

Les données administratives d'abord comme paradigme statistique pour les statistiques officielles canadiennes : signification, défis et possibilités¹

Eric Rancourt²

Résumé

Depuis des décennies, les organismes nationaux de statistique affirment qu'ils ont l'intention d'utiliser de plus en plus de données administratives, ce qu'ils ont d'ailleurs fait à divers degrés d'un programme à l'autre. Cependant, avec l'avènement de la révolution des données, il ne s'agit plus d'un souhait, d'une question secondaire, d'une méthode marginale ou d'une tendance croissante : c'est devenu le centre d'attention pour l'avenir des programmes. Que l'objectif soit d'accroître la pertinence, de réduire le fardeau de réponse, d'accroître l'efficacité ou de produire plus rapidement avec plus de détails, l'utilisation des données administratives (au sens le plus large) prolifère à un rythme effréné dans les systèmes statistiques et sans systèmes statistiques. Statistique Canada doit faire face au nouveau monde des données en se modernisant et en adoptant un paradigme des données administratives d'abord. Dans le présent article, nous tentons d'expliquer ce que cela signifie du point de vue statistique, de mettre en évidence quelques-uns des défis théoriques à relever et de souligner les possibilités qui y sont associées. Cela soulève aussi des questions législatives ainsi que d'importantes considérations liées aux éléments de l'acceptabilité sociale comme la protection de la vie privée et le fardeau des répondants, mais le présent article n'y est pas consacré.

Mots-clés : Approche déductive, approche inductive, inférence statistique, cadre théorique.

1. Contexte

Le monde est en constante évolution et, comme les organismes nationaux de statistique (ONS) ont pour mandat de mesurer la société, ils doivent s'adapter. Ces dernières années, c'est non seulement le paysage socioéconomique qui se transforme, mais aussi le paysage des données. En fait, la transformation du paysage des données est encore plus marquée. Aujourd'hui, les données sont présentes partout : dans les lieux de travail, sur le marché, dans l'air (capteurs et images satellites), à la maison (ordinateurs, systèmes de sécurité, appareils, etc.), sur notre corps, voire dans notre corps (montres intelligentes, stimulateurs cardiaques, ADN, etc.). En bref, pour être pertinents, les systèmes statistiques et les méthodes s'y rattachant doivent être constamment repensés et adaptés afin que la société dispose des meilleurs renseignements possible pour prendre des décisions éclairées. Le paradigme statistique d'il y a quelques décennies ne peut être le même aujourd'hui.

1.1 Contexte historique

¹ Le présent article expose des points de vue théoriques qui pourraient évoluer avant d'être mis en œuvre dans un programme statistique.

² Eric Rancourt, directeur général de la Méthodologie, Statistique Canada. 100, promenade Tunney's Pasture, Ottawa (Ontario), Canada K1A 0T6. Eric.rancourt@canada.ca.

Depuis le début de la production des statistiques officielles, les statisticiens qui les produisent utilisent tant des méthodes de collecte primaire (soit des mécanismes de collecte que l'ONS contrôle, principalement des enquêtes) que des méthodes de collecte secondaire (soit l'obtention de données recueillies par une autre entité, p. ex. des données administratives). Avant la Seconde Guerre mondiale, les tentatives de mettre en place une capacité d'inférence statistique dans le contexte des statistiques officielles consistaient essentiellement à tenter de la façon la plus disciplinée possible de recueillir des données auprès de toutes les unités au moyen d'un recensement ou en obtenant des fichiers de données transactionnelles (p. ex. les statistiques de l'état civil) d'entités connexes. Parfois, des pseudo-recensements ont été effectués lorsque les unités les plus grandes d'une population couvraient un pourcentage jugé suffisamment élevé (p. ex. dans le cas des revenus commerciaux des manufactures).

En 1934, un article de Jersey Neyman faisant autorité (Neyman, 1934) a jeté les bases d'une méthode d'inférence statistique valide qu'on appelle aujourd'hui la théorie de l'échantillonnage. Cette méthode a rapidement proliféré dans le monde entier après la Seconde Guerre mondiale. Au Canada, on a commencé à utiliser l'échantillonnage dans le cadre de l'Enquête sur la population active en 1945 et, tôt ou tard, presque tous les programmes statistiques étaient basés sur des enquêtes. C'est essentiellement la situation qui a prévalu jusqu'à maintenant. Accroître l'utilisation des données administratives dans les programmes statistiques était une grande priorité par le passé et récemment (Statistique Canada, 1984; 2012; 2018).

