

## **UN ENVIRONNEMENT RICHE EN INFORMATION : LA QUALITÉ DES DONNÉES DES SYSTÈMES D'APPARIEMENT DE DOSSIERS AU CANADA**

Leslie L. Roos, Ruth-Ann Soodeen et  
Laurel Jebamani<sup>1</sup>

### **RÉSUMÉ**

Des programmes de recherche fondés sur l'utilisation de données administratives longitudinales représentatives de la population et sur les techniques de couplage d'enregistrements sont en cours à Oxford, en Écosse, à Rochester (Mayo Clinic), en Australie-Occidentale et au Canada. Ces programmes représentent des progrès importants pour la recherche, méthodologique et autre, sur la santé et les soins de santé.

Nous résumons ici des études publiées au Canada sur la qualité des données concernant les registres, les dossiers de sortie des hôpitaux, les médicaments délivrés sur ordonnance et les demandes de remboursement des médecins. Des suggestions d'amélioration des registres permettent de faciliter le couplage et d'étendre la recherche à l'épidémiologie sociale. Nous faisons état de nouvelles tendances en matière de dépistage des cas et d'évaluation de l'état de santé à l'aide de données administratives, ainsi que des besoins de chaque province en matière de recherche sur la qualité des données.

### **1. INTRODUCTION**

En science, les « points d'entrée » sont de nouvelles techniques ou de nouveaux concepts qui ouvrent des perspectives et permettent de réaliser des progrès dans un domaine (Weiner, 1999). L'étude de la santé et des soins de santé bénéficie de deux « points d'entrée » : les données administratives longitudinales concernant des populations définies et les techniques de couplage d'enregistrements, qui ont joué un rôle essentiel dans l'établissement d'environnements riches en information (Roos et Roos, 2001). Plusieurs provinces canadiennes ont mis au point les éléments nécessaires – registres, fichiers de demandes de remboursement, etc. – pour produire un environnement de recherche riche en information en se fondant sur des systèmes d'appariement de dossiers représentatifs de la population.

Tout système d'appariement de dossiers dépend de la qualité de ses composantes. La comparaison de dossiers avec des données administratives, des enquêtes et des renseignements cliniques en vue d'assurer la qualité est facilitée par le couplage d'enregistrements et par les données longitudinales.

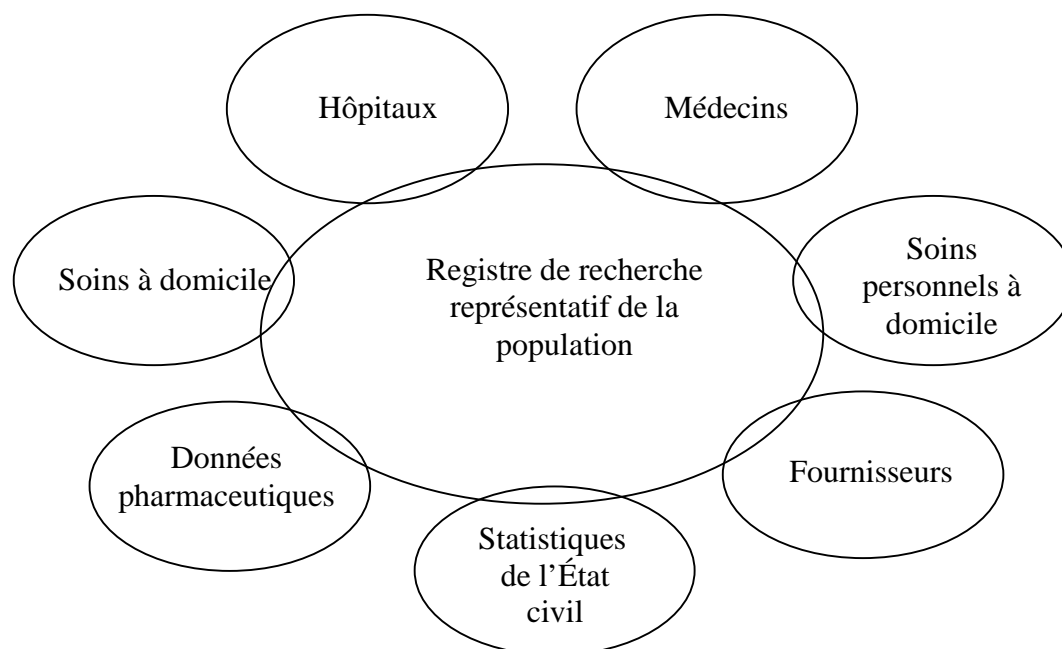
Dans la présente communication, nous passons en revue les plus importants ensembles de données mentionnés dans la figure 1 : registres, dossiers de sortie des hôpitaux, visites de médecins et médicaments délivrés sur ordonnance. Nous étudions deux méthodes d'établissement de la qualité des données :

