs pour répondre à leurs besoins en données. Rien ne remplace le fait de sonder les

---

1. Constance F. Citro, chercheuse principale, Committee on National Statistics, National Academies of Sciences, Engineering, and Medicine, Keck Center, 500 5th St, NW (Washington), DC 20001, États-Unis. Courriel : ccitro@nas.edu.

gens à propos de leurs opinions, de leurs attitudes et de leurs caractéristiques, par exemple en leur posant des questions sur leur race ou leur origine ethnique, leur état de santé, ou encore les tâches qu'ils accomplissent simultanément (par exemple s'occuper des enfants et faire de l'exercice en courant ou en marchant avec une poussette). De plus, malgré les défis que posent les plans probabilistes, les échantillons non probabilistes n'ont pas, selon moi, démontré leur capacité à produire des estimations suffisamment robustes pour les statistiques officielles. Les avantages théoriques et souvent encore pratiques des enquêtes probabilistes à des fins d'estimation sont la raison pour laquelle j'ai inclus « probabiliste » comme modificateur dans le terme « paradigme des enquêtes par échantillonnage probabiliste ».

Dans le paradigme des données multiples ou mélangées que je préconise, les enquêtes-échantillons, conçues pour être aussi robustes que possible compte tenu des limites des ressources et des difficultés à obtenir une réponse adéquate, sont souvent la « colonne vertébrale », pour ainsi dire, des ensembles de données mélangées. Un aspect important du paradigme des données mélangées est qu'il comprend aussi d'autres sources de données, comme des dossiers administratifs des pouvoirs publics et d'entreprises ou des données extraites d'Internet, lesquelles permettent de régler le problème croissant que pose le fait que les répondants aux enquêtes omettent souvent de répondre à des questions clés ou fournissent des réponses inexactes. Des années de recherche et d'expérimentation ont permis de cerner des éléments cruciaux dont les niveaux d'erreur sont complexes et, souvent, de plus en plus importants : il s'agit notamment du revenu, de la richesse, des dépenses et des problèmes de santé. Mon impatience à l'égard de « la façon de penser habituelle » provient de l'effort disproportionné qui avait été consacré, à mon avis, aux questions de pondération et d'estimation des enquêtes pour neutraliser les effets du biais et de la variabilité de la non-réponse des unités, comparativement à l'effort consacré à l'intégration de sources de données qui seraient plus exactes que les réponses à une enquête pour des éléments clés.

Encore une fois, mon intention n'était pas de m'opposer aux travaux en matière de pondération et d'estimation, mais plutôt de plaider en faveur d'un effort méthodologique proportionnel de la part des organismes nationaux de statistique, dans le but d'intégrer de façon appropriée d'autres sources de données qui pourraient régler de graves problèmes de déclaration concernant des statistiques d'enquête clés. Des travaux sont nécessaires dans des domaines comme : la documentation et l'évaluation d'autres sources de données pour bien comprendre leurs utilisations et leurs limites; l'adaptation des systèmes de traitement pour intégrer harmonieusement d'autres sources aux réponses d'enquête, notamment en prévoyant que certaines sources ne sont disponibles que de manière décalée dans le temps; et l'élaboration de mesures de l'erreur pour d'autres sources qui contribuent à une mesure de l'erreur totale en plus des mesures traditionnelles de l'erreur d'échantillonnage et de non-réponse.

J'ai rédigé l'article cité par Särndal en 2014 dans le contexte des États-Unis. À l'époque, les séries statistiques officielles des États-Unis fondées sur des enquêtes, comme bon nombre d'entre elles l'étaient et

le sont toujours, présentait des résultats d'enquête pondérés comme s'il s'agissait des résultats souhaités, malgré de graves erreurs de déclaration et souvent de graves erreurs de couverture (par exemple l'incapacité de saisir la strate la plus élevée des personnes riches dans l'échantillon). En général, les organismes de statistique ne commençaient pas par déterminer les données dont avaient besoin les utilisateurs, pour ensuite trouver la combinaison de sources de données la plus susceptible de fournir ces résultats. Les seules exceptions à la publication d'estimations d'enquête malgré de graves erreurs sont les estimations du produit intérieur brut, du revenu personnel, de la consommation personnelle, et d'autres statistiques similaires, produites dans le cadre conceptuel des National Income and Product Accounts (comptes nationaux des revenus et des produits). L'objectif de ces estimations a toujours été l'exhaustivité, ce qui implique nécessairement de s'appuyer sur plusieurs enquêtes et dossiers administratifs pour obtenir la meilleure estimation pour chaque composante (voir [NIPA Handbook: Concepts and Methods of the U.S. National Income and Product Accounts | U.S. Bureau of Economic Analysis \[BEA\]](#)).

