

Sources de données des populations utilisées en surveillance des maladies chroniques

LM Lix, Ph.D. (1); MS Yogendran, M.Sc. (2); SY Shaw, M.A. (3); C Burchill, M.Sc. (2); C Metge, Ph.D. (2,4); R Bond, M.A. (2)

RÉSUMÉ

Dans cette étude, on a estimé la concordance entre les données des populations administratives et les données des populations d'enquête lorsqu'il s'agit de déterminer les cas d'arthrite, d'asthme, de diabète, de cardiopathie, d'hypertension et d'accident vasculaire cérébral. Des définitions de cas de maladies chroniques, qui variaient de par la source des données, le nombre d'années de données et les codes de diagnostic et de médicaments prescrits employés, ont été élaborées à partir des données administratives du Manitoba. Un recouplement a ensuite été fait entre ces données et celles tirées de l'Enquête sur la santé dans les collectivités canadiennes. La concordance entre ces deux sources de données, estimée par le coefficient κ , a été calculée pour chaque définition de cas, et les différences ont été testées. Les variables socio-démographiques et les variables liées à la comorbidité en lien avec la concordance ont été testées au moyen d'une régression logistique pondérée. C'est pour le diabète et l'hypertension que l'on a observé la concordance la plus élevée, et pour l'arthrite que l'on a observé la concordance la plus faible. Les éléments de la définition de cas qui contribuaient à la concordance la plus élevée entre les deux sources de données des populations variaient selon les maladies chroniques. Il est probable que les données administratives et les données d'enquête concordent peu dans le cas des affections qui sont difficiles à diagnostiquer, mais cette concordance est aussi influencée par les caractéristiques socio-démographiques et l'état de santé des individus. Toute définition de cas de maladie chronique fondée sur des données administratives devrait être accompagnée d'une justification du choix de chacun de ses éléments.

Mots clés : maladie chronique, détermination des cas, recouplement de registres, santé de la population, surveillance

Introduction

Les données des populations sur la prévalence des maladies chroniques sont essentielles pour décrire le fardeau de la maladie et pour planifier et évaluer les stratégies de prévention, de traitement et de prise en charge des maladies¹. Les bases de données administratives et les enquêtes sur la santé fondées sur l'auto-déclaration sont deux sources importantes de données des populations permettant d'estimer la prévalence des maladies chroniques²⁻⁵. Toutes deux présentent toutefois des limites : les bases de données pour des

considérations liées à l'exactitude de l'information relative au diagnostic, et les enquêtes pour des considérations liées à la validité de l'auto-déclaration des diagnostics^{2,6-8}. La concordance entre ces deux sources de données a été évaluée dans quelques études au moyen de techniques de recouplement de registres sur des sujets précis^{2,9,10}. Le recouplement de registres permet aussi d'étudier les caractéristiques individuelles qui ont pour effet de modifier la concordance entre les deux sources de données.

Coordonnées des auteurs

1 School of Public Health, University of Saskatchewan, Saskatoon (Saskatchewan)

2 Centre manitobain des politiques en matière de santé, Université du Manitoba, Winnipeg (Manitoba)

3 Département des sciences de la santé communautaire, Université du Manitoba, Winnipeg (Manitoba)

4 Faculté de pharmacie, Université du Manitoba, Winnipeg (Manitoba)

Correspondance : Lisa M. Lix, School of Public Health, University of Saskatchewan, Health Sciences Building, 107 Wiggins Road, Saskatoon SK S7N 5E5, Tél. : 306-966-1617,

Télécopieur : 306-966-7920, Courriel : lisa.lix@cpe.umanitoba.ca; lisa.lix@usask.ca

On trouve très peu d'études systématiques sur l'effet qu'a la définition de cas administrative choisie sur la concordance entre les deux sources de données des populations. Les définitions de cas administratives des maladies chroniques sont établies en choisissant une combinaison des éléments de données suivants : source des données, codes de diagnostic ou de traitement, nombre d'années de données et nombre de consultations affichant les codes sélectionnés dans les registres administratifs⁹⁻¹¹. Il n'y a aucun consensus quant à la définition de cas optimale, aussi la définition de cas est elle souvent fondée sur les données dont on dispose au sein de l'administration concernée.

Dans cette étude, on s'est penché sur de multiples définitions utilisées pour déterminer les cas d'arthrite, d'asthme, de diabète, de cardiopathie, d'hypertension et d'accident vasculaire cérébral (AVC) à partir de données administratives et on a comparé les cas ainsi déterminés avec les cas auto-déclarés de maladies chroniques tirés des données d'enquête en population. Les objectifs étaient les suivants : 1) tester la concordance entre les données administratives et les données d'enquête sur les maladies chroniques lorsque les éléments de la définition de cas de maladie chronique sont systématiquement manipulés; 2) étudier les caractéristiques démographiques, géographiques et socio-économiques et les caractéristiques liées à l'état de santé des individus pouvant influer sur la concordance entre ces deux sources de données.

Méthodologie

L'étude a été menée à l'aide des données des populations du Manitoba, une province du Centre du Canada jouissant d'un système de santé universel. Les archives des données de recherche du Centre manitobain des politiques en matière de santé renferment les registres administratifs fournis par le ministère de la Santé de la province relativement au Régime d'assurance-maladie du Manitoba. Ces archives renferment également les données des populations tirées de l'Enquête sur la santé dans les collectivités canadiennes (ESCC), et ces deux sources peuvent être reliées directement au moyen d'un numéro d'identification sanitaire personnel unique et anonyme.

Pour établir les définitions de cas, on a choisi les bases de données sur les services hospitaliers, sur les consultations médicales et sur les médicaments de prescription. Dans d'autres études, ces bases de données administratives ont été utilisées pour déterminer les cas de maladies chroniques^{7,10,12}. Lorsqu'un patient obtient son congé d'un établissement de soins actifs, un résumé de dossier d'hospitalisation est rempli. Chaque résumé contient jusqu'à 16 codes de diagnostic tirés de la Classification internationale des maladies, 9^e révision, modifications cliniques (CIM-9-MC). On a fait une recherche sur ces 16 codes pour déterminer les cas de maladie pour cette étude. Tous les médecins qui sont rémunérés à l'acte présentent des factures au ministère de la Santé de la province. Ces factures font état de tous les services externes, y compris ceux offerts dans les hôpitaux au service des urgences et au service des consultations externes, et ceux offerts aux résidants des établissements de soins de longue durée. Tandis que certains médecins sont salariés (soit environ 7 % des médecins de famille du Manitoba¹³), environ 90 % de ces médecins présentent aussi des factures parallèles à des fins administratives. Par conséquent, les bases de données sur la facturation font état de presque toutes les consultations avec des médecins au Manitoba. Les demandes d'honoraires des médecins contiennent un code CIM-9-MC. Les fiches sur les médicaments prescrits aux patients externes sont saisies dans le Réseau pharmaceutique informatisé (RPI),

