
Utilisation des données administratives pour comprendre la distribution géographique de la détermination des cas

N. Yiannakoulias, Ph. D. (1); D. P. Schopflocher, Ph. D. (2); L. W. Svenson, B.Sc. (3)

Résumé

Nous avons examiné la variabilité géographique de l'information générée à partir de différentes définitions de cas d'asthme infantile, provenant des données administratives sur la santé utilisées en Alberta, au Canada. Notre objectif était de déterminer si les analyses fondées sur différents algorithmes de détermination des cas permettent ou non de définir les groupes géographiques dans la même région que celle sur laquelle porte l'étude. Notre groupe d'étude reposait sur une cohorte fermée d'enfants asthmatiques nés en 1988. Nous avons utilisé une statistique sur l'analyse spatiale pour établir les variations de l'emplacement approximatif des groupes géographiques liés à l'asthme selon différentes définitions de cas. Nos résultats indiquent que l'algorithme de détermination des cas et la source de données n'ont pas une très grande incidence sur les profils géographiques. Par exemple, les asthmatiques identifiés à partir des données sur les réclamations au titre de frais médicaux ont démontré un regroupement semblable à celui des asthmatiques définis à partir des données sur l'hospitalisation et provenant du service d'urgence. Cependant, il faut bien prendre en considération les estimations de la prévalence et de l'incidence et les valider par rapport aux autres sources de données.

Mots clés : *analyse spatiale, données administratives sur la santé, asthme, surveillance de la santé publique*

Introduction

L'augmentation de l'accessibilité aux données électroniques sur la santé et des capacités informatiques permettant de stocker et d'analyser ces données offrent de nouvelles possibilités de recherche sur une vaste gamme de résultats en matière de santé. Il existe un ensemble croissant de données probantes suggérant que les données administratives sur la santé peuvent constituer des ressources d'une grande importance pour la recherche sur les maladies chroniques et leur surveillance¹⁻⁴. L'une des plus importantes contributions des données administratives sur la santé a été de fournir de l'information qui facilite l'analyse de populations entières vivant

dans de grandes régions géographiques. Au Canada, cela comprend souvent une analyse à l'échelle provinciale (c'est-à-dire la comparaison des régions au sein d'une province), qui peut s'avérer particulièrement utile pour faire le lien entre les systèmes d'information sur la santé et le processus décisionnel en matière de santé publique. Les variations géographiques entre les régions d'une province peuvent indiquer des différences dans l'épidémiologie, les caractéristiques des populations, les services offerts, l'exposition aux dangers environnementaux, la pratique diagnostique et un grand nombre d'autres facteurs.

L'un des défis à relever quant à l'utilisation des données administratives sur la santé en recherche et en surveillance est que les différentes méthodes de détermination des cas peuvent confondre les différences entre les groupes, particulièrement si la façon dont ces données sont recueillies ou générées varie. Par exemple, il peut y avoir des différences rurales ou urbaines quant à l'efficacité de certaines données pour déterminer les accidents vasculaires cérébraux⁵⁻⁷. Ces différences pourraient être liées à la prestation de soins (surtout dans les régions rurales, où les établissements de soins de courte durée assument généralement le rôle de centres de soins primaires), à la disponibilité des ressources, aux variations géographiques en ce qui concerne la spécialisation des médecins et à de nombreux autres facteurs. La solution dont il est le plus souvent question pour régler ces problèmes liés aux données est la création d'algorithmes de détermination des cas qui combinent une multitude de sources de données administratives sur la santé ou de dossiers dans une source unique^{3,8}. Grâce à ce concept, plutôt que d'utiliser un seul dossier provenant d'une seule source pour déterminer un cas, de multiples dossiers et de multiples sources de données sont regroupés. L'intégration de nombreuses sources de données peut améliorer la détermination des cas faite à partir de mesures traditionnelles (comme la sensibilité et la spécificité), mais aussi améliorer l'uniformité géographique pour le processus de choix des cas, en garantissant que les singularités locales ou régionales associées à une seule source de données n'ont pas une trop

Coordonnées des auteurs

1 School of Geography and Earth Sciences, McMaster University, Hamilton (Ontario)

2 Faculty of Nursing, University of Alberta, Edmonton (Alberta)

3 Surveillance and Environmental Health, Alberta Health and Wellness, Edmonton (Alberta)

Correspondance : Nikolaos Yiannakoulias, School of Geography and Earth Sciences, McMaster University, 1280 Main Street West, Hamilton (Ontario) L8S 4K1; tél. : 905-525-9140, poste 20117; téléc. : 905-546-0463; courriel : yiannan@mcmaster.ca

grande incidence sur les comparaisons géographiques.

La plupart des algorithmes de définition de cas utilisent des données administratives qui ont été validées par rapport aux examens des dossiers médicaux^{9,10} ou aux réponses à un sondage^{4,11}. Bien que les premiers constituent une importante référence pour évaluer un algorithme de définition de cas, ils ne permettent pas de bien caractériser les façons dont les propriétés d'un algorithme peuvent varier entre les régions. Cela se révèle particulièrement vrai lorsque la validation repose sur un site précis de l'étude, plutôt que sur un large échantillon de sites dans une vaste région géographique. Bien que la comparaison des données administratives avec les données d'un sondage puisse être informative, le processus porte sur deux sources de données faillibles et sur aucun « exemple idéal ».