- 1) La première méthode consiste à appairer deux ou plusieurs sources de renseignements dont on dispose déjà et à comparer l'étendue de leur concordance à l'égard d'éléments clés. Souvent, les deux sources sont tirées

---

<sup>1</sup> Manitoba Centre for Health Policy  
Département des sciences de la santé communautaire, Faculté de médecine  
Université du Manitoba, 4<sup>e</sup> étage, Centre de recherche John Buhler  
408 – 727 McDermot Avenue  
Winnipeg (Manitoba) R3E 3P5 Canada

Fig. 1 La base de données administratives idéale



de données administratives (habituellement les dossiers de sortie des hôpitaux et les demandes de remboursement des médecins). L'appariement permet également de mieux connaître le nombre d'enregistrements appariés convenablement; ce renseignement est important pour évaluer la qualité globale du système.

- 2) La deuxième méthode, habituellement fondée sur une seconde saisie de dossiers hospitaliers ou de dossiers de médecins, consacre des ressources considérables à mesurer l'exactitude des renseignements utilisés (Potvin et Champagne, 1986). Outre qu'on peut évaluer les erreurs de transcription, on peut balayer les enregistrements physiques pour chercher d'autres résultats de test ou d'autres entrées.

## 2. REGISTRE

Comme le montre la figure 1, un registre fondé sur une population définie est essentiel à un système d'appariement de dossiers; bon nombre d'études sont conçues pour utiliser un registre afin de retracer des personnes d'un dossier à l'autre. L'exhaustivité d'un registre est particulièrement essentielle à l'étude d'une population définie.

Presque toutes les analyses de la qualité des registres provinciaux sont transversales et utilisent des vérifications par appariement avec une autre source (données du recensement, statistiques de l'état civil ou enquêtes). Les articles publiés font état de dix-huit vérifications de ce genre. On compare habituellement les bases de données en fonction du nombre d'« individus » qu'on trouve dans les deux dossiers. On a déclaré des niveaux de concordance élevés. Toutefois, ces comparaisons ne constituent qu'un aspect de l'ensemble. Il est essentiel de préciser la mortalité individuelle et le lieu de résidence (à l'intérieur ou en dehors de la province) pour effectuer de nombreux types de recherche. Il importe de disposer de données exactes sur le lieu de résidence pour évaluer l'état de santé et l'utilisation des soins de santé à l'égard de différentes populations et pour effectuer des études à long terme sur des cohortes. Le plus souvent, on n'effectue pas ces vérifications de la mortalité et du lieu de résidence; une seule étude (menée au Manitoba) a utilisé la collecte de données primaires pour analyser la qualité des registres.

### 3. HÔPITAUX ET MÉDECINS

Plusieurs provinces ont mené des études par appariement portant sur la qualité des dossiers de sortie des hôpitaux et des dossiers de médecins, en utilisant souvent un grand nombre d'enregistrements, de méthodes et de diagnostics. On a signalé des taux de concordance élevés pour les renseignements démographiques et pour les méthodes; la concordance relative aux diagnostics est habituellement quelque peu inférieure. La méthode de l'appariement est polyvalente : les auteurs d'études hospitalières sur la qualité des données ont procédé à des appariements avec des données sur la santé mentale, les statistiques de l'état civil et des données financières, en plus des demandes de remboursement des médecins. Nous avons passé en revue neuf études par appariement fondées sur les dossiers de sortie des hôpitaux. On a également apparié les demandes de remboursement des médecins avec des données pharmaceutiques, des registres de maladies, des enquêtes et des enregistrements de laboratoire. Lors de cette étude, un problème s'est posé à quelques reprises : lorsque deux dossiers utilisent des schémas de codage un peu différents, le niveau d'agrégation des données – la mesure dans laquelle un code est hautement spécifique – devient un enjeu de ces études sur la qualité des données.

