

Utilisation des banques de données médico-administratives : forces et défis

Marc Corbière

Volume 43, Number 2, Fall 2018

URI: <https://id.erudit.org/iderudit/1058606ar>

DOI: <https://doi.org/10.7202/1058606ar>

[See table of contents](#)

Publisher(s)

Revue Santé mentale au Québec

ISSN

1708-3923 (digital)

[Explore this journal](#)

Cite this document

Corbière, M. (2018). Utilisation des banques de données médico-administratives : forces et défis. *Santé mentale au Québec*, 43(2), 9–13. <https://doi.org/10.7202/1058606ar>

ÉDITORIAL

Utilisation des banques de données médico-administratives : forces et défis

Marc Corbière, rédacteur en chef (Thématiques)

Dans le cadre de ce numéro thématique, Christophe Huÿnh, Alain Lesage et Sylvanne Daniels nous proposent d'aborder l'exploitation des grandes banques de données médico-administratives sous la facture de huit articles. Tout en soulignant les défis rencontrés lors de l'analyse des données des banques médico-administratives, les auteurs nous sensibilisent au portrait de divers troubles de santé mentale au Québec. Les données de banques médico-administratives sont souvent appelées « données secondaires », recueillies pour des fins qui ne sont pas relatives à la recherche comme telle, mais qui peuvent être analysées de façon rétrospective par les chercheurs¹. Ces banques de données peuvent, en ce sens, faciliter le développement de questions et hypothèses de recherche à posteriori dont la vérification serait difficile, voire impossible, par l'usage de méthodes de recherche conventionnelles¹. En effet, elles sont souvent conçues pour des cibles pharmaceutiques et d'assurance et, par conséquent, les chercheurs s'évertuent à utiliser des algorithmes pour traiter au mieux les données recueillies.

Les banques de données médico-administratives représentent une source d'information exhaustive, soutenue par des tailles d'échantillons conséquentes, pouvant venir en appui à des résultats d'études épidémiologiques et expérimentales. Qui plus est, de nature longitudinale, ces banques de données systématiques offrent une surveillance sans égale, étayée par un recul temporel important (plusieurs décennies parfois) pour mieux saisir l'évolution des maladies dans un territoire circonscrit ainsi que la trajectoire des patients dans l'utilisation

des services en santé au Québec et ailleurs. Ces données peuvent donc fournir aux chercheurs et cliniciens des pistes utiles pour les interventions futures à mettre en place, en dégagant des profils d'individus et leur santé comme nous pourrions le constater par exemple dans ce numéro thématique avec la proposition de stratégies de détection et de prévention du suicide. Par ailleurs, elles offrent l'avantage de coupler des données cliniques à des données géographiques notamment. Certaines banques de données sont parfois le résultat de jumelages de plusieurs fichiers de nature médico-administrative comme c'est le cas du Système intégré de surveillance des maladies chroniques du Québec (SISMACQ) qui représente une source inestimable de données sur les maladies chroniques telles que certaines maladies mentales². En bref, les banques de données médico-administratives ouvrent de multiples champs d'investigation pour les chercheurs.

Nonobstant les divers apports des banques de données médico-administratives, les défis de leur utilisation sont de taille, et ce, à plusieurs niveaux. Tout d'abord, lorsqu'une équipe de recherche souhaite entreprendre l'analyse des données provenant de ces banques, un investissement substantiel en temps est souvent requis pour obtenir l'approbation de leur utilisation. De longs délais d'approbation peuvent occasionner des hiatus dans le déroulement de la recherche sachant que les objectifs de recherche et les financements des projets sont circonscrits dans le temps³. D'autres encore déplorent le manque de clarté à propos des variables disponibles dans les banques de données médico-administratives, contraignant ainsi la formulation et la vérification d'hypothèses de recherche. Des ajustements ou une bonne planification de l'utilisation de ces données s'avèrent donc nécessaires.