1.2 Contexte actuel

De nos jours, les données sont omniprésentes. Dans la foulée de la révolution des données (voir par exemple Organisation des Nations Unies, 2014), le contexte des statistiques officielles et des programmes statistiques a complètement changé. L'appétit pour les données a énormément augmenté, et la variété des données produites (qu'elles soient accessibles ou non par un tiers) a décuplé. Tandis que les décideurs, les entreprises et les gens exigent de plus en plus de données actuelles et gratuites, on s'attend à ce qu'il faille alléger davantage le fardeau qui incombe aux citoyens de fournir des données, et à ce que les données soient bien protégées. Il est donc nécessaire de repenser l'approche actuelle en matière de production des statistiques officielles. La théorie de l'échantillonnage demeurera-t-elle appropriée et utile? Tout à fait. Mais la théorie de l'échantillonnage suffit-elle pour répondre aux nouveaux besoins? Pas du tout. Comment donc la théorie permettra-t-elle de répondre aux nouveaux besoins? Selon toute vraisemblance, il s'agira d'élargir des volets existants de la théorie et/ou de développer de nouvelles théories.

Récemment (Arora, 2018), Statistique Canada a entrepris un processus de modernisation qui vise effectivement à mieux positionner l'organisme national de statistique afin qu'il puisse répondre aux nouvelles demandes statistiques dans le contexte du monde moderne des données. La stratégie repose sur cinq piliers : la prestation de services qui est axée sur l'utilisateur; des méthodes de pointe et l'intégration des données; le renforcement de la capacité statistique et le leadership; le partage et la collaboration; une main-d'œuvre moderne et un milieu de travail flexible.

Quant à la stratégie globale devant appuyer les programmes statistiques, celle qui consiste à utiliser et/ou à recueillir les renseignements en misant sur les données administratives d'abord est considérée comme le paradigme statistique à privilégier.

2. Un paradigme des données administratives d'abord : Qu'est-ce que cela veut dire?

Dans cette section, nous définissons le concept du paradigme des données administratives d'abord et nous présentons les concepts et les termes nécessaires pour donner un sens à cette approche.

Enquête : Une enquête est un outil conçu par l'ONS pour réaliser des activités de collecte primaire visant à obtenir des renseignements sur des concepts clairement définis. On suppose ici que l'enquête serait conçue, menée et utilisée de façon à permettre une inférence statistique valide selon la théorie de l'échantillonnage. De ce fait, les enquêtes sont normalement de nature probabiliste. D'autres outils de collecte primaire pourraient être un recensement, des groupes de discussion, des applications mobiles, des appareils électroniques, etc.

Données administratives : Données utilisées pour administrer un programme ou une organisation. Ce terme est utilisé au sens le plus large, c'est-à-dire qu'il pourrait s'agir de données provenant de ministères, mais aussi de toute autre organisation publique ou privée. Comme ces données ne sont pas produites à des fins d'inférence (mais plutôt à des

fins de gestion d'un programme, de prestation de services ou de réglementation), leur conception est presque toujours non probabiliste. Aux fins de la discussion conceptuelle d'un cadre de données, nous supposons que les données administratives sont utilisées à des fins secondaires, par opposition aux données obtenues au moyen d'activités de collecte primaire.

Paradigme des données d'enquête seulement : Cadre dans lequel l'inférence appliquée à une population relativement à des concepts est effectuée exclusivement à partir d'une enquête.

Paradigme des données d'enquête d'abord : Cadre dans lequel l'inférence appliquée à une population relativement à des concepts est effectuée à partir d'une enquête, appuyée de diverses façons par des données administratives. Les façons dont les données administratives peuvent améliorer l'enquête sont nombreuses (voir par exemple Brackstone, 1987 ou Beaumont, 2018). Du point de vue du plan, ce paradigme suppose d'envisager d'abord une enquête, puis d'utiliser des données administratives pour l'améliorer. Quant à l'inférence, elle pourrait être fondée sur le plan, appuyé ou non d'un modèle, ou bien elle pourrait être fondée sur un modèle.

Paradigme des données administratives seulement : Cadre dans lequel l'inférence appliquée à une population relativement à des concepts est effectuée à partir de données administratives seulement. Il se pourrait que plusieurs ensembles de données administratives soient requis.