une base de données électronique centralisée reliant toutes les pharmacies de détail du Manitoba. Le RPI renferme une variété d'information sur chaque ordonnance délivrée, y compris la date, le nom du médicament et son numéro d'identification (DIN). Les DIN sont liés à des codes de classification anatomique, thérapeutique et chimique (ATC) mis au point par l'Organisation mondiale de la santé. Ces codes classent les médicaments de prescription dans divers groupes à chacun des cinq niveaux, selon l'organe ou le système sur lequel ils agissent et/ou selon leurs caractéristiques thérapeutiques et chimiques¹⁴. On a considéré que les données du RPI étaient exactes pour ce qui est de consigner tant les médicaments délivrés que les détails sur les ordonnances¹⁵. La base de données sur les médicaments prescrits ne contient pas d'information sur les médicaments obtenus sans ordonnance, y compris les médicaments en vente libre, les échantillons de médicaments gratuits offerts directement aux patients par les médecins et les médicaments fournis aux patients hospitalisés ou aux résidants des établissements de soins de longue durée.

L'ESCC fournit des estimations transversales de l'état de santé, des déterminants de la santé et de l'utilisation du système de santé concernant les habitants de 136 régions sanitaires au Canada, dont 11 sont au Manitoba¹⁶. L'Enquête est fondée sur un plan multi-degrés stratifié avec grappes. Elle

représente environ 98 % de la population canadienne âgée de 12 ans et plus. Dans le cycle 1.1, recueilli entre septembre 2000 et novembre 2001, on comptait 8 120 répondants au Manitoba.

On a fait un recouplement anonyme entre les données administratives et les données d'enquête sur les répondants qui ont donné leur consentement ($n = 7 560$; 93,1 %). Après avoir retiré les répondants dont le numéro d'identification sanitaire personnel était non valide ou manquant, on a pu établir un lien direct entre les données administratives et les données d'enquête pour 6 812 répondants (soit 83,9 % des répondants). On a formé une cohorte de répondants qui étaient couverts de façon continue par le régime d'assurance-maladie du Manitoba depuis au moins cinq ans au moment de leur entrevue d'enquête ($n = 6 422$). On a exclu les répondants couverts depuis moins de cinq ans par ce régime parce que les analyses préliminaires ont révélé qu'une couverture discontinue par l'assurance-maladie était associée à une concordance inférieure entre les données administratives et les données d'enquête. Les données administratives visent les exercices financiers 1997-1998 à 2001-2002.

Dans le cadre de l'entrevue d'enquête, on donnait aux répondants les directives suivantes : « Les questions qui suivent portent sur certains problèmes de santé de longue durée que vous pouvez avoir. Par "problème de santé de longue durée", on

TABLEAU 1
Questions d'enquête, codes de diagnostic de la CIM-9-MC et codes de classification ATC des médicaments de prescription utilisés pour déterminer les cas de maladie chronique

Maladie	Questions d'enquête	Codes CIM-9-MC	Codes ATC ^b
Arthrite	Êtes-vous atteint d'arthrite ou de rhumatisme, à l'exclusion de la fibromyalgie?	714, 715, 446, 710, 720, 274, 711-713, 716, 717, 718 719, 721, 725-729, 739	A07, J01, L01, L04, M01, P01, N02, R05, H02
Asthme	Êtes-vous atteint d'asthme?	493	R03, R06
Diabète	Êtes-vous atteint de diabète?	250	A10
Cardiopathie	Êtes-vous atteint d'une cardiopathie? ^a	410-414	C01, C07, C08, C09
Hypertension	Faites-vous de l'hypertension?	401	C02, C03, C07, C08, C09
AVC	Présentez-vous des séquelles à la suite d'un AVC?	430-438	B01

^a Les personnes atteintes d'insuffisance cardiaque congestive ont été exclues de la catégorie « cardiopathie ».

^b Pour chaque classe ATC, ce ne sont pas tous les médicaments qui ont été sélectionnés. Une liste complète des médicaments de prescription inclus et exclus est offerte à l'adresse suivante : <http://mchp-appserv.cpe.umanitoba.ca/reference/chronic.disease.pdf>

entend un état qui dure ou qui devrait durer six mois ou plus, et qui a été diagnostiquée par un professionnel de la santé. » Les questions pertinentes sont présentées dans le tableau 1.

Un total de 152 définitions de cas ont été étudiées (la liste complète peut être obtenue auprès de l'auteur-ressource sur demande), et ont été mises au point en s'inspirant des recherches antérieures^{5,9-11,17-27}. Certaines définitions n'étaient fondées que sur une seule source de données administratives, tandis que d'autres étaient fondées sur plusieurs sources. Les définitions de cas variaient de par le nombre d'années de données administratives (une, deux, trois ou cinq), conformément aux recherches antérieures⁹. Certaines définitions n'exigeaient qu'une seule consultation soit consignée dans les données administratives, tandis que d'autres demandaient plus d'une consultation. Les éléments constants dans les définitions de cas étaient le diagnostic (code de la CIM-9-MC) et les codes de médicaments prescrits (codes de classification ATC) (tableau 1); ces éléments ont été déterminés au moyen d'une analyse documentaire exhaustive²⁸.

On s'est servi du coefficient kappa de Cohen (κ) pour quantifier la concordance entre les deux sources de données; des intervalles de confiance (IC) à 95 % ont aussi été calculés. Les estimations et les IC ont été calculés à l'aide des poids d'échantillonnage issus des données d'enquête. Le coefficient kappa est une mesure adoptée couramment car elle corrige la concordance entre deux sources en tenant compte de la proportion de concordance due au hasard. L'ampleur de la concordance a été évaluée de la façon suivante²⁹ : concordance faible : $\kappa < 0,20$; concordance passable : $0,20 \leq \kappa < 0,40$; concordance modérée : $0,40 \leq \kappa < 0,60$; bonne concordance : $0,60 \leq \kappa < 0,80$; très bonne concordance : $\kappa \geq 0,80$.

On a mis à l'épreuve les différences entre les coefficients κ sélectionnés (c.-à-d. H_0 : $\kappa_1 = \kappa_2$) au moyen d'un test de validité de l'ajustement³⁰; la distribution suivie était d'environ χ^2 . Les tests ont été faits pour les définitions de cas comprenant différentes sources, différents nombres d'années de données ou différents nombres de

consultations. Toutes les comparaisons ont été identifiées a priori, et le taux d'erreurs par comparaison a été fixé à $\alpha = 0,05$.