Quatre études portant sur des hôpitaux de la Saskatchewan montrent des résultats généralement semblables lorsqu'on peut comparer, dans un même document, des données obtenues par appariement et par une seconde saisie (Rawson et Malcolm, 1995a; Rawson et Malcolm, 1995b; Edouard et Rawson, 1996; Rawson et coll., 1997). Nous avons également analysé neuf études de données hospitalières obtenues par une seconde saisie. Huit études portant sur les demandes de remboursement des médecins se répartissaient à peu près également entre la méthode de l'appariement et celle de la seconde saisie des données. Les résultats concernant la qualité des renseignements sur la démographie, les méthodes et les diagnostics étaient généralement semblables à ceux des études portant sur les dossiers de sortie des hôpitaux.

Les possibilités de divergence dans les études par seconde saisie sont limitées; la fiabilité formelle se rapporte à ces écarts dans la transcription et le codage. Les comparaisons établies dans les études par appariement (fiabilité factuelle) montrent plus de possibilités de divergence. Des erreurs de transcription peuvent survenir dans le codage des dossiers de sortie et des demandes de remboursement des médecins; une modification à court terme de l'état de santé des malades ainsi que la mémoire des médecins pourraient avoir une incidence sur les diagnostics précis codés dans l'un ou l'autre ensemble de données. Les comparaisons établies par appariement semblent produire des estimations prudentes de la fiabilité des données, car la modification de l'état des malades pourrait constituer une cause légitime d'écarts entre deux types de dossiers.

Notre analyse propose aux chercheurs une stratégie pratique : vérifier la qualité des données des hôpitaux et des médecins en utilisant, dans la mesure du possible, des méthodes fondées sur l'appariement. Au besoin, on pourrait approfondir l'étude des problèmes révélés par l'appariement en procédant à une seconde saisie. Cette stratégie réduirait de façon marquée la seconde saisie et la collecte d'autres données primaires.

### 4. DONNÉES PHARMACEUTIQUES

Cinq études de données pharmaceutiques indiquent des niveaux généralement élevés d'exhaustivité et d'exactitude, quoique la couverture des données pharmaceutiques canadiennes ne s'étende habituellement qu'à une certaine partie de la population. Seuls la Saskatchewan et le Manitoba possèdent des bases de données sur les ordonnances qui sont pertinentes pour presque toute la population (Miller et coll., 1996).

Les bases de données pharmaceutiques provinciales présentent certaines lacunes. Il n'est pas toujours possible de calculer la posologie. Les médicaments administrés dans les hôpitaux, les maisons de soins infirmiers ou les services de consultations externes n'ont pas été enregistrés dans les bases de données provinciales. Bien que les bases de données sur les demandes de remboursement des médicaments délivrés sur ordonnance manquent de renseignements diagnostiques, on commence à utiliser les données liées aux demandes de remboursement des médecins (qui fournissent ordinairement au moins un diagnostic) pour le dépistage des cas.

## 5. ÉTENDRE LA PORTÉE

Pour étendre la portée des systèmes d'appariement de dossiers, il faut disposer de données sur l'utilisation pour cerner non seulement les soins de santé, mais aussi certains états de santé (Roos et Roos, 2001; Virnig et McBean, 2001). Parfois, on peut cerner de manière fiable un état particulier en utilisant uniquement des données hospitalières; avec les algorithmes pertinents, l'infarctus aigu du myocarde semble en représenter un bon exemple (Cox et coll., 1997; Levy et coll., 1999). Dans quatre études axées sur d'autres états de santé, on a employé une méthode de validité conceptuelle utilisant plus d'une source de données (demandes de remboursement des médecins, dossiers de sortie des hôpitaux) ou plus d'un point dans le temps pour obtenir la meilleure approximation possible de la prévalence révélée par les enquêtes ou les examens physiques. Plus précisément, l'appariement des enquêtes et des mesures de laboratoire avec des données administratives a permis de déterminer de manière indépendante des cas de maladies chroniques (comme le diabète et l'hypertension), alors qu'il faut habituellement communiquer avec le système de soins de santé pendant une certaine période (Robinson et coll., 1997; Muhajarine et coll., 1997; Hebert et coll., 1999). On met également au point des mesures plus générales de l'état de santé. Par exemple, une mesure de l'état de santé individuel axée sur les contacts, le système casuistique Adjusted Clinical Group (ACG), a réussi à utiliser des renseignements diagnostiques disponibles dans deux provinces. La corrélation élevée entre la mesure ACG (agrégée pour les petites régions) et la mortalité prématurée, qui est la meilleure mesure de l'état de santé en petite région, laisse supposer une grande validité conceptuelle (Reid et coll., 2001a et 2001b).