Dans la présente étude, nous avons examiné la variabilité géographique de l'information générée par différentes définitions de cas d'asthme infantile, provenant d'un système de données administratives sur la santé. Notre objectif était d'observer la façon dont l'utilisation de différents algorithmes de détermination des cas peut influencer sur l'apparition de profils et la variabilité sur une carte. Plutôt que de comparer les définitions de cas à un exemple idéal ou à un sondage de répondants, nous avons comparé la façon dont l'information analytique varie dans un ensemble d'algorithmes de définition de cas. Plus précisément, notre approche était de rechercher des groupes géographiques d'enfants représentant des cas d'asthme. Si les analyses basées sur différents algorithmes de définition de cas permettent de déterminer les groupes dans une région de la zone d'étude, l'information analytique sur les profils géographiques est fiable, même si les critères de détermination des cas diffèrent. D'un autre côté, d'importantes variations dans les analyses résultantes pourraient suggérer que les algorithmes de définition de cas doivent être plus précis sur le plan géographique et être conçus de façon à prendre en considération les

différences locales ou régionales quant à l'utilisation des services, la conformité aux normes en matière de données ou à d'autres facteurs.

L'information provenant de la présente étude contribuera à établir des définitions de cas pour l'asthme infantile fondées sur des données administratives, en déterminant dans quelle mesure les algorithmes doivent être propres à une région géographique. Cependant, à plus grande échelle, nous présentons un cadre de travail simple visant à évaluer la fiabilité géographique des données administratives sur la santé pour une vaste gamme de maladies chroniques.

Méthodes

Données

Notre zone d'étude, la province de l'Alberta au Canada, convient parfaitement à l'analyse, en raison de l'accès à de multiples sources de données administratives sur la santé et de l'existence d'un registre de la population qui peut servir à déterminer le lieu de résidence au fil du temps. À l'instar d'autres provinces canadiennes, l'Alberta possède un régime d'assurance-maladie à payeur unique financé par l'État, qui couvre la plupart des services médicaux. Les habitants de la province qui choisissent de participer à ce régime provincial doivent donc souscrire au régime d'assurance-maladie de l'Alberta (Alberta Health Care Insurance Plan). On leur fournit ensuite un identifiant unique valide pour toute leur vie et qui permet de regrouper les diverses sources de données sur la santé. Plus de 99 % de la population de la province est inscrite à ce régime.

Alberta Health and Wellness a conçu un système pour générer des profils de données longitudinales sur la santé à l'appui d'un vaste éventail d'activités de surveillance de la santé publique. Ces profils sont liés à quatre bases de données : les demandes de remboursement pour des services médicaux payés à l'acte (demandes de remboursement), le système de classification des soins ambulatoires qui contient les données sur les admissions au service

d'urgence (urgences), un système sur les services hospitaliers aux patients hospitalisés (patients hospitalisés) et le système de registre de l'Alberta Health Care Insurance. Les données provenant de ces systèmes sont associées à un identifiant numérique unique, puis classées dans le dénombrement annuel des services afférents à une maladie donnée (définie selon les codes de la CIM-9* ou de la CIM-10†) au fil du temps. Lorsque le système est jumelé à celui du registre de la population (qui contient l'information sur le sexe, l'âge et le lieu de résidence), il fournit une méthode simple pour observer les changements des prévisions démographiques de l'incidence ou de la prévalence selon différents algorithmes de définition de cas.

Dans le cadre de l'étude, nous avons utilisé un profil de données pour les services liés à l'asthme (code 493 de la CIM-9 ou code J45 de la CIM-10 dans le champ sur le diagnostic primaire) comme source de données primaires. La période de notre étude s'échelonne entre les années civiles 1998 et 2004. Afin de contrôler les variations en matière de pratique de la prise en charge de l'asthme survenues au cours des dernières années, notre étude se limite aux personnes nées la même année. Bien que le système de demandes de remboursement contienne des dossiers remontant au début des années 80, les dossiers électroniques du système sur les patients hospitalisés remontent à 1993 et ceux du système sur les soins ambulatoires aux urgences à 1998. En 1998, il y avait 44 651 enfants nés en 1988 habitant dans la province et inscrits au régime d'assurance-maladie. Nous avons restreint la population étudiée aux enfants nés en 1988 qui ont résidé de façon continue dans la province entre 1998 et 2004. Ainsi, les sujets étaient suffisamment âgés pour être diagnostiqués en tant qu'asthmatiques, mais avaient moins de 18 ans au cours de la période d'étude. Cela a donné une cohorte fermée de 38 905 enfants. Parmi eux, 8 965 avaient au moins un service médical enregistré à leur nom dans l'une des trois bases de données sur la santé et pour lequel l'asthme était le

* Classification statistique internationale des maladies et des problèmes de santé connexes, 9^e révision.