Pour chaque maladie chronique, on a sélectionné la définition de cas pour laquelle la concordance estimée était la plus élevée afin d'étudier plus à fond les caractéristiques des répondants associées à une concordance entre les données des populations administratives et les données des populations d'enquête. Des analyses de régression logistique pondérée ont permis de modéliser la concordance comme une fonction des variables liées aux caractéristiques socio-démographiques et aux maladies, y compris

le groupe d'âge (12 à 18, 19 à 49, 50 à 64, 65 à 74, 75+ [référence]), le sexe (homme, femme [référence]), la région de résidence (rurale, urbaine [référence]), le quintile du revenu adéquat (la catégorie de revenu la plus élevée était la référence) et la présence/absence d'affections comorbidies (la présence de telles affections était la référence)^{31,32}. Les spécialistes en méthodologie de l'ESCC ont établi des quintiles du revenu adéquat à l'aide du revenu total auto-déclaré du ménage et du nombre de personnes dans le ménage. Les affections comorbidies suivantes ont été sélectionnées : allergies, emphysème ou maladie pulmonaire obstructive chronique dans le

Figure 1
La valeur de $\hat{\kappa}$ pondérée associée aux définitions de cas de maladies chroniques fondées sur des données administratives.

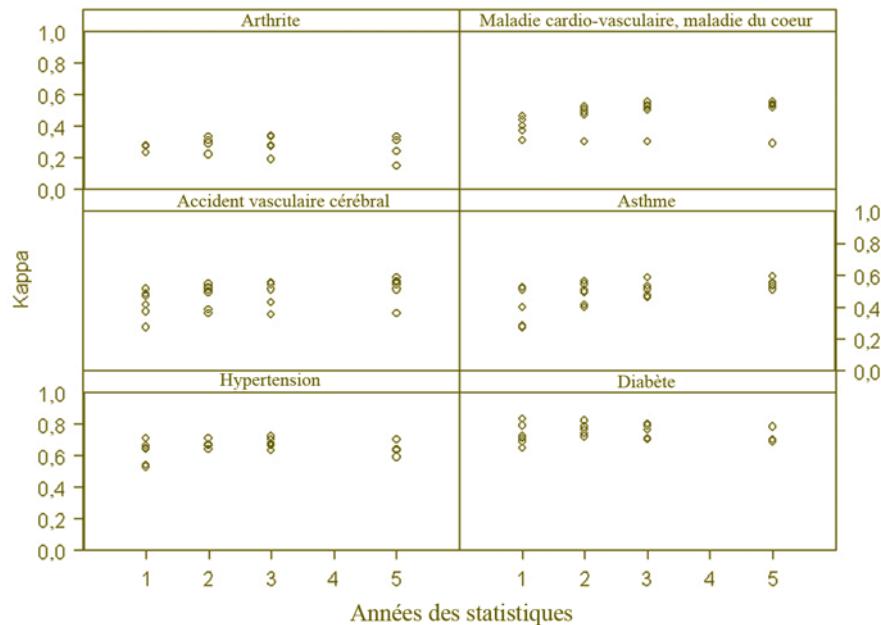


TABLEAU 2

Nombre (et pourcentage) de membres de la cohorte atteints d'une maladie chronique auto-déclarée et estimations de la prévalence tirées de l'Enquête sur la santé dans les collectivités canadiennes, cycle 1.1

Maladie	Cohorte d'adultes (n = 5 589)	Cohorte de jeunes (n = 833)	Prévalence (%) (IC à 95 %) ^a
Arthrite	1 344 (24,0)	–	18,6 (17,4 to 19,8)
Asthme	418 (7,5)	111 (13,5)	8,6 (7,6 to 9,6)
Diabète	337 (6,0)	–	4,5 (3,9 to 5,2)
Cardiopathie	371 (6,6)	–	5,1 (4,4 to 5,7)
Hypertension	1 033 (18,5)	–	15,3 (14,2 to 16,5)
AVC	108 (1,9)	–	1,5 (1,1 to 1,8)

^a Les estimations de la prévalence ont été calculées pour la population âgée de 19 ans et plus, sauf dans le cas de l'asthme, dont la prévalence a été calculée pour la population âgée de 12 ans et plus. Les estimations de la prévalence sont fondées sur les pondérations de l'enquête, et les intervalles de confiance ont été calculés à l'aide de méthodes d'estimation *bootstrap* des variances.

Tableau 3

Valeurs de κ (IC à 95 %) pondérées associées aux définitions de cas de maladies chroniques sélectionnées fondées sur des données administratives^{a,b,c,d}