## 6. RÉALISATIONS ET PERSPECTIVES

D'une province canadienne à l'autre, l'élaboration des bases de données sur les soins de santé présente une grande hétérogénéité (tableau 1). On observe une disparité semblable entre les systèmes d'appariement de dossiers de certains États australiens. Ces disparités sont attribuables à l'« esprit d'entreprise » d'une ou de plusieurs personnes sur un certain nombre d'années, ainsi qu'à divers degrés de collaboration entre le monde universitaire et l'administration publique.

Les registres ne s'imposent pas comme étant essentiels à la recherche sur la santé et les soins de santé au même titre que les dossiers détaillés concernant les hôpitaux, les médecins, etc. Les registres sont axés sur des études à long terme et sur l'épidémiologie plutôt que sur certains types de recherche sur les services de santé. Le manque de recherche sur la qualité des suivis et de renseignements sur le lieu de résidence constitue un grave problème pour les chercheurs canadiens.

De nouvelles études bénéficieront également des progrès considérables réalisés sur le plan du recours aux dossiers multiples pour le dépistage de cas et l'évaluation de l'état de santé. Par contre, on a constaté que l'intensité de la concordance entre les demandes de remboursement, les auto-évaluations et les mesures cliniques variait considérablement selon l'âge, la profession, l'obésité et l'état de santé. Il faudrait sans doute porter attention à ces écarts.

À l'avenir, il est probable que le domaine en pleine croissance de l'épidémiologie sociale bénéficiera des améliorations apportées aux données des registres (Berkman et Kawachi, 2000). Le lieu de résidence constitue un indicateur important du statut social, qui peut s'avérer essentiel pour de telles études. On a estimé au micro-niveau les données provinciales et nationales sur les coûts et on les a intégrées à la recherche permanente. On évalue et on organise les données sur les soins à domicile. On s'efforce de surveiller l'immunisation des adultes, en plus de celle des enfants, au Manitoba. Des efforts importants pour appairer les données administratives et cliniques sont en cours au Québec.

Il faut envisager la préoccupation nécessaire pour la mesure à la lumière des grandes économies d'efforts, de l'aptitude à considérer des populations entières et du nombre élevé de cas faisant l'objet de données administratives. En raison des effets multiplicateurs qui se manifestent lorsqu'on dispose d'un plus grand nombre d'ensembles de données pour l'analyse, il est très important de comprendre la qualité de ces ensembles de données. Si la recherche axée sur une population peut tolérer un degré d'erreur plus élevé que des travaux axés plus étroitement sur l'individu, les environnements riches en information qui sont créés présentent

cependant certaines lacunes. Si nous en sommes conscients, nous pouvons maintenir et améliorer une capacité de recherche provinciale et nationale, ce qui nous permettra de mieux connaître un certain nombre de domaines.

Tableau 1  
Aperçu de la vérification de la qualité des données dans cinq provinces

	Registres	Hôpitaux	Ordonnances	Médecins
Colombie-Britannique	il faut vérifier le suivi et les identificateurs du lieu de résidence	les vérifications des diagnostics et des méthodes laissent à désirer	renseignements non disponibles	renseignements non disponibles
Saskatchewan	renseignements non disponibles; il faut publier les vérifications effectuées	on vérifie les diagnostics et les méthodes	on effectue certaines vérifications importantes	on vérifie les diagnostics et les méthodes
Manitoba	il faut vérifier les identificateurs du lieu de résidence	on vérifie les diagnostics et les méthodes	on vérifie les demandes de remboursement produites; on effectue certaines vérifications importantes	on vérifie les diagnostics et les méthodes
Ontario	renseignements non disponibles	on vérifie les diagnostics et les méthodes	renseignements non disponibles	on vérifie les méthodes
Québec	renseignements non disponibles	on vérifie les diagnostics et les méthodes	on vérifie les demandes de remboursement produites	on vérifie les diagnostics et les méthodes

Cet aperçu tient compte des études publiées jusqu'en juin 2001.