† Classification statistique internationale des maladies et des problèmes de santé connexes, 10^e révision.

diagnostic primaire. Nous avons supposé que deux services liés à l'asthme ou plus (qui n'ont pas été fournis le même jour) au cours de la période de l'étude représentent un seuil minimal raisonnable pour déterminer qu'un enfant est asthmatique. C'est ce que nous avons appelé le groupe d'asthmatiques « de référence ». Selon cette définition, 5 110 enfants de la cohorte étaient considérés comme des asthmatiques de référence en 2004, ce qui équivaut à une prévalence d'environ 13 % pour l'asthme dans la cohorte.

Nous avons utilisé le système de profil de données pour définir les cas selon différents algorithmes de définition de cas (tableau 1). Nous faisons référence à ces algorithmes en tant que « définitions de cas d'essai ». Les définitions varient selon le nombre et le type de services dans chaque système de données. Ces définitions ont été choisies afin d'accentuer les différences entre les algorithmes de détermination des cas : pour qualifier une personne en tant que cas, la plupart des algorithmes exigeraient moins de contacts (avec le système de santé) que les définitions présentées dans le présent document. L'interprétation de nos résultats doit donc être vue comme un exemple extrême de la façon dont différents algorithmes de détermination des cas peuvent présenter de l'information différente quant à la distribution géographique de la maladie. Nous avons évité le surdénombrement des contacts au cours d'épisodes précis en ne comptant qu'un seul de ces contacts dans une journée. Lorsqu'il y avait plusieurs contacts entre les systèmes de données pour une journée en particulier, nous avons privilégié les dossiers sur les patients hospitalisés aux dossiers provenant du service d'urgence, et privilégié ces derniers à ceux de demandes de remboursement.

En utilisant les codes postaux résidentiels de 2004 associés au profil de données, toutes les données furent regroupées au niveau municipal. Les municipalités (grandes villes, villes et villages) furent limitées à celles qui comprenaient au moins dix membres de la cohorte, pour un total de 294 municipalités. Les municipalités plus petites ont été jumelées aux plus grandes pour s'assurer qu'aucun membre

de la cohorte n'était laissé de côté dans l'analyse. Les enfants vivant à l'extérieur des municipalités (p. ex. entreprises agricoles) ont été assignés à la municipalité liée à leur code postal résidentiel (généralement celle où le courrier est ramassé). Par conséquent, certains enfants vivant dans des régions rurales peuvent avoir été jumelés à une municipalité pas nécessairement la plus proche de leur lieu de résidence ou qui n'est pas dans le secteur où ils reçoivent la majorité de leurs services médicaux. Cependant, toute erreur de géoréférencement est commune au numérateur et au dénominateur, ce qui ne devrait pas avoir d'incidence sur nos résultats.

Analyse

À des fins de visualisation, nous avons établi la carte de la prévalence des asthmatiques de référence en Alberta. Plutôt que de dresser la carte des taux de prévalence bruts, qui seraient grandement touchés par les nombres peu élevés, nous avons utilisé une approche modélisée afin d'estimer le risque relatif de façon à gérer la variation stochastique des données. Nous avons donc fait appel au modèle de Poisson pour faire des prévisions en fonction du nombre moyen d'asthmatiques de référence et avons inclus la population d'enfants comme facteur de compensation pour contrôler les variations de la distribution géographique de la cohorte. Le modèle comprend également un effet d'interception aléatoire pour chaque municipalité et des estimations des paramètres spatiaux afin d'éliminer les variations locales de la prévalence. En termes simples, cette méthode fait une moyenne entre les observations géographiques avoisinantes d'une façon conceptuelle semblable à la moyenne sur une fenêtre mobile à une dimension. On appelle cette méthode de modélisation une approche de modèle linéaire mixte généralisé (GLMM)¹². Nous avons utilisé la procédure PROC GLIMMIX de SAS pour résoudre le modèle¹³. Nous avons dressé la carte du taux attendu de morbidité de référence pour l'asthme dans la cohorte au niveau municipal. La carte obtenue est composée de polygones ombragés selon le risque relatif, chaque polygone représentant la région entourant la municipalité (figure 1).

Notre analyse primaire comprenait des vérifications explicites des groupes géographiques selon les différents algorithmes de définitions de cas d'essai, et nous avons utilisé la statistique sur l'analyse spatiale¹⁴ pour déterminer ces groupes géographiques. L'approche d'analyse spatiale utilise une fenêtre mobile et à dimension variable (de forme circulaire dans la présente étude) pour chercher un grand nombre de groupes potentiels. La méthode permet ensuite de définir le groupe, contenu dans la fenêtre, qui infirmera le plus probablement l'hypothèse nulle de risque constant. Ce groupe représente le groupe de cas « le plus probable ». La signification statistique d'un groupe le plus probable fut évaluée avec la simulation de Monte-Carlo. En n'évaluant la signification que du seul groupe rejetant le plus probablement l'hypothèse nulle de risque constant, la méthode contourne les problèmes d'analyses multiples, qui sont courants avec d'autres méthodes de détection des groupes locaux¹⁴.