Source et nombre d'années de données	Arthrite	Asthme	Diabète	Cardiopathie	Hypertension	AVC
Médecin (M) ^a	1 M	1 M	1 M	1 M	1 M	
1	0,27 (0,26-0,27)	0,40 (0,35-0,44)	0,69 (0,68-0,69)	0,44 (0,39-0,49)	0,64 (0,64-0,64)	0,47 (0,47-0,48)
2	0,29 (0,28-0,29)	0,49 (0,45-0,54)	0,72 (0,72-0,73)	0,51 (0,46-0,55)	0,66 (0,66-0,67)	0,54 (0,53-0,54)
3	0,27 (0,27-0,28)	0,53 (0,49-0,57)	0,70 (0,69-0,70)	0,52 (0,48-0,56)	0,66 (0,66-0,67)	0,51 (0,50-0,52)
5	0,24 (0,24-0,24)	0,55 (0,51-0,59)	0,69 (0,69-0,69)	0,52 (0,48-0,57)	0,63 (0,63-0,64)	0,54 (0,54-0,55)
	2 + M	2 + M	2 + M	2 + M	2 + M	2 + M
1	0,23 (0,23-0,23)*	0,27 (0,23-0,32)*	0,65 (0,65-0,66)*	0,37 (0,31-0,42)*	0,53 (0,52-0,53)*	0,41 (0,40-0,42)*
2	0,31 (0,30-0,31)*	0,40 (0,35-0,44)*	0,76 (0,75-0,76)*	0,47 (0,42-0,52)*	0,64 (0,64-0,65)*	0,49 (0,48-0,50)*
3	0,34 (0,33-0,34)*	0,46 (0,42-0,51)*	0,78 (0,78-0,79)*	0,50 (0,46-0,55)*	0,68 (0,68-0,68)*	0,55 (0,54-0,56)*
5	0,33 (0,33-0,34)*	0,54 (0,50-0,58)	0,78 (0,78-0,79)*	0,54 (0,49-0,58)*	0,70 (0,70-0,70)*	0,58 (0,57-0,59)*
Hôpital (H), Médecin (M) ^b	1+ H ou 2+ M	1+ H ou 1+ M	1+ H ou 1+ M	1+ H ou 2+ M	1+ H ou 2+ M	1+ H ou 2+ M
1	0,23 (0,23-0,24)	0,40 (0,36-0,45)	0,72 (0,72-0,73)†	0,40 (0,34-0,45)†	0,54 (0,54-0,54)†	0,48 (0,47-0,48)†
2	0,31 (0,31-0,31)	0,50 (0,46-0,54)	0,74 (0,73-0,74)†	0,49 (0,44-0,54)†	0,66 (0,65-0,66)†	0,52 (0,51-0,53)†
3	0,34 (0,34-0,34)	0,53 (0,50-0,57)	0,70 (0,70-0,71)†	0,53 (0,49-0,58)†	0,70 (0,69-0,70)†	0,55 (0,54-0,55)
5	0,33 (0,33-0,34)	0,55 (0,51-0,59)	0,70 (0,69-0,70)†	0,55 (0,51-0,59)†	0,71 (0,71-0,71)†	0,56 (0,55-0,57)†
Hôpital (H), Médecin (M), Ordonnance (Rx) ^c	1+ H ou 2+ M ou (1+ P et 2+ Rx)	1+ H ou 1+ M ou 1+ Rx	1+ H ou 1+ M ou 1+ Rx	1+ H ou 2+ M ou (1+ M et 2+ Rx)	1+ H ou 2+ M ou (1+ M et 2+ Rx)	1+ H ou 2+ M ou (1+ M et 2+ Rx)
1	0,28 (0,28-0,28)‡	0,52 (0,48-0,56)‡	0,79 (0,79-0,80)‡	0,46 (0,41-0,51)‡	0,66 (0,66-0,67)‡	0,51 (0,50-0,52)‡
2	0,33 (0,33-0,33)‡	0,54 (0,50-0,57)‡	0,76 (0,76-0,77)‡	0,52 (0,48-0,57)‡	0,71 (0,71-0,71)‡	0,51 (0,51-0,52)
3	0,33 (0,33-0,34)‡	0,51 (0,48-0,55)	0,71 (0,71-0,72)‡	0,55 (0,50-0,59)‡	0,72 (0,71-0,72)‡	0,54 (0,53-0,55)
5	0,31 (0,31-0,31)‡	0,48 (0,45-0,51)‡	0,69 (0,69-0,70)	0,55 (0,51-0,60)	0,70 (0,70-0,71)‡	0,55 (0,54-0,55)‡

^a Les valeurs présentées en caractères gras représentent les valeurs de kappa les plus élevées pour chaque définition de cas.

^b L'astérisque (*) indique une différence statistiquement significative entre un algorithme fondé sur une consultation chez le médecin (selon le registre des consultations médicales) et un algorithme fondé sur deux consultations et plus chez le médecin ($p < 0,05$); toutes les comparaisons sont fondées sur le même nombre d'années de données.

^c La croix (†) indique une différence statistiquement significative entre un algorithme fondé sur le registre des consultations médicales (une ou deux consultations) et un algorithme fondé sur le registre des services hospitaliers et le registre des consultations médicales ($p < 0,05$); toutes les comparaisons sont fondées sur le même nombre d'années de données.

^d La croix double (‡) indique une différence statistiquement significative entre un algorithme fondé sur le registre des services hospitaliers et le registre des consultations médicales et un algorithme fondé sur le registre des services hospitaliers, le registre des consultations médicales et le registre sur les médicaments de prescription ($p < 0,05$); toutes les comparaisons sont fondées sur le même nombre d'années de données.

^e Double Dagger (†) denotes a statistically significant difference between an algorithm based on hospital and physician data and an algorithm based on hospital, physician and prescription drug data ($p < 0,05$); all comparisons are based on the same number of years of data.

cas de l'asthme; cardiopathie ou hypertension dans le cas du diabète; diabète ou hypertension dans le cas des cardiopathies; diabète ou cardiopathie dans le cas de l'hypertension; cardiopathie ou diabète dans le cas de l'accident vasculaire cérébral (AVC). Les rapports de cotes (RC) et les IC à 95 % sont indiqués. Toutes les analyses ont été menées en utilisant le logiciel SAS, version 9,1.

Sur le plan éthique, l'étude a été approuvée par le conseil d'éthique de la recherche en santé de l'Université du Manitoba. L'accès aux données a été approuvé par le comité de protection des renseignements personnels de Santé Manitoba.

Résultats

Dans le cas du diabète, des cardiopathies, de l'hypertension, de l'arthrite et de l'AVC, seuls les répondants adultes (19 ans et plus) ($n = 5\,589$; 87,0 %) ont fait l'objet de l'étude sur la concordance entre les données administratives et les données d'enquête, tandis qu'on s'est servi tant de la cohorte d'adultes que de la cohorte de jeunes dans les analyses sur l'asthme. Près de la moitié (46,1 %) des sujets adultes étaient âgés de 50 ans et plus. Un peu moins de la moitié de tous les sujets de cette cohorte (45,3 %) étaient des hommes, et près du quart (23,1 %) vivaient en milieu urbain. Dans l'ESCC, on a suréchantillonné

les personnes vivant en milieu rural pour pouvoir générer des estimations pour les régions sanitaires individuelles.

Le nombre et le pourcentage de répondants ayant déclaré chacune des maladies chroniques sont présentés dans le tableau 2. Les estimations pondérées de la prévalence des maladies chroniques ont été calculées à partir des données d'enquête (voir le tableau 2), et allaient de 1,5 % dans le cas de l'AVC à 18,6 % dans le cas de l'arthrite. Ces estimations provinciales correspondent aux estimations faites à partir des données d'enquête auto-déclarées provenant d'autres études provinciales et nationales³³⁻³⁵.