Je suis heureuse de souligner qu'au cours des 10 années qui se sont écoulées depuis la rédaction de mon article sur le paradigme des données multiples, le système statistique des États-Unis a fait d'importants progrès dans cette direction. Or, ces progrès ont été ralentis par les obstacles juridiques qui empêchent les organismes du système décentralisé des États-Unis d'avoir facilement accès aux dossiers administratifs fédéraux, étatiques et locaux dont ils ont besoin. De plus, à ce stade, l'élaboration d'ensembles de données mélangées par les organismes statistiques des États-Unis est souvent expérimentale, ce qui signifie que des changements de priorités ou de budgets pourraient mettre un terme à des avancées prometteuses. Néanmoins, le changement de point de vue est réel (voir, par exemple, National Academies of Sciences, Engineering, and Medicine, 2017a, b; 2023a, b) et les nombreuses validations de concept sont encourageantes. L'encadré 1 résume un exemple de progrès qui est comparable à l'un des exemples de problèmes mentionnés dans mon article de 2014, soit celui du revenu du ménage. Citons entre autres exemples : l'utilisation prévue de dossiers administratifs en remplacement de certaines questions sur le logement dans l'American Community Survey (voir [Federal Register: Agency Information Collection Activities; Submission to the Office of Management and Budget \(OMB\) for Review and Approval; Comment Request; American Community Survey and Puerto Rico Community Survey](#)), ainsi que les couplages courants de dossiers administratifs (par exemple Medicare et Medicaid) avec les enquêtes sur la santé du National Center for Health Statistics (NCHS, voir le [tableau de données couplées du NCHS \(cdc.gov\)](#)).

**ENCADRÉ 1 – Données mélangées aux fins d'amélioration des estimations du revenu**

L'enquête phare du U.S. Census Bureau (Bureau du recensement des États-Unis) sur le revenu et la pauvreté, intitulée Current Population Survey Annual Social and Economic Supplement, est connue pour ses niveaux élevés de données manquantes et mal déclarées (largement sous-déclarées) sur le revenu, constatés pour de nombreuses sources. À l'heure actuelle, près de 40 % du revenu total des ménages est estimé au moyen de cette enquête, et pourtant, pour le revenu de la propriété, le revenu de retraite et le revenu sous forme de transferts, la sous-déclaration nette reste importante. Le Bureau du recensement des États-Unis a mené des recherches approfondies pour documenter ces phénomènes, mais jusqu'à tout récemment, il ne s'était pas engagé à utiliser des données administratives pour remédier à la situation. En février 2023, s'appuyant sur des années de recherche, le Bureau a lancé le programme intitulé National Experimental Wellbeing Statistics afin d'améliorer les estimations du revenu en combinant des données administratives et des données d'enquête. Ce premier ensemble de statistiques a permis d'estimer le revenu à partir de la Current Population Survey de 2018 en utilisant une grande variété de sources administratives afin d'améliorer la pondération et les imputations et de neutraliser la sous-déclaration. Le résultat a été spectaculaire : le revenu médian des ménages avant impôt a augmenté de plus de 6 %, et le taux officiel de pauvreté a diminué de 1,1 point de pourcentage par rapport aux statistiques publiées; le revenu médian avant impôt des ménages composés de personnes de 65 ans et plus a augmenté de plus de 27 %, et leur taux de pauvreté a diminué de 3,3 points de pourcentage. Les prochaines étapes comprennent la mise à jour de la série et le traitement des sources de revenus – en particulier le travail autonome – qui ne sont pas correctement mesurées dans les enquêtes et les dossiers administratifs.

Référence : [sehsd-wp2023-02.pdf \(census.gov\)](#)

L'essai de Särndal met l'accent sur les difficultés que posent le plan de sondage et l'estimation probabilistes, compte tenu des coûts engendrés pour obtenir un taux de réponse acceptable et du fait que le biais et la variance demeurent souvent plus élevés que ce que les utilisateurs souhaiteraient. Je soutiens le cadre probabiliste, mais je n'ai pas de solution à ces problèmes. Je propose toutefois, pour les plans probabilistes et non probabilistes, que la profession des méthodes statistiques et d'enquête élabore des normes qui aideront les utilisateurs à comprendre le poids qu'il faut attribuer à chacun des ensembles d'estimations. Par exemple, les normes pourraient préciser la quantité minimale et les types d'analyses de sensibilité (comme les diverses hypothèses au sujet de l'exactitude de la déclaration d'un élément clé ou de la suffisance de la couverture d'un groupe de population) que les producteurs de données doivent effectuer et documenter pour que les utilisateurs fassent suffisamment confiance aux données. Lorsque les taux de réponse sont très faibles et que la représentativité est suspecte, les normes pourraient également exiger que