Dans notre analyse, nous avons examiné deux hypothèses générales pour chacune des six définitions de cas d'essai d'asthme décrites dans le tableau 1. La première hypothèse était que la distribution spatiale des asthmatiques de l'ensemble des définitions (y compris celle des asthmatiques de référence) différait de celle de la population de la cohorte d'étude. Cela correspond à l'hypothèse nulle de risque constant, à savoir qu'il n'y a aucun sous-ensemble de municipalités regroupées géographiquement où le risque d'asthme est excessivement élevé. C'est ce que nous appelons *l'hypothèse nulle de risque constant*. Notre variable à tester était le ratio de probabilité du modèle de Poisson¹⁴ :

$$\left(\frac{c_i}{e(c_i)} \right) c_i \left(\frac{C - c_i}{C - e(c_i)} \right) C - c_i$$

où C est le nombre total de cas définis comme asthmatiques selon l'une des définitions, c_i le nombre de ces cas présents

dans la municipalité i et $e(c_i)$ le nombre de tels cas prévus dans la municipalité i . Le calcul de $e(c_i)$ est le suivant :

$$e(c_i) = m_i g.$$

Pour toutes les définitions de cas, m_i est le nombre d'enfants de la cohorte d'étude qui habitent dans la municipalité i , et g le taux général d'asthmatiques pour une définition en particulier dans la cohorte d'étude.

Les résultats de l'analyse ci-dessus indiquaient si nous devions rejeter ou non l'hypothèse nulle de risque constant et approximativement où se trouvaient les groupes d'asthmatiques de la cohorte d'étude pour les différentes définitions de cas d'essai. Pour déterminer de façon plus explicite la variation géographique des définitions de cas d'essai, nous avons également déterminé si la distribution géographique des asthmatiques, selon chaque définition de cas d'essai, différait de celle des asthmatiques de référence. Comme précédemment, nous avons utilisé le ratio de probabilité du modèle de Poisson, mais cette fois :

$$e(c_i) = a_i h_i,$$

où a_i est le nombre d'enfants asthmatiques selon une définition de cas précise et qui résident dans la municipalité i , et h_i la proportion d'asthmatiques selon cette définition parmi la population d'asthmatiques de référence. Dans ce cas, le rejet de l'hypothèse nulle pour une définition de cas d'essai indique que la distribution géographique des asthmatiques définis par cet algorithme de définition de cas est la même que celle des asthmatiques de référence. Cela correspond à un essai d'hypothèse nulle pour une définition de cas constante, soit *l'hypothèse nulle de définition de cas constante*.

Nous avons utilisé la version 6.1 de SaTScan pour chercher des groupes¹⁵. La taille des groupes était limitée à 50 % ou moins de la population de l'Alberta et les groupes étaient de forme circulaire. Dans tous les cas, un seuil de signification de 0,05 fut utilisé pour évaluer l'existence éventuelle d'un groupe le plus probable par

rapport aux hypothèses nulles de risque constant et de définition de cas constante.

Résultats

Selon la définition de cas d'essai A, la prévalence s'élevait à 4,4 %, inférieure à la moitié de celle du groupe de référence (tableau 2). À l'autre extrême, la définition de cas d'essai F évaluait la prévalence à moins de 0,2 %. Pour chacune des définitions de cas d'essai, nous avons également calculé le nombre moyen de services que chaque enfant a reçu, par type de service dans le système de la santé. Dans toutes les définitions de cas d'essai, les enfants semblaient avoir fréquemment des contacts avec les services d'urgence concernant l'asthme, comparativement à la cohorte dans son ensemble. Les enfants de toutes les définitions de cas d'essai semblaient davantage fréquenter les établissements hospitaliers (pour une raison quelconque) que les asthmatiques de référence et les non-asthmatiques de la cohorte.

La figure 1 illustre la distribution géographique des asthmatiques de référence de la cohorte, selon le modèle. Le taux d'asthme varie parmi les asthmatiques de référence; et le risque relatif le plus élevé s'observe à Calgary, où il dépasse de 22 % la moyenne provinciale. Le risque relatif de la région d'Edmonton, quant à lui, est très près de la moyenne provinciale (il lui est supérieur de 0,3 %). Finalement, les zones rurales du centre et du nord de l'Alberta ont la prévalence d'asthme la plus faible de la cohorte.

Les groupes statistiquement significatifs d'après les deux hypothèses nulles sont délimités aux figures 2 et 3. En fonction de notre hypothèse nulle de risque constant, chaque définition de cas d'essai d'asthme, à l'exception de B, fut associée au groupe statistiquement significatif le plus probable. Le groupe B, même s'il n'est pas statistiquement significatif, a été circonscrit dans une région semblable à celle du groupe D. Le risque relatif lié à chaque groupe est relativement faible; dans la plupart des cas, la population à l'étude qui habite dans le groupe était 25 % plus susceptible d'être atteinte d'asthme que la population à l'étude vivant à l'extérieur

du groupe. La définition de cas F faisait exception : le risque relatif d'asthme dans le groupe était supérieur au double du risque hors du groupe. Tous les groupes délimités représentaient les régions où la probabilité de rejeter l'hypothèse nulle de risque constant était la plus élevée pour chacune des définitions de cas.