TABLEAU 4
Rapports de cotes* (IC à 95 %) des prédicteurs de concordance entre les données administratives et les données d'enquête sur les maladies chroniques

	Arthrite	Asthme	Diabète ^{a,b}	Cardiopathie ^{a,b}	Hypertension	AVC ^a
Âge						
12 à 18 ans	—	1,38 (1,33 à 1,43)	—	—	—	—
19 à 49 ans	3,67 (3,62 à 3,77)	2,17 (2,11 à 2,24)	—	—	6,00 (5,83 à 6,17)	—
50 à 64 ans	1,57 (1,53 à 1,60)	1,86 (1,79 à 1,92)	2,34 (2,28 à 3,46)	8,46 (8,21 à 8,72)	2,42 (2,35 à 2,48)	13,47 (12,75 à 14,23)
65 à 74 ans	1,39 (1,35 à 1,42)	1,85 (1,78 à 1,93)	0,85 (0,81 à 0,89)	1,77 (1,71 à 1,82)	1,93 (1,88 à 1,99)	1,90 (1,81 à 1,99)
75 ans et +	Réf.	Réf.	Réf.	Réf.	Réf.	Réf.
Sexe						
Hommes	1,09 (1,08 à 1,11)	1,05 (1,03 à 1,07)	1,40 (1,35 à 1,45)	0,83 (0,81 à 0,85)	0,81 (0,79 à 0,82)	0,44 (0,42 à 0,46)
Femmes	Réf.	Réf.	Réf.	Réf.	Réf.	Réf.
Résidence						
Région rurale	1,06 (1,04 à 1,07)	1,01 (0,99 à 1,02)	1,18 (1,14 à 1,22)	1,31 (1,28 à 1,34)	0,89 (0,88 à 0,91)	1,28 (1,23 à 1,33)
Rég. urbaine	Réf.	Réf.	Réf.	Réf.	Réf.	Réf.
Quintile du revenu						
Le plus faible	0,66 (0,68 à 0,71)	0,33 (0,31 à 0,34)	—	—	0,84 (0,80 à 0,89)	0,07 (0,06 à 0,09)
Moyen-faible	1,06 (1,03 à 1,09)	0,71 (0,68 à 0,74)	1,07 (1,00 à 1,13)	0,87 (0,83 à 0,91)	1,06 (1,02 à 1,11)	0,10 (0,09 à 0,13)
Moyen	0,98 (0,96 à 1,00)	0,59 (0,57 à 0,60)	1,04 (0,99 à 1,09)	0,66 (0,63 à 0,68)	1,10 (1,07 à 1,13)	0,19 (0,17 à 0,21)
Moyen-élévé	0,91 (0,90 à 0,93)	0,66 (0,64 à 0,68)	1,32 (1,26 à 1,38)	0,78 (0,76 à 0,82)	1,15 (1,12 à 1,18)	0,18 (0,16 à 0,20)
Le plus élevé	Réf.	Réf.	Réf.	Réf.	Réf.	Réf.
Affections comorbidies						
Absence	—	1,90 (1,82 à 1,94)	3,09 (2,97 à 3,21)	2,73 (2,66 à 2,80)	2,96 (2,89 à 3,03)	2,97 (2,84 à 3,10)
Présence	Réf.	Réf.	Réf.	Réf.	Réf.	Réf.

* Statistiquement significatif à $\alpha = 0,05$.

^a On a combiné le groupe d'âge des 19 à 49 ans à celui des 50 à 64 ans en raison de sa petite taille.

^b On a combiné le quintile moyen-faible au quintile moyen en raison de sa petite taille.

La figure 1 illustre la variation entre les estimations du coefficient κ pour toutes les définitions de cas étudiées. Dans le cas de l'arthrite, la valeur de $\hat{\kappa}$ allait de 0,15 à 0,34, ce qui indique une concordance faible à passable. Dans le cas de l'asthme, les estimations allaient de 0,27 à 0,59, ce qui représente une concordance passable à modérée. Pour ce qui est du diabète, la concordance était bonne ou très bonne, allant de 0,65 à 0,83. La valeur de $\hat{\kappa}$ allait de 0,29 à 0,55 dans le cas des cardiopathies, ce qui représente une concordance passable à modérée. En ce qui concerne l'hypertension, l'intervalle allait de 0,53 à 0,72, ce qui indique une concordance modérée à bonne. Pour finir, la valeur de $\hat{\kappa}$ allait de 0,27 à 0,58 dans le cas de l'AVC, ce qui représente une concordance passable à modérée.

Le tableau 3 indique la valeur de $\hat{\kappa}$ pour les définitions de cas sélectionnées, ainsi que

les résultats des analyses inférentielles. On a testé les différences de concordance entre les définitions de cas ne comportant qu'une seule consultation (selon le registre des consultations médicales) et les définitions comportant au moins deux consultations. En se servant d'une année de données administratives, on a observé des valeurs de $\hat{\kappa}$ beaucoup plus faibles lorsqu'au moins deux consultations étaient nécessaires, et ce pour toutes les maladies chroniques. Toutefois, lorsqu'on a utilisé au moins deux années de données administratives dans la définition de cas, le profil des différences de concordance a changé. Pour ce qui est de l'arthrite et du diabète, la concordance estimée était toujours beaucoup plus élevée lorsque la définition de cas incluait deux consultations que lorsqu'elle n'en incluait qu'une seule, tandis que c'était l'inverse en ce qui concerne l'asthme. Pour ce qui est des cardiopathies, de l'hypertension et de l'AVC, la concordance était

beaucoup plus élevée ou beaucoup plus faible, selon le nombre d'années de données utilisé dans la définition de cas.

On a testé les différences entre les définitions de cas fondées sur les diagnostics figurant dans le registre des consultations médicales (une ou deux consultations, selon la maladie) et les définitions de cas fondées à la fois sur le registre des services hospitaliers et le registre des consultations médicales, pour une et pour plusieurs années de données administratives (tableau 3). Lorsque le registre des services hospitaliers était inclus, on a observé des améliorations statistiquement significatives de la concordance, et ce, pour toutes les maladies à l'exception de l'arthrite et de l'asthme. Pour ce qui est des autres maladies, la concordance était presque toujours meilleure, peu importe le nombre d'années de données utilisé dans les définitions de cas.

Le tableau 3 fait état des tests sur les différences entre les définitions de cas fondées sur les diagnostics figurant dans le registre des services hospitaliers et le registre des consultations médicales et les définitions de cas fondées à la fois sur les codes de diagnostic et sur les codes de médicaments de prescription. Pour toutes les maladies chroniques, la combinaison des données sur les diagnostics et des données sur les médicaments de prescription a entraîné une amélioration statistiquement significative de la concordance lorsqu'une seule année de données était utilisée dans la définition de cas. Cela n'était pas toujours le cas lorsqu'on utilisait deux années de données ou plus dans les définitions. En fait, lorsqu'on utilisait cinq années de données, la concordance ne s'améliorait pas ou empirait, et ce, pour toutes les maladies.

Les RC pour les analyses de régression logistique pondérée figurent dans le tableau 4. L'âge était statistiquement significatif dans tous les modèles ($p < 0,0001$). Les cotes de concordance entre les données administratives et les données d'enquête étaient presque toujours plus élevées dans les groupes d'âge plus jeunes que dans les groupes plus âgés. La concordance était aussi liée au sexe pour toutes les maladies chroniques, quoique l'ampleur et la direction de l'effet variaient. L'effet était plus fort dans le cas du diabète et de l'AVC. En ce qui concerne l'arthrite, l'asthme, le diabète et l'AVC, les cotes de concordance étaient beaucoup plus élevées chez les hommes que chez les femmes, mais c'était l'inverse dans le cas des cardiopathies, de l'hypertension et de l'AVC.