Paradigme des données administratives d'abord : Cadre dans lequel l'inférence appliquée à une population relativement à des concepts est effectuée à partir de données administratives, appuyées de diverses façons par des enquêtes (ou plus généralement par des activités de collecte primaire). Autrement dit, avant de mener des enquêtes, on fait tous les efforts raisonnables pour tirer le meilleur parti possible des renseignements préexistants qui peuvent être mis à la disposition de l'ONS ou qui lui sont accessibles. Du point de vue du plan, ce paradigme suppose d'envisager d'abord des données administratives, puis de mener ou non une enquête pour compléter la portée des données administratives ou pour en évaluer la qualité. Quant à l'inférence, il est probable qu'elle soit fondée sur un modèle, mais le choix de ce paradigme ne sous-entend pas nécessairement un choix séquentiel de données administratives suivies de données d'enquête. Comme on peut élaborer le plan en envisageant conjointement les deux types de données, l'inférence pourrait aussi être fondée sur le plan dans certains cas. Les statistiques trimestrielles des entreprises de la Nouvelle-Zélande sont un exemple de ce paradigme (Liken et coll., 2018).

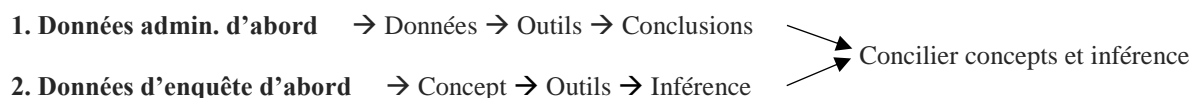
3. Certains défis et possibilités associés à un paradigme des données administratives d'abord

Dans le contexte de la production des statistiques officielles, le travail commence normalement lorsque les décideurs et/ou la société expriment un besoin en données à l'ONS. Citro (2014) adopte un tel point de vue sur le rôle du programme statistique de l'ONS. Dans le paradigme des données d'enquête d'abord, c'est par définition l'approche suivie. Le paradigme des données administratives d'abord suit une approche plus générale et offre de nouvelles possibilités en ce sens qu'il devient maintenant possible d'explorer les données existantes pour déterminer lesquelles pourraient être consultées ou utilisées. Toutefois, comme l'utilisation de données est censée découler d'un but défini, il est difficile d'élargir les possibilités entourant ces données tout en respectant le but en question.

Les données administratives sont définies au moyen de concepts appropriés ou propres au programme pour lequel elles ont été produites. Ces concepts pourraient 1) être différents des concepts ciblés dans le programme statistique et 2) être différents des concepts utilisés dans les enquêtes. Le défi consiste alors à concilier ces divergences. À mesure que la variété des sources de données augmente, la probabilité d'être mis en présence de concepts différents augmente aussi. On peut sans doute recourir à la modélisation ou à d'autres techniques pour tenir compte des différences, mais il faudra à tout le moins reconnaître l'existence des divergences et, idéalement, les mesurer.

Il n'existe peut-être pas de recueil complet de tous les fichiers de données qui ont été produits dans un pays. Par conséquent, la conception d'un programme statistique selon le paradigme des données administratives d'abord présuppose de connaître l'existence des éventuelles sources de données qui pourraient être liées à un programme statistique donné. De plus, même si l'on est parfaitement au courant des fonds de données qui existent, il faut ensuite déterminer ou élaborer les protocoles d'accès aux données.

Une fois que les données sont accessibles, il s'agit d'en évaluer la structure, la nature et la qualité. Comme les données administratives sont loin d'être parfaites, on est confronté à la difficulté de comprendre le désordre présent dans les données (voir par exemple Baker et coll., 2013 au sujet des limites des données administratives) et de mettre en place des mécanismes pour composer avec cette situation. Meng (2018) a donné un exemple frappant de la façon dont des données non probabilistes de qualité apparemment élevée pourraient en fait mener à des conclusions erronées si l'on n'est pas prudent. Il est donc important de définir et d'adopter un cadre de qualité totale qui informera convenablement les gestionnaires des programmes statistiques et les utilisateurs des données qu'ils produisent. Mais, d'une façon plus complète, il faut veiller à ce que les estimations soient produites selon un cadre théorique explicite qui permet une inférence statistique valide.