Selon notre hypothèse nulle de définition de cas constante, seule la définition F atteignait un niveau de signification statistique. Elle constitue la définition la plus rigoureuse de toutes les définitions de cas d'essai, et le nombre total de cas dans ce groupe était minime. Pour cette définition, seulement 52 cas du nombre total de cas ont été relevés à l'intérieur de ce groupe. Les enfants de ce groupe couraient 87 % plus de risques d'être des cas (conformément à la définition F) que les autres.

Analyse

Selon l'hypothèse nulle de risque constant, on dénombrait peu de différences manifestes dans la répartition des groupes entre les définitions de cas. Tous les groupes, sauf un, se situaient au sud-ouest de la province, même si leur étendue géographique variait considérablement. Par exemple, les groupes A et C étaient plus petits que les autres. Cette similarité apparente se fondait sur l'évaluation qualitative d'une carte recouvrant une vaste superficie. À elle seule, cette carte ne permet pas d'établir si ces observations représentent une géographie systématiquement différente dans les définitions de cas d'essai lorsqu'on les compare les unes aux autres ou au groupe d'asthmatiques de référence. La recherche de groupes au moyen de l'hypothèse nulle de définition de cas constante offrait un test plus explicite quant à la fluctuation éventuelle de la répartition des groupes d'asthme en fonction d'un algorithme de détermination des cas particulier. Toutes ces recherches, à part une, n'ont pas réussi à rejeter l'hypothèse nulle de définition de cas constante, ce qui suggère que le profil géographique de l'asthme différait légèrement selon les différentes définitions de cas d'essai et que la détection des groupes résistait assez bien à la définition précise sélectionnée.

TABLEAU 1
Définitions de cas d'essai

Code de la définition	Définition de cas d'essai
A	Six services ou plus de n'importe quel type entre 1998 et 2004
B	Six services ou plus, y compris au moins deux admissions au service d'urgence ou hospitalisations entre 1998 et 2004
C	Douze services ou plus de n'importe quel type entre 1998 et 2004
D	Douze services ou plus, y compris au moins deux admissions au service d'urgence ou hospitalisations entre 1998 et 2004
E	Un minimum de six admissions au service d'urgence ou hospitalisations entre 1998 et 2004
F	Un minimum de douze admissions au service d'urgence ou hospitalisations entre 1998 et 2004

TABLEAU 2
Utilisation des services liés à l'asthme ou de tous les services, par définition de cas (1998-2004)

	Définition de référence	Définitions de cas d'essai d'asthme						Non-asthmatiques
		A n = 1710	B n = 685	C n = 638	D n = 390	E n = 244	F n = 71	
Pourcentage d'asthmatiques dans la cohorte	100,00	30,81	12,34	11,49	7,03	4,40	1,28	S/O
Pourcentage de la population totale de la cohorte	13,13	4,40	1,76	1,64	1,00	0,63	0,18	100,00
Services liés à l'asthme (1998-2004)								
Nombre moyen de réclamations	5,04	9,67	10,47	15,52	14,87	12,09	16,12	S/O
Nombre médian de réclamations	3	8	8	14	13	9	13	S/O
Nombre moyen d'admissions au service d'urgence	1	2,47	5,75	4,76	7,59	10,95	19,26	S/O
Nombre médian d'admissions au service d'urgence	0	1	4	2	5,5	8,5	16	S/O
Nombre moyen d'hospitalisations	0,05	0,16	0,37	0,322	0,523	0,63	1,12	S/O
Nombre médian d'hospitalisations	0	0	0	0	0	0	0	S/O
Tous les services (1998-2004)								
Nombre moyen de réclamations	43,74	49,7	49,18	56,95	55,03	50,35	60,51	29,44
Nombre médian de réclamations	36	42	41	47	46	42	46	23
Nombre moyen d'admissions au service d'urgence	10,64	12,2	17,76	15,02	19,24	24,95	32,59	7,08
Nombre médian d'admissions au service d'urgence	6	7	13	10	14	20	27	3
Nombre moyen d'hospitalisations	0,33	0,44	0,75	0,66	0,93	1,11	1,55	0,21
Nombre médian d'hospitalisations	0	0	0	0	0	1	1	0

FIGURE 1
Risque relatif estimé pour la définition des asthmatiques de référence, selon le modèle