La région de résidence était statistiquement significative pour toutes les maladies chroniques ($p < 0,0001$) sauf l'asthme. C'est pour l'arthrite et l'hypertension que la puissance de l'association était la plus faible. Les cotes de concordance étaient supérieures chez les habitants des régions rurales que chez ceux des régions urbaines en ce qui concerne l'arthrite, le diabète, les cardiopathies et l'AVC, tandis que c'était l'inverse dans le cas de l'hypertension.

Le quintile du revenu était aussi associé à la concordance entre les données administratives et les données d'enquête ($p < 0,0001$). Les cotes de concordance étaient en général plus faibles parmi les quintiles plus pauvres que parmi les quintiles plus riches, sauf dans le cas du diabète et de l'hypertension pour lesquels c'était l'inverse.

Pour finir, la comorbidité était statistiquement significative dans tous les modèles ($p < 0,0001$); les cotes de concordance étaient invariablement plus élevées en l'absence qu'en présence d'une affection comorbide. L'association entre la comorbidité et la concordance était d'une importance similaire pour toutes les maladies chroniques à l'exception de l'asthme, pour lequel la relation était la plus faible.

Conclusions

Dans cette étude de recouplement de registres, on a comparé les cas de maladies chroniques déterminés au moyen de données administratives et de données d'enquête auto-déclarées lorsque les éléments de la définition de cas fondée sur les données administratives ont été systématiquement manipulés. On s'est aussi penché sur les caractéristiques individuelles ayant pour effet de modérer la concordance entre ces deux sources de données. Les résultats indiquent que peu importe la définition de cas employée, le diabète et l'hypertension étaient associés à la concordance la plus élevée entre ces deux sources de données, tandis que l'arthrite était liée à la concordance la plus faible. Ces résultats correspondent à ceux obtenus dans une recherche antérieure dans laquelle on a comparé les cas de maladies chroniques déterminés au moyen de définitions administratives et les cas auto-déclarés parmi des échantillons des populations et cliniques^{2,9,10,19,22,31}. Selon l'hypothèse émise par Okura et ses collaborateurs, le diabète et l'hypertension, quoique ces maladies ne soient pas habituellement caractérisées par des signes cliniques distincts et spectaculaires, sont des affections chroniques qui requièrent un recours continu et répété au système de santé^{31(p. 1101)}, ce qui augmente la probabilité qu'elles soient relevées au moyen des données administratives. Pour

ce qui est de l'arthrite, le choix de codes de diagnostic non spécifiques par les praticiens, les diagnostics potentiellement erronés posés par des praticiens non spécialisés et la faible probabilité que cette affection entraîne un séjour à l'hôpital pourraient tous être des facteurs contribuant à une mauvaise concordance entre les deux sources de données³⁶. De plus, il est possible que certaines maladies ne soient pas consignées avec exactitude au moyen de l'auto-déclaration. Par exemple, Kriegsman et ses collaborateurs³⁷ soutiennent que l'arthrite, parce qu'elle n'est pas une affection constituant un danger de mort, pourrait bien être sur-déclarée dans les enquêtes, et sous-déclarée dans les données administratives, ce qui contribuerait à une mauvaise concordance entre ces deux sources. Le choix des termes employés dans le questionnaire est aussi un facteur important pour évaluer l'exactitude des données d'enquête.

La concordance entre les données administratives et les données d'enquête variait de façon importante en fonction du type de données, du nombre d'années de données et du nombre de consultations nécessaire^{9,10}, quoique les différences de la valeur estimée de κ entre les diverses définitions de cas étaient plutôt faibles. L'emploi de plusieurs types de données et/ou de plus d'une année de données avait habituellement pour effet d'améliorer la concordance entre les données administratives et les données d'enquête, mais cette observation ne peut être généralisée à toutes les maladies étudiées. Les améliorations de la concordance n'étaient pas toujours évidentes lorsqu'on utilisait trois ou cinq années de données dans la définition de cas. L'usage des données sur les médicaments prescrits en plus des données sur les services hospitaliers et les consultations médicales pour déterminer les cas de maladie a eu des effets contrastés sur la concordance. Pour ce qui est de l'asthme, l'usage de données sur le diagnostic et sur les médicaments prescrits a entraîné une concordance supérieure lorsque la définition de cas comprenait une ou deux années de données administratives, tandis que cette amélioration de la concordance était moins importante en ce qui concerne le diabète. L'usage des données sur le diagnostic et

des données sur les médicaments prescrits pour déterminer les cas a également amélioré la concordance pour ce qui est de l'hypertension, mais pas pour ce qui est d'autres maladies. On utilise un ensemble précis de médicaments de prescription pour traiter l'asthme et le diabète, tandis que pour d'autres maladies chroniques comme l'hypertension ou l'arthrite, les médicaments prescrits à une personne peuvent servir à traiter plus d'une maladie chronique, et ne sont par conséquent pas toujours utiles pour relever les cas.

L'effet modérateur des variables démographiques, géographiques et socio-économiques et des variables liées à l'état de santé sur la concordance entre les données administratives et les données d'enquête a été observé dans d'autres études³⁸⁻⁴⁰. Simpson et ses collaborateurs³⁹ ont observé que l'absence d'affections comorbides était associée tantôt à une hausse tantôt à une baisse de la concordance, selon la maladie, tandis que dans la présente étude, on a observé une amélioration constante de la concordance en l'absence d'affections comorbides.

Une des limites possibles de cette recherche tient au fait qu'un ensemble fixe de codes de diagnostic et de codes de médicaments prescrits a été sélectionné. Dans d'autres études, on a comparé les résultats de la détermination des cas obtenus à partir de différents ensembles de codes de diagnostic. Par exemple, dans des études sur la détermination des cas d'AVC, on a parfois exclu les codes de diagnostic non spécifiques comme le code 437, qui représente l'AVC de cause indéterminée^{24,41}. Il se peut que l'exclusion des codes non spécifiques améliore les estimations de la concordance entre les données administratives et les données d'enquête. Toutefois, lorsque nous avons répété nos analyses sur l'AVC en nous limitant à un ensemble restreint de codes de diagnostic (c. à d. 430, 431, 434, 345, 436), les estimations du coefficient κ n'ont dans aucun cas été plus élevées que celles obtenues au moyen de l'ensemble complet de codes. Ensuite, cette étude a été menée à l'aide de données administratives codées selon la CIM-9-MC. Au cours de l'exercice 2004-2005, on est passé à la CIM-10-CA dans les bases de données des hôpitaux au Manitoba. D'autres recherches

seront menées sur l'effet de ce changement sur la détermination des cas de maladie chronique. La concordance entre les deux sources de données a peut-être été modérée par la date de l'entrevue d'enquête, un facteur qui n'a pas été pris en considération dans cette étude. Enfin, cette étude était limitée à un ensemble d'affections pour lesquelles ont disposait de suffisamment de données pour étudier la concordance. Ainsi, d'autres affections respiratoires chroniques comme la maladie pulmonaire obstructive chronique, d'autres affections musculosquelettiques comme l'ostéoporose et certaines affections gastro-intestinales comme les maladies inflammatoires de l'intestin ont été exclues de cette étude.