les estimations publiées s'accompagnent de la mention « 10 % à 15 % [ou même 10 % à 20 %] de la population déclare ceci ou cela ». De telles formulations tiendraient compte non seulement de l'erreur découlant d'un échantillonnage traditionnel, mais aussi des résultats des analyses de sensibilité ou d'autres façons de calculer l'erreur totale (voir l'encadré 2 pour obtenir un exemple du type de série de données qui pourrait tirer parti de ce genre de normes).

### **ENCADRÉ 2 – Sondage éclair auprès des ménages des États-Unis**

Plusieurs organismes fédéraux, dirigés par le Bureau du recensement des États-Unis, ont pensé, conçu et mis en œuvre en un temps record un sondage éclair auprès des ménages, soit le Household « Pulse » Survey (HPS), afin de fournir des renseignements actuels sur les effets de la pandémie de COVID-19 sur la santé mentale, l'emploi, les dispositions prises en rapport avec la garde des enfants et de nombreux autres sujets. (Il y a aussi un sondage éclair mené auprès des petites entreprises, intitulé le Small Business Pulse Survey, et un sondage éclair auprès des écoles a récemment été mis œuvre, soit le School Pulse Panel.) Dans sa première phase, le HPS a été mené chaque semaine, du 23 avril 2020 au 21 juillet 2020, au moyen d'un questionnaire en ligne. Les phases suivantes consistent en deux semaines de collecte de données suivies de deux semaines de « pause ». Les données sont diffusées dans les 10 à 14 jours suivant leur collecte. Les échantillons d'environ 1 million de ménages pour chaque période de collecte proviennent d'adresses figurant dans le fichier principal des adresses du Bureau du recensement des États-Unis, qui contient des courriels et des numéros de téléphone (environ 86 % du total). Les taux de réponse sont faibles – de 5 % à 10 % – et la représentativité ne peut pas être garantie, comme l'indique le Bureau sur son site Web, même malgré les ajustements apportés aux poids pour compenser la non-réponse et la couverture. Les estimations sont qualifiées d'« expérimentales ». Bien que cela n'ait pas été réalisable au départ, étant donné l'objectif primordial d'une mise en œuvre et d'une diffusion rapides des estimations, le Household Pulse Survey pourrait tirer profit d'analyses de sensibilité qui saisiraient la variabilité plus large et le biais plus grand qui sont inévitables étant donné le plan utilisé.

Référence : [Household Pulse Survey \(COVID-19\) \(census.gov\)](https://www.census.gov/hhes/health/mental/2020/hps.html)

En conclusion, l'essai de Särndal est réfléchi et incite à la réflexion. J'ai hâte de voir comment évolueront les enquêtes par échantillonnage probabiliste, à la fois en tant qu'ensembles de données autonomes, ainsi que dans le cadre d'une approche fondée sur des données mélangées dans le but de produire des estimations plus complètes et plus exactes des principaux indicateurs des organismes nationaux de statistique.

## Remerciements

Les points de vue exprimés dans les commentaires sont uniquement ceux de l'auteure et ne doivent pas être attribués aux National Academies.

## Bibliographie

Citro, C.F. (2014). [Des modes multiples pour les enquêtes à des sources de données multiples pour les estimations](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2014002/article/14128-fra.pdf). *Techniques d'enquête*, 40(2), 151-181. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2014002/article/14128-fra.pdf>.

National Academies of Sciences, Engineering, and Medicine (2017a). *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps*. Washington, DC: The National Academies Press. DOI: <https://doi.org/10.17226/24893>.

National Academies of Sciences, Engineering, and Medicine (2017b). *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy*. Washington, DC: The National Academies Press. DOI: <https://doi.org/10.17226/24652>.

National Academies of Sciences, Engineering, and Medicine (2023a). *Toward a 21st Century National Data Infrastructure: Enhancing Survey Programs by Using Multiple Data Sources*. Washington, DC: The National Academies Press. DOI: <https://doi.org/10.17226/26804>.

National Academies of Sciences, Engineering, and Medicine (2023b). *Toward a 21st Century National Data Infrastructure: Mobilizing Information for the Common Good*. Washington, DC: The National Academies Press. DOI: <https://doi.org/10.17226/26688>.