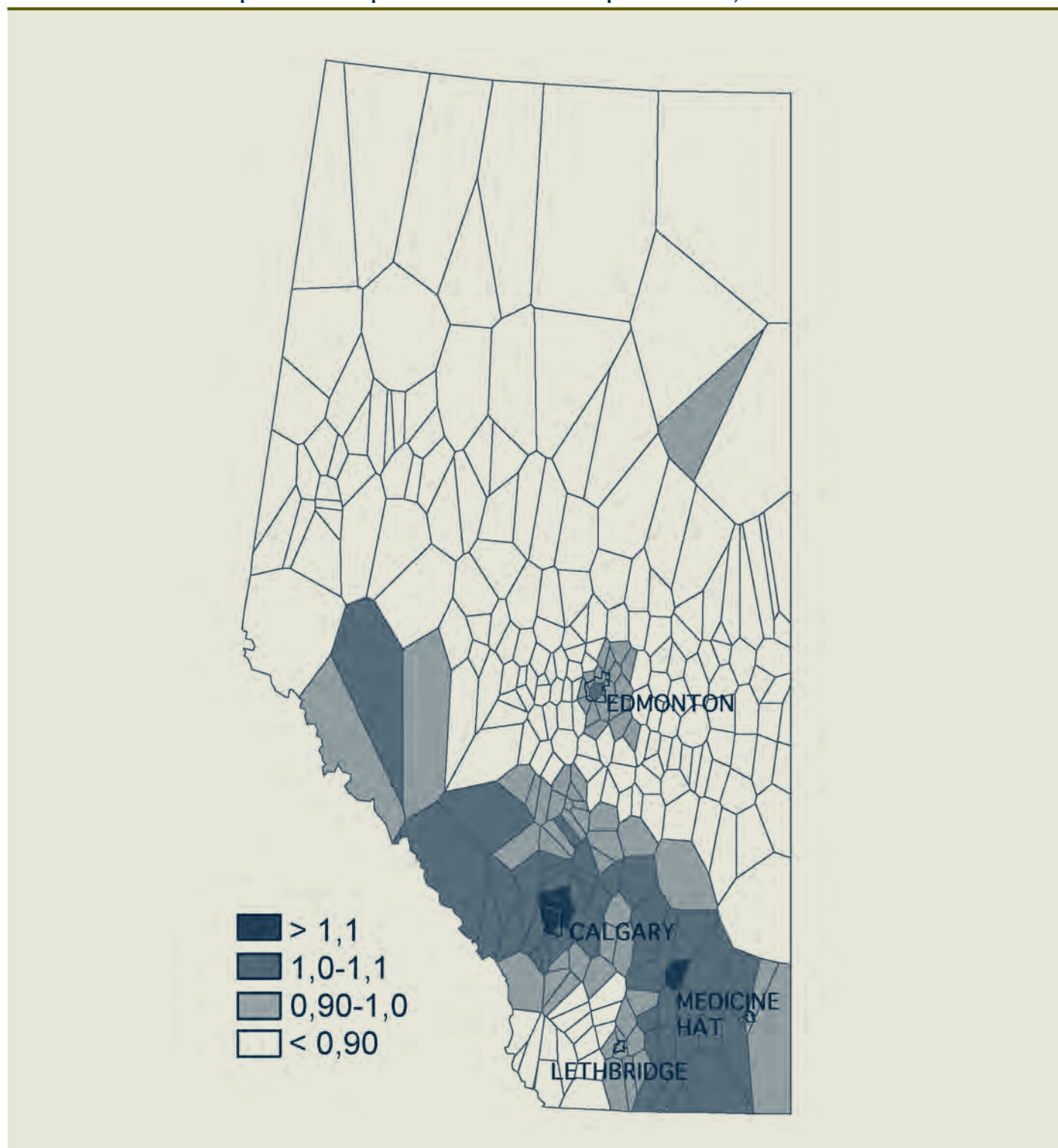


FIGURE 2
Groupes significatifs d'après l'hypothèse nulle de définition de cas constante