Dans un paradigme des données d'enquête d'abord, l'inférence est déductive en ce sens que, dans un premier temps, un concept à mesurer est défini, puis des outils de mesure sont élaborés. L'inférence est appliquée à la population à partir des données résultantes selon une théorie d'échantillonnage claire. Dans le contexte du paradigme des données administratives d'abord, il arrive parfois que le même type d'inférence soit souhaité; la combinaison de données administratives et de données d'enquête permettra l'approche déductive. Toutefois, certaines personnes souhaitent commencer par des données administratives, élaborer des outils et essayer de tirer des conclusions selon une approche inductive. Ce n'est pas une erreur en soi, mais la combinaison de cette approche avec une approche traditionnelle déductive fondée sur une enquête pose deux défis importants. Premièrement, si les données ou les estimations qui en résultent sont produites au moyen des deux approches, il faut veiller à donner des indications et des explications appropriées aux utilisateurs afin qu'ils comprennent le contexte de l'inférence et puissent tirer les bonnes conclusions. Deuxièmement, il est difficile d'intégrer les deux approches de manière à maximiser le pouvoir conjoint des données d'enquête et des données administratives. À cette fin, il faut un cadre théorique plus complet et/ou plus général.

4. Progrès liés à l'élargissement et à la définition du cadre théorique

Au cours des dernières décennies, il s'est accompli beaucoup de progrès vers une plus grande utilisation des données administratives dans le cadre des programmes statistiques. Depuis les débuts de la théorie de l'échantillonnage, le contexte s'est élargi. En outre, les méthodes d'estimation de la qualité des enquêtes et de la qualité des données en général se sont multipliées, et la communauté statistique a fait de grands pas vers un cadre de qualité totale.

En fait d'avancées théoriques, on pourrait notamment penser à Särndal, Swanson et Wretman (1992), qui ont proposé l'approche assistée par un modèle comme stratégie globale relative à l'utilisation des données administratives. Zhang (2012) fait avancer la théorie de l'intégration des données dans le contexte des registres, tandis que Lohr et Raghunathan (2017) traitent de la façon de combiner des enquêtes avec d'autres sources de données. En examinant le problème de l'utilisation des données administratives (ou de renseignements auxiliaires) au moyen de modèles de données régionales, Rao et Molina (2015) présentent un compte rendu complet des méthodes d'estimation sur petits domaines. Ensuite, Beaumont (2018) présente un examen rigoureux des pratiques existantes concernant la combinaison d'échantillons probabilistes et non probabilistes.

Un important travail a été abattu en vue de l'élaboration d'un cadre de qualité, et Beamer (2016) décrit la théorie et la pratique entourant un paradigme de l'erreur d'enquête totale.

Plusieurs autres auteurs ont contribué à l'avancement de l'utilisation des données administratives à des fins d'inférence valide (p. ex. Mercer et coll. (2017)) et pour l'estimation de plusieurs dimensions de la qualité (p. ex. Bosa et coll. (2018) pour la variance de non-réponse; Pankowska et coll. (2018) pour l'erreur de mesure), mais l'objectif ici est simplement de présenter le contexte général menant à une tentative de préciser les éventuels éléments d'un cadre théorique du paradigme des données administratives d'abord.

5. Tentative de préciser les éléments d'un cadre théorique

Dans la présente section, nous exposons succinctement les éléments qui pourraient ensemble former le squelette de ce qu'on pourrait appeler une « théorie sociostatistique » aux fins de l'inférence selon un paradigme des données administratives d'abord. Il s'agit d'une modeste tentative d'unifier les paradigmes statistiques des données d'enquête d'abord et des données administratives d'abord avec des activités de gérance des données et des éléments tels que l'accès aux données et les registres statistiques. De plus, le cadre pourrait également servir à contextualiser la mesure de la qualité.

Si les neuf hypothèses suivantes étaient jugées valides dans un contexte particulier, alors on serait en mesure de produire des estimations « parfaites ». En pratique, il y aura inévitablement certains écarts par rapport à ces hypothèses. Toute inférence doit alors tenir compte de la mesure de l'écart par rapport à ces hypothèses. Il en découle une structure qui est implicitement conforme à celle de Biemer (2016) pour l'évaluation de la qualité des données, et Berka et coll. (2011) fournit l'exemple autrichien d'un cadre de qualité pour les données administratives.

Les éléments d'une théorie sociostatistique

Hypothèse 1 : Les données de toutes les unités existent sous forme numérique. En supposant que les gens et les entreprises trouveront une raison impérieuse pour au moins un des programmes qui existent à la fois dans les secteurs privé et public, ils s'y conformeront en s'inspirant peut-être de la théorie de l'utilité de von Neumann et Morgenstern (1953). Sinon, les éléments de la population cible seraient captés par un quelconque capteur, satellite ou autre appareil, y compris les médias sociaux.