Un autre défi s'illustre par la création d'indicateurs précis lorsqu'un thème d'analyse est choisi, notamment lorsque des diagnostics psychiatriques ou groupes de diagnostics particuliers sont ciblés. Les codes diagnostics sont parfois difficiles à interpréter et certains auteurs préconisent de rendre disponibles les syntaxes qu'ils ont utilisées pour définir de façon robuste la population sous étude et ainsi, répliquer l'étude au besoin⁴. Dans ce cas de figure, la notion d'objectivité reste complexe, car l'établissement d'un diagnostic psychiatrique est réalisé parfois dans des situations d'urgence où le médecin n'a pas toute la latitude et le temps requis pour considérer l'ensemble des critères diagnostiques⁵. C'est pourquoi, par ce manque de standardisation, chercheurs et cliniciens sont invités à vérifier scrupuleusement la validité des banques de données médico-administratives pour des fins pra-

tiques de nature clinique¹. Un nouveau défi s'ajoute lorsqu'il s'agit d'entreprendre des comparaisons avec les diagnostics du *Diagnostic and Statistical Manual of Mental Disorders* (DSM) source incontournable des cliniciens nord-américains avec ceux de la Classification internationale des maladies (CIM); des codes sensiblement différents selon la problématique de santé. De nouveaux défis qui exigent chez les utilisateurs-chercheurs une double lecture, celle appartenant au contexte dans lequel ils œuvrent et celle particulière des banques de données médico-administratives; une navigation des compétences sur le plan terminologique s'impose donc chez les utilisateurs des banques de données administratives en Amérique du Nord, notamment lorsqu'il s'agit de santé mentale.

D'autres défis sont relatifs au jumelage des banques de données comme la RAMQ avec d'autres données cliniques des patients ou des résultats d'examen de laboratoire. Le couplage de données (*data linkage*) de divers fichiers peut être effectué par le nom de la personne, sa date de naissance, son sexe, son adresse, son numéro d'assurance sociale ou encore par d'autres identifiants^{2, 6}. Puisque les données administratives n'ont pas été recueillies pour des objectifs de recherche, le public en général pourrait donc avoir des préoccupations quant à l'utilisation par les chercheurs de ces informations provenant de diverses sources⁷. Toutefois, les procédures de confidentialité de l'utilisation des données sont extrêmement strictes. Les chercheurs n'ont accès qu'à un ensemble restreint de données « anonymisées » et doivent faire la démonstration avant même l'analyse et la publication des résultats que les individus ne peuvent pas être identifiés (p. ex. croisement de variables) et ainsi respecter les règles de confidentialité en vigueur au Québec et au Canada, et ce, quels que soient les pays.

Par ailleurs, une des controverses liées aux banques de données médico-administratives est l'avantage d'un recueil de données plus objectif et non autorapporté par les personnes directement concernées. Ces deux types de données devraient pourtant être considérés comme complémentaires plutôt que mutuellement exclusifs. L'utilisation de ces deux types de données est judicieuse, mais la plupart du temps coûteuse. En l'occurrence, un plan solide de la pérennité des recueils de données est essentiel. La Banque Signature du Centre de recherche de l'Institut universitaire en santé mentale de Montréal, une base de données biologique, psychosociale et clinique de personnes atteintes de troubles de santé mentale (<http://www.iusmm.ca/recherche/signature.html>) est un bon exemple de cette possible opérationnalisation.