L'avantage de cette étude vient du fait qu'elle évalue la concordance au moyen de trois bases de données administratives qui sont en place dans de nombreuses administrations. Elle a aussi l'avantage d'utiliser différents éléments de données dans les définitions de cas, et elle examine systématiquement l'effet de l'usage combiné des données sur le diagnostic et des données sur les médicaments prescrits. Finalement, on emploie dans cette étude des techniques de recouvrement de registres pour étudier les variables associées à la concordance entre les deux sources données des populations. Les chercheurs qui élaborent des définitions de cas fondées sur des données administratives devraient connaître l'effet qu'a le choix des éléments de données sur la détermination des cas. La sélection d'une définition de cas devrait être justifiée par une description des effets potentiels liés à la manipulation de chaque élément de données.

Les résultats de cette étude donnent à penser qu'il existe des possibilités de recherches futures sur l'élaboration de définitions de cas fondées sur des données administratives. D'abord, on pourrait stratifier la population au moyen d'importantes variables socio-démographiques et/ou liées à l'état de santé, puis établir et mettre à l'épreuve des définitions de cas pour chacune de ces strates. Par exemple, un algorithme fondé sur une seule consultation (selon la base de données sur les honoraires des médecins) au cours d'une année de

données pourrait donner lieu à une concordance élevée chez les jeunes atteints d'asthme, tandis qu'un algorithme fondé sur au moins deux consultations pourrait donner une concordance plus élevée chez les adultes atteints d'asthme. On pourrait aussi envisager d'adopter une approche modélisée utilisant des méthodes de classification pour établir une définition de cas⁴²⁻⁴³. Une approche modélisée peut intégrer un vaste ensemble de variables (c.-à-d. de caractéristiques de données), y compris les affections comorbides, les consultations avec des spécialistes, les variables socio-démographiques et les indicateurs de diagnostic et de traitement, et évaluer la contribution relative de ces variables à l'amélioration de la détermination des cas. En terminant, des études de validation additionnelles utilisant les bases de données d'autres provinces ou une source de données constituant une « norme d'or », comme les résultats d'analyses de laboratoire ou des données cliniques sur les diagnostics, aideront à comprendre les avantages et les limites de l'emploi de données administratives à des fins de surveillance des maladies chroniques. Plus particulièrement, les études utilisant une norme d'or peuvent permettre d'estimer la sensibilité et la spécificité des définitions de cas fondées sur des données administratives. Toutefois, étant donné qu'il n'existe aucune norme d'or sans biais pour certaines maladies chroniques, comme l'arthrite et le syndrome du côlon irritable, d'autres recherches sont aussi nécessaires au sujet des techniques de validation en présence d'erreurs de mesure⁴⁴⁻⁴⁵.

Remerciements

Cette recherche a été financée par Santé et mode de vie sain Manitoba et par une subvention des Instituts de recherche en santé du Canada (n° PPR79240). Les auteurs sont redevables à Santé et mode de vie sain Manitoba pour avoir fourni des données dans le cadre du projet n° 2004/05-01. Les résultats et les conclusions sont ceux des auteurs; ils ne sont pas assujettis à la sanction officielle de Santé et mode de vie sain Manitoba ni n'ont reçu une telle sanction. Les auteurs n'ont pas d'intérêts divergents à déclarer.