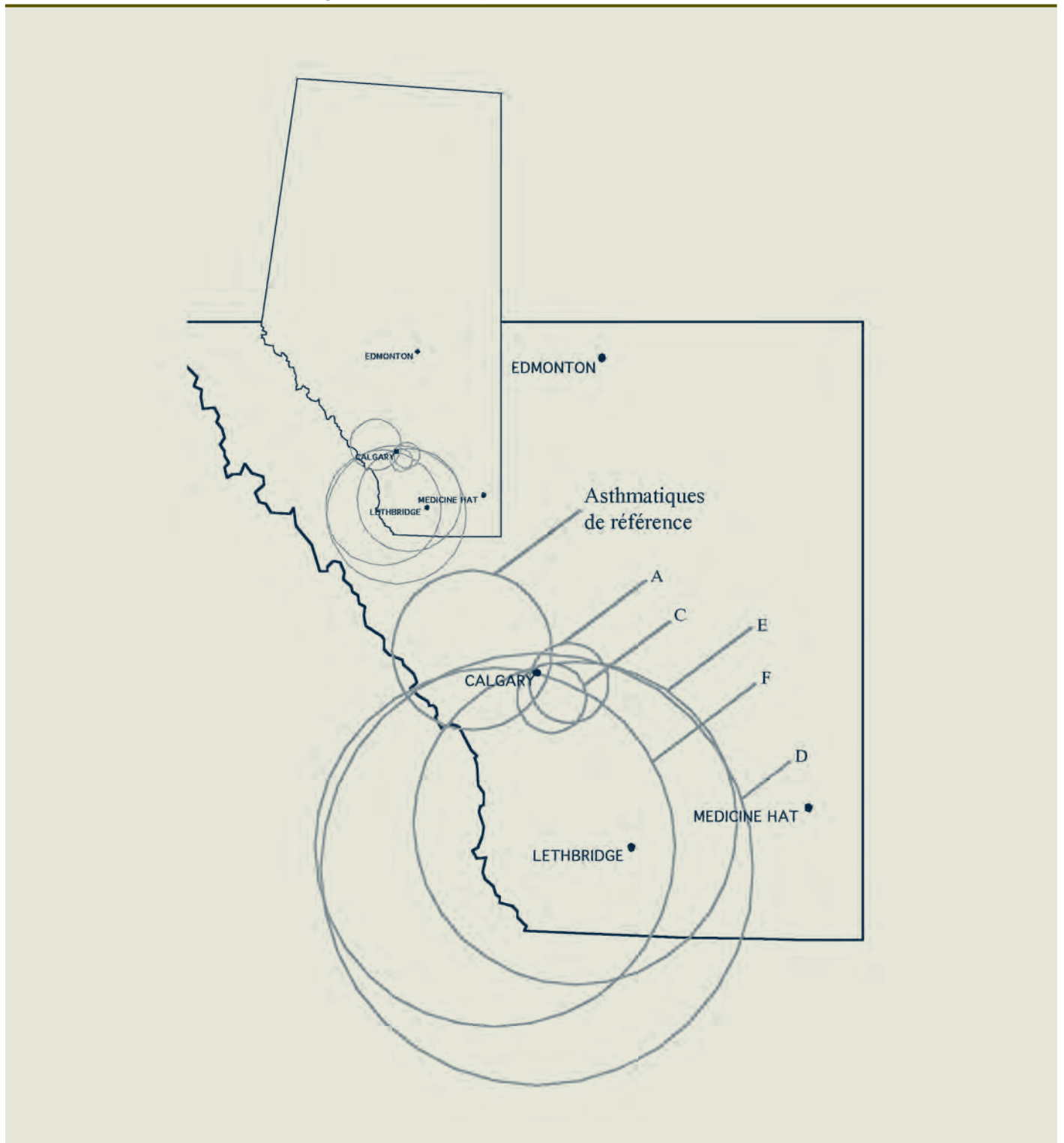
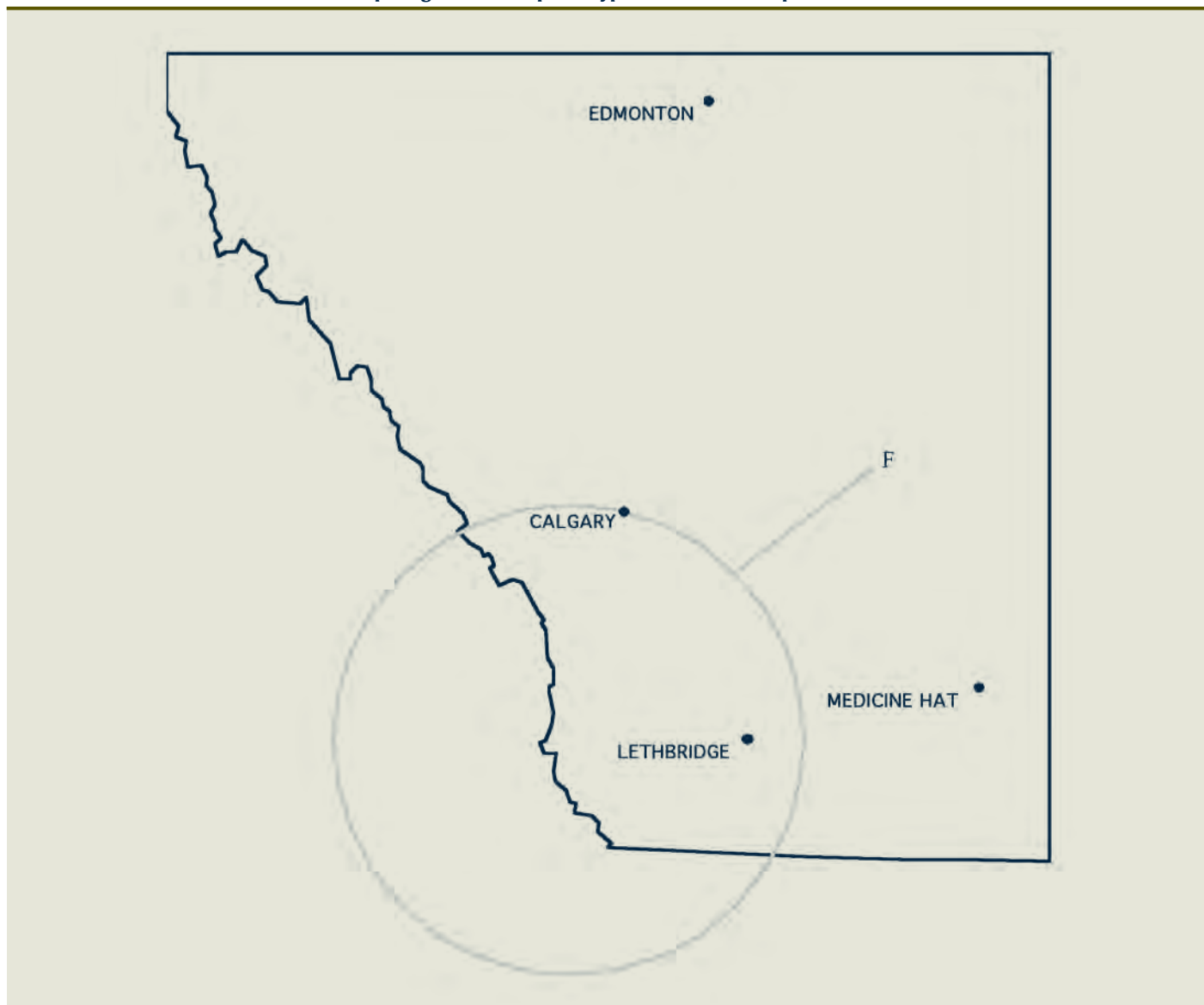