La deuxième décennie de ce nouveau siècle s'illustre par l'explosion de nouvelles formes de données et de leur exploitation. Bien que l'utilisation des banques de données médico-administratives n'a pas été centrale dans les discussions concernant les données massives ou *big data*, certains auteurs précisent que ces premières correspondent à la définition des *big data* notamment en termes de Volume (*Volume*), Variété (*Variety*), Vitesse (*Velocity*), auxquels s'ajoutent Valeur (*Value*), Variabilité (*Variability*) et Vérité (*Veracity*) pour un total de 6V⁷⁻⁹. Autrement dit, à l'instar des *big data*, les banques de données médico-administratives se caractérisent par de grands bassins d'individus et une taille volumineuse d'informations par individu, un recueil rapide de divers types de données provenant de multiples sources, même si ces dernières ne sont pas nécessairement recueillies en temps réel, des données qui ne sont pas statiques et qui peuvent évoluer à travers le temps en vue d'une meilleure adaptation aux changements conceptuels, technologiques et sociaux. La notion de « vérité » est probablement la plus complexe à respecter au sens où les données n'ont pas été recueillies dans le cadre d'objectifs de recherche⁹. Il demeure donc important de bien connaître les facteurs influençant le processus de recueil et la précision des données; des éléments incontournables qui sont soulevés par les auteurs non seulement pour l'utilisation d'algorithmes au sein de *big data*, mais aussi lors de l'utilisation des banques de données médico-administratives. *Big data* et banques de données médico-administratives sont d'ailleurs à la croisée des réflexions chez les chercheurs et cliniciens à savoir : Comment utiliser de façon judicieuse, précautionneuse et rigoureuse les données qui sont disponibles en vue d'améliorer les conditions de santé des personnes en besoin ?

Le comité éditorial de *Santé mentale au Québec* invite donc les lecteurs, à travers ce numéro thématique, à mieux saisir l'ampleur des résultats et retombées issus de ces banques de données médico-administratives tout en étant attentifs aux limites inhérentes et au grand potentiel pour les futures interventions cliniques et sociales dans le domaine de la santé mentale.

RÉFÉRENCES

- 1 Johnson, E. K. et Nelson, C. P. (2013). Utility and Pitfalls in the Use of Administrative Databases for Outcomes Assessment. *The Journal of Urology*, 190(1), 17-18. doi:10.1016/j.juro.2013.04.048

- ² Blais, C., Jean, S., Sirois, C., Rochette, L., Plante, C., Larocque, I., ... Emond, V. (2014). Quebec Integrated Chronic Disease Surveillance System (QICDSS), an innovative approach. *Chronic Diseases and Injuries in Canada*, 34(4), 10.
- ³ Harron, K., Dibben, C., Boyd, J., Hjern, A., Azimae, M., Barreto, M. L. et Goldstein, H. (2017). Challenges in administrative data linkage for research. *Big Data & Society*, 4(2), 2053951717745678. doi:10.1177/2053951717745678
- ⁴ Playford, C. J., Gayle, V., Connelly, R. et Gray, A. J. (2016). Administrative social science data: The challenge of reproducible research. *Big Data & Society*, 3(2), 2053951716684143. doi:10.1177/2053951716684143
- ⁵ Davis, K. A. S., Sudlow, C. L. M. et Hotopf, M. (2016). Can mental health diagnoses in administrative data be used for research? A systematic review of the accuracy of routinely collected diagnoses. *BMC Psychiatry*, 16(1), 263. doi:10.1186/s12888-016-0963-x
- ⁶ Jutte, D. P., Roos, L. L. et Brownell, M. D. (2011). Administrative Record Linkage as a Tool for Public Health Research. *Annual Review of Public Health*, 32(1), 91-108. doi:10.1146/annurev-publhealth-031210-100700
- ⁷ Connelly, R., Playford, C. J., Gayle, V. et Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, 59, 1-12. doi:10.1016/j.ssresearch.2016.04.015
- ⁸ Ehrenstein, V., Nielsen, H., Pedersen, A. B., Johnsen, S. P. et Pedersen, L. (2017). Clinical epidemiology in the era of big data: new opportunities, familiar challenges. *Clinical Epidemiology*, 9, 245-250. doi:10.2147/CLEP.S129779
- ⁹ Stewart, R. et Davis, K. (2016). 'Big data' in mental health research: current status and emerging possibilities. *Social Psychiatry and Psychiatric Epidemiology*, 51(8), 1055-1072. doi:10.1007/s00127-016-1266-8