Hypothèse 2 : L'ONS a à sa disposition des bases de sondage ou des registres complets. Que la population soit constituée d'entreprises, de personnes, d'exploitations agricoles, d'animaux, de bâtiments, etc., le fait d'avoir une liste complète permet d'assurer une couverture complète.

Hypothèse 3 : Soit que l'ONS a accès à l'ensemble des données au moyen d'un mécanisme qu'il contrôle (p. ex. un échantillonnage probabiliste), soit qu'il est en mesure de recueillir les renseignements nécessaires à une inférence statistique valide et qu'il est autorisé à le faire. Il peut s'agir d'un seul fichier ou de plusieurs fichiers ou sources.

Hypothèse 4 : Toutes les unités peuvent être appariées sans erreur. Il est ainsi possible d'éviter les enregistrements en double tout en ayant une parfaite connaissance de la couverture.

Hypothèse 5 : Les concepts liés aux données correspondent aux concepts cibles, ce qui signifie qu'il n'existe pas de divergence conceptuelle entre les données recueillies et l'objectif cible, ni entre les sources de données, ni entre les données administratives et les données d'enquête.

Hypothèse 6 : Il n'y a pas d'erreur de non-réponse.

Hypothèse 7 : Il n'y a pas d'erreur de mesure.

Hypothèse 8 : Il existe des liens étroits entre les variables. Autrement dit, des modèles peuvent être élaborés à des fins d'estimation ou d'analyse.

Hypothèse 9 : Les erreurs types sont petites, et/ou la quantité de renseignements est suffisamment grande pour permettre une inférence précise.

On pourrait tirer un certain nombre de corollaires de ces hypothèses. En voici quelques-uns :

Corollaire 1 : Des registres devraient être élaborés et tenus à jour en temps réel (ou assez fréquemment pour qu'ils puissent répondre aux besoins des programmes statistiques à mesure qu'ils se présentent).

Corollaire 2 : L'ONS devrait viser à obtenir l'accès à toutes les données.

Corollaire 3 : Une fois qu'un paradigme des données administratives d'abord est pleinement mis en œuvre, la collecte primaire devrait servir à évaluer les incertitudes et les écarts par rapport à toutes les hypothèses susmentionnées.

Corollaire 4 : L'amélioration de la qualité de l'appariement et la mesure des erreurs s'y rattachant sont essentielles à l'inférence.

En fait, les efforts pour s'écarter le moins possible de chacune des neuf hypothèses et la mesure de l'ampleur des écarts pourraient tous constituer un corollaire.

6. Conclusion

Nous avons expliqué ce qu'on entend par un paradigme des données administratives d'abord pour les programmes statistiques. Ce paradigme offre de nombreuses possibilités de production de renseignements et il est appuyé par un certain nombre de méthodes statistiques comportant une capacité d'inférence. Nous avons présenté les défis et les possibilités ainsi que les éléments d'un cadre théorique, mais ce cadre théorique pourrait être élargi considérablement de façon à inclure un plus vaste éventail de possibilités. Par exemple, on pourrait envisager d'utiliser des données administratives pour établir des antécédents, puis de concevoir des enquêtes de façon qu'elles soient optimales pour l'estimation du maximum de vraisemblance compte tenu de ces antécédents. Les premiers travaux de Rao et Ghangurde (1972) peuvent s'y rattacher. À l'inverse, il serait peut-être possible d'estimer les antécédents à partir de petites enquêtes rapides complétées par des données administratives exhaustives. Il faudrait examiner attentivement les questions du biais de sélection et de l'incidence de la taille des variances. De même, des enquêtes pourraient être conçues de manière à combler les lacunes statistiques présentes dans les données administratives et être élaborées de façon optimale pour la mesure d'un ou de plusieurs aspects de la qualité des données administratives. Il se pourrait aussi qu'une théorie complète n'existe pas ou qu'elle ne soit pas probabiliste. Dans le présent article, nous avons (presque) seulement examiné des approches quantitatives. Par ailleurs, il serait possible et il y aurait lieu d'aller plus loin et de tenter de combiner l'utilisation de données qualitatives avec des données quantitatives au moyen d'approches fondées sur des méthodes mixtes (voir par exemple Poth, 2018).

Remerciements

L'auteur tient à remercier Martin Beaulieu, Jean-François Beaumont, Linda Howatson-Leo, Andrea-Leigh MacMillan, J.N.K. Rao et Siu-Ming Tam de leurs commentaires et de leurs réflexions concernant le présent article.