## REMERCIEMENTS

Cette recherche a été soutenue par l'Institute of Health Economics d'Edmonton (Canada), par le Centre d'élaboration et d'évaluation de la politique des soins de santé du Manitoba et par une bourse de recherche scientifique accordée au professeur L.L. Roos par le Programme national de recherche et développement en matière de santé. Les auteurs remercient le Centre de recherche de l'Hôpital général de Saint-Boniface pour l'aide qu'il leur a apportée.

## BIBLIOGRAPHIE

- Berkman, LF. Kawachi, I. (éds). (2000), *Social Epidemiology*, New York: Oxford University Press.
- Cox, JL. Melady, MP. Chen, E. Naylor, CD. (1997), "Towards improved coding of acute myocardial infarction in hospital discharge abstracts: A pilot project", *Can J Cardiol*, 13(4), pp. 351-358.
- Edouard, L. et Rawson NSB. (1996), "Reliability of the recording of hysterectomy in the Saskatchewan health care system", *Br J Obstet Gynecol*, 103(9), pp. 891-897.

- Hebert, PL. Geiss, LS. Tierney, EF. Engelgau, MM. Yawn, BP. McBean, AM. (1999), "Identifying persons with diabetes using medicare claims data", *Am J Med Qual*, 14(6), pp. 270-277.
- Levy, AR. Tamblyn, RM. Fitchett, D. McLeod, PJ. Hanley, JA. (1999), "Coding accuracy of hospital discharge data for elderly survivors of myocardial infarction", *Can J Cardiol*, 15(11), pp. 1277-1282.
- Miller, E. Blatman, B. Einerson, TR. (1996), "A survey of population-based drug databases in Canada", *Can Med Assoc J*, 154(12), pp. 1855-1864.
- Muhajarine, N. Mustard, CA. Roos, LL. Young, TK. Gelskey, DE. (1997), "Comparison of survey data and physician claims data for detecting hypertension", *J Clin Epidemiol*, 50(6), pp. 711-718.
- Potvin, L. et Champagne, F. (1986), "Utilization of administrative data files in health research", *Social Indicators Research*, 18, pp. 409-423.
- Rawson, NSB. et Malcolm, E. (1995), *Validity of the Recording of Cholecystectomy and Hysterectomy in the Saskatchewan Health Care Datafiles. Technical Report Series (Report #3)*. Saskatoon, SK: Pharmacoepidemiology Research Unit, Technical report series.
- Rawson, NSB. et Malcolm, E. (1995), "Validity of the recording of ischaemic heart disease and chronic obstructive pulmonary disease in the Saskatchewan health care datafiles", *Stat Med*, 14(24), pp. 2627-2643.
- Rawson, NSB. Malcolm, E. D'Arcy, C. (1997), "Reliability of the recording of schizophrenia and depressive disorder in the Saskatchewan health care datafiles," *Soc Psychiatry Psychiatr Epidemiol*, 32(4), pp. 191-199.
- Reid, RJ. MacWilliam, L. Verhulst, L. Roos, NP. Atkinson, M. (2001) "Performance of the ACG Case-mix System in two Canadian provinces", *Med Care*, 39(1), pp. 86-99.
- Reid, RJ. Roos, NP. MacWilliam, L. Frohlich, N. Black C. (2001). "Assessing population need using a claims-based ACG morbidity measure: A validation analysis in the province of Manitoba.", (Unpub).
- Robinson, JR. Young, TK. Roos, LL. Gelskey, DE. (1997), "Estimating the burden of disease: Comparing administrative data and self-reports", *Med Care*, 35(9), pp. 932-947.
- Roos, LL. et Roos, NP. (2001), "Of space and time, of health care and health", *J Health Serv Res Pol*, 6(2), pp. 120-122.
- Virnig, BA et McBean, M. (2001), "Administrative data for public health surveillance and planning", *Annu Rev Public Health*, 22, pp. 213-230.
- Weiner, J. (1999), *Time, Love, Memory: A Great Biologist and His Quest for the Origins of Behavior*, New York: Alfred A. Knopf.