Références

1. Thacker SB, Stroup DF, Rothenberg RB, Brownson RC. Public health surveillance for chronic conditions: a scientific basis for decisions. *Stat Med.* 1995;14:629-641.
2. Cricelli C, Mazzaglia G, Samani F, Marchi M, Sabatini A, Nardi R, Ventriglia G, Caputi AP. Prevalence estimates for chronic diseases in Italy: exploring the differences between self-report and primary care databases. *J Public Health Med.* 2003;25:254-257.
3. Powell KE, Diseker RA, Presley RJ, Tolsma D, Harris S, Mertz KJ, Viel K, Conn DL, McClellan W. Administrative data as a tool for arthritis surveillance: estimating prevalence and utilization of services. *J Public Health Manag Pract.* 2003;9:291-298.
4. Umphrey GJ, Kendall O, MacNeill IB. Assessing the surveillance capability of Canada's national health surveys. *Chronic Dis Canada.* 2001;22:50-56.
5. Mayo NE, Chockalingam A, Reeder BA, Phillips S. Surveillance for stroke in Canada. *Health Rep.* 1994;6:62-72.
6. Saydah SH, Geiss LS, Tierney E, Benjamin SM, Engelgau M, Brancati F. Review of the performance of methods to identify diabetes cases among vital statistics, administrative and survey data. *AEP.* 2004;14:507-516.
7. Maio V, Yuen E, Rabinowitz C, Louis D, Jimbo M, Donatini A, Mall S, Taroni F. Using pharmacy data to identify those with chronic conditions in Emilia Romagna, Italy. *J Health Serv Res Policy.* 2005;10:232-238.
8. Mahonen M, Salomaa V, Brommels M, Molarius A, Miettinen H, Pyorala K, Tuomilehto J, Arstila M, Kaarsalo E, Ketonen M, Kuulasmaa K, Lehto S, Mustaniemi H, Niemela M, Palomaki P, Torppa J, Vuorenmaa T. The validity of hospital discharge register data on coronary heart disease in Finland. *Eur J Epidemiol.* 1997;13:403-415.
9. Robinson JR, Young TK, Roos LL, Gelskey DE. Estimating the burden of disease. Comparing administrative data and self-reports. *Med Care.* 1997;35:932-947.
10. Rector TS, Wickstrom SL, Shah M, Greenlee NT, Rheault P, Rogowski J, Freedman V, Adams J, Escarce JJ. Specificity and sensitivity of claims-based algorithms for identifying members of Medicare plus Choice health plans that have chronic medical conditions. *Health Serv Res.* 2004;39:1839-1861.
11. Hux JE, Ivis F, Flintoft V, Bica A. Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes Care.* 2002;25:512-516.
12. Virnig BA, McBean M. Administrative data for public health surveillance and planning. *Annu Rev Public Health.* 2001;22:213-230.
13. Watson DE, Katz A, Reid RJ, Bogdanovic B, Roos NP, Heppner P. Family physician workloads and access to care in Winnipeg: 1991 to 2001. *CMAJ.* 2004;171:339-342.
14. Organisation mondiale de la Santé. WHO Collaborating Centre for Drug Statistics Methodology: ATC classification index with DDDs and Guidelines for ATC classification and DDD assignment, Oslo, Norvège, Norwegian Institute of Public Health, 2006.
15. Kozyrskyj AL, Mustard CA. Validation of an electronic, population-based prescription database. *Ann Pharmacother.* 1998;32:1152-57.
16. Statistique Canada. L'Enquête sur la santé dans les collectivités canadiennes (ESCC) – Cycle 1.1. Ottawa, Statistique Canada, 2005. URL : http://www.statcan.ca/francais/concepts/hs/index_f.htm.
17. Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. *J Clin Epidemiol.* 2004;57:131-141.
18. Kozyrskyj AL, Mustard CA, Becker AB. Identifying children with persistent asthma from health care administrative records. *Can Respir J.* 2004;11:141-145.
19. Huzel L, Roos LL, Anthonisen NR, Manfreda J. Diagnosing asthma: the fit between survey and administrative database. *Can Respir J.* 2002;9:407-412.
20. Macy E, Schatz M, Gibbons C, Zeiger R. The prevalence of reversible airflow obstruction and/or methacholine hyperreactivity in random adult asthma patients identified by administrative data. *J Asthma.* 2005;42:213-220.
21. Morgan CL, Currie CJ, Stott NC, Smithers M, Butler CC, Peters JR. Estimating the prevalence of diagnosed diabetes in a health district of Wales: the importance of adjustment for death and migration. *Diabet Med.* 2000;17:141-145.
22. Muhajarine N, Mustard C, Roos LL, Young TK, Gelskey DE. Comparison of survey and physician claims data for detecting hypertension. *J Clin Epidemiol.* 1997;50:711-718.
23. Leibson CL, Maessens JM, Brown RD, Whisnant JP. Accuracy of hospital discharge abstracts for identifying stroke. *Stroke.* 1994;25:2348-2355.
24. Tirschwell DL, Longstreth WT. Validating administrative data in stroke research. *Stroke.* 2002;33:2465-2470.
25. Rawson NS, Malcolm E. Validity of the recording of ischaemic heart disease and chronic obstructive pulmonary disease in the Saskatchewan Health care data files. *Stat Med.* 1995;14:2627-2643.
26. Shah BR, Hux JE, Zinman B. Increasing rates of ischemic heart disease in the native population of Ontario, Canada. *Arch Intern Med.* 2000;160:1862-1866.
27. Bullano MF, Kamat S, Willey VJ, Barlas S, Watson DJ, Brenneman SK. Agreement between administrative claims and the medical record in identifying persons with a diagnosis of hypertension. *Med Care.* 2006;44:486-490.
28. Lix LM, Yogendran M, Burchill C, McKeen N, Metge C, Moore D, Bond R. Defining and validating chronic diseases: An administrative data approach, University of Manitoba, Manitoba Centre for Health Policy, 2006.
29. Altman DG. Practical Statistics for Medical Research, London, Chapman & Hall, 1991.

30. Donner A, Shoukri MM, Klar N, Bartfay E. Testing the equality of two dependent kappa statistics. *Stat Med*. 2000;19:373-387.
31. Okura Y, Urban LH, Mahoney DW, Jacobsen SJ, Rodeheffer RJ. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *J Clin Epidemiol*. 2004;57:1096-1103.
32. Borzecki AM, Wong AT, Hickey EC, Ash AS, Berlowitz DR. Identifying hypertension-related comorbidities from administrative data: what's the optimal approach? *Am J Med Qual*. 2004;19:201-206.
33. Fondation des maladies du cœur du Canada. Le nouveau visage des maladies cardio-vasculaires et des accidents vasculaires cérébraux au Canada 2000. Ottawa, Santé Canada, 2000.
34. Santé Canada. Le diabète au Canada, Deuxième édition, Ottawa, Santé Canada, 2002.
35. Wolff HK, Andreou P, Bata IR, Comeau DG, Gregor RD, Gephart G, MacLean DR, Sketris I. Trends in the prevalence and treatment of hypertension in Halifax county from 1985-1995. *CMAJ*. 1999;161:699-704.
36. Harrold LR, Yood RA, Andrade SE, Reed JI, Cernieus J, Straus W, Weeks M, Lewis B, Gurwitz JH. Evaluating the predictive value of osteoarthritis diagnoses in an administrative database. *Arthritis Rheum*. 2000; 43:1881-1885.
37. Kriegsman DM, Penninx BW, van Eijk JT, Boeke AJ, Deeq DJ. Self-reports and general practitioner information on the presence of chronic diseases in community dwelling elderly. *J Clin Epidemiol*. 1996;49:1407-1417.
38. Haapanen N, Miiunpalo S, Pasanen M, Oja P, Vuori I. Agreement between questionnaire data and medical records of chronic diseases in middle-aged and elderly Finnish men and women. *Am J Epidemiol*. 1997;145:762-769.
39. Simpson CF, Boyd CM, Carlson MC, Griswold ME, Guralnik JM, Fried LP. Agreement between self-report of disease diagnoses and medical record validation in disabled older women: factors that modify agreement. *J Am Geriatr Soc*. 2004; 52:123-127.
40. Verbrugge LM, Lepkowski JM, Imanaka Y. Comorbidity and its impact on disability. *Milbank Q*. 1989;67:450-484.
41. Benesch C, Witter DM, Wilder AL, Duncan PW, Samsa GP, Matchar DB. Inaccuracy of the International Classification of Disease (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease. *Neurology*. 1997;49:660-664.
42. Hirsch S, Shapiro JL, Turega MA, Frank TL, Niven RM, Frank PI. Using a neural network to screen a population for asthma. *Ann Epidemiol*. 2001;11:369-376.
43. Shanker M, Hu MY, Hung MS. Estimating probabilities of diabetes mellitus using neural networks. *SAR QSAR Environ Res*. 2000;11:133-147.
44. Ladouceur M, Rahme E, Pineau CA, Joseph L. Robustness of prevalence estimates derived from misclassified data from administrative databases. *Biometrics* 2007; 63(1):272-279.
45. Cole SR, Haitao C, Greenland S. Multiple-imputation for measurement-error correction. *Int J Epidemiol*. 2006;35:1074-1081.