Références

ARORA, A. (2018). *Modernisation du système statistique national – Consultations avec les partenaires*, produit n° 89200003 au catalogue. Statistique Canada. ISBN 978-0-660-31581-2.

BAKER, R., J.M. BRICK, N.A. BATES, M. BATTAGLIA, M.P. COUPER, J.A. DEVER, K. GILE, et R. TOURANGEAU. (2013). « Summary report of the AAPOR task force on non-probability sampling », *Journal of Survey Statistics and Methodology*, 1, p. 90 à 143.

BEAUMONT, J.F. (2018). *Les enquêtes probabilistes sont-elles vouées à disparaître pour la production de statistiques officielles?*, présentation au 10^e Colloque francophone sur les sondages, Lyon, France.

BERKA, C., S. HUMER, M. LENK, M. MOSER, H. RECHTA, E. SCHWERER. (2011). « A quality framework for statistics based on administrative data sources using the example of the Austrian census 2011 », *Austrian Journal of Statistics*, volume 39 (2010), numéro 4, p. 299 à 308.

BIEMER, P.P. (2016). « Total survey error paradigm: Theory and practice », publié dans *The SAGE Handbook of Survey Methodology* sous la direction de C. Wolf, D. Joye, T.W. Smith et F. Yang-chih, Londres, Sage.

BOSA, K., S. GODBOUT, F. MILLS, F. PICARD. (2018). « Comment décomposer la variance due à la non-réponse : une méthode fondée sur l'erreur d'enquête totale », *Techniques d'enquête*, 44, p. 319 à 337.

- BRACKSTONE, G. J. (1987). « Utilisation des dossiers administratifs à des fins statistiques », *Techniques d'enquête*, 13, p. 35 à 51.
- CITRO, C.F. (2014). « Des modes multiples pour les enquêtes à des sources de données multiples pour les estimations », *Techniques d'enquête*, 40, p. 151 à 181.
- LIKEN, C., M. PAGE, et J. STUART. (2018). « Realisation of 'administrative data first' in quarterly business statistics' », *Statistical Journal of the IAOS*, 34, p. 567 à 576. DOI 10.3233/SJI-180456.
- LOHR, S. et T.E. RAGHUNATHAN. (2017). « Combining survey data with other data sources », *Statistical Science*, 32, p. 293 à 312.
- MENG, X.-L. (2018). « Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election », *Annals of Applied Statistics*.
- MERCER, A.W., F. KREUTER, S. KEETER, et E.A. STUART. (2017). « Theory and practice in nonprobability surveys: Parallels between causal inference and survey inference », *Public Opinion Quarterly*, 81, p. 250 à 271.
- NEYMAN, J. (1934). « On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection », *Journal of the Royal Statistical Society*, 97, p. 558 à 625.
- ORGANISATION DES NATIONS UNIES. (2014). *A World that Counts – Mobilising the data revolution for sustainable development*, rapport préparé à la demande du Secrétaire général des Nations Unies par le Groupe consultatif d'experts indépendants sur la révolution des données pour le développement durable.
- PANKOWSKA, P., B. BAKKER, D.L. OBERSKI, D. PAVLOPOULOS. (2018). « Reconciliation of inconsistent data sources by correction for measurement error: the feasibility of parameter re-use », *Statistical Journal of the IAOS*, volume 34, numéro 3.
- POTH, C. N. (2018). *Innovation in mixed methods research. A practical guide to integrative thinking with complexity*, Londres, Sage.
- RAO, J.N.K., et I. MOLINA. (2015). *Small area estimation*, Second Edition, Wiley, Hoboken, NJ.
- RAO, J.N.K., et P.D. GHANGURDE. (1972). « Bayesian optimization in sampling finite populations », *Journal of the American Statistical Society*, vol. 67. p. 439 à 443.
- SÄRNDAL, C.-E., B. SWENSSON, et J. WRETMAN. (1992). *Model-assisted survey sampling*, Springer-Verlag, New York.
- STATISTIQUE CANADA. (1984). *Rapport annuel 1983-1984*, 1-0000-502, produit n° 11-201 au catalogue, Statistique Canada, Ottawa.
- STATISTIQUE CANADA. (2012). *Statistique Canada – Rapport sur les plans et les priorités 2012-2013*.
- STATISTIQUE CANADA. (2018). *Statistique Canada – Plan ministériel 2018-2019*, produit n° 11-635-X au catalogue. ISSN 2371-7718.
- VON NEUMANN, J., et O. MORGENSTERN. (1953). *Theory of games and economic behavior*, 3rd Ed., Princeton University Press.