FIGURE 3
Groupes significatifs d'après l'hypothèse nulle de risque constant



En ce qui concerne l'hypothèse nulle de définition de cas constante, le seul groupe statistiquement significatif reposait sur la définition F. Ce groupe se trouvait au sud-ouest de l'Alberta, dans la même région que les groupes découverts à l'aide de l'hypothèse nulle de risque constant. Il est possible que ce groupe représente une région géographiquement concentrée représentant un usage extensif des services ou des cas graves d'asthme dans la population asthmatique de la cohorte. Le groupe pourrait aussi être lié à des changements dans la population ou dans les pratiques d'utilisation des services de santé dans la région, où une croissance démographique

notable a été observée au cours des dernières années. Somme toute, il est important d'être attentif au fait que cette définition correspond à une prévalence estimative de moins de 1 % et est beaucoup plus précise que toute autre définition de cas susceptible d'être utilisée dans des applications d'épidémiologie ou de surveillance. Par conséquent, même s'il y avait une différence statistiquement significative entre le risque relatif des régions à l'intérieur et à l'extérieur du groupe, la différence en matière de risque absolu était très faible.

Sur la base de ces observations, il semble que l'algorithme de détermination des

cas utilisé n'a pas une énorme incidence sur l'analyse géographique comparative du risque d'asthme. Plus généralement, nos résultats indiquent que les données géographiques sur les écarts relatifs de prévalence (ou risque relatif) restent constantes quelle que soit la définition de cas choisie, même si la prévalence (risque absolu) fluctue selon ces définitions de cas. Nos résultats, s'ils peuvent s'étendre à d'autres contextes ou maladies chroniques, sont donc importants pour les applications touchées par les variations géographiques de la maladie. Au moment d'utiliser des sources de données non validées à l'aide d'un exemple idéal, il serait

peut-être plus approprié de signaler les mesures géographiques du risque relatif plutôt que celles du risque absolu. Ces mesures conviennent aux applications dans lesquelles le rapport entre le risque et les facteurs de risque présente un intérêt capital. Par exemple, une analyse de corrélation écologique concernant le lien entre l'asthme et les facteurs de risque sociaux et environnementaux produira probablement des coefficients modèles similaires, indépendamment de l'algorithme de détermination des cas. Dans les applications de surveillance, où les variations du risque absolu sont généralement intéressantes, la mesure précise de la prévalence et de l'incidence est toujours importante. Nos algorithmes de détermination des cas ont généré des estimations très diverses de la prévalence de l'asthme dans la cohorte. Des estimations précises de la prévalence sont nécessaires pour comprendre le fardeau actuel de la maladie pour la population et, conséquemment, nécessitent des données validées à l'aide d'une norme de définition de cas acceptable sur les plans médical et social.

Même s'il ne s'agissait pas de l'un des objectifs de notre étude, nos résultats révèlent des profils géographiques de l'asthme infantile en Alberta qui piquent l'intérêt. Premièrement, la distribution du risque d'asthme relatif, d'après la figure 1, suggère que le risque est plus faible dans les zones rurales du centre et du nord de l'Alberta, et plus élevé dans le sud de l'Alberta, particulièrement à Calgary. Les groupes d'asthmatiques reposant sur les définitions de cas d'essai désignent également la région de Calgary comme celle associée au risque le plus élevé. Mises ensemble, ces observations reflètent une prévalence élevée d'asthme ainsi qu'une forte incidence d'admissions au service d'urgence et d'hospitalisations. L'utilisation de ces services en particulier montre un lourd fardeau de traitement dans une sous-population d'enfants de la région de Calgary. De plus, elle pourrait indiquer que l'asthme dans cette région de la province est plus grave (et, par conséquent, nécessite plus de soins d'urgence et de soins hospitaliers). Comme nous l'avons indiqué ci-dessus, ce profil pourrait s'apparenter à l'absence de soins primaires pour les

enfants dans une région où la population croît rapidement, et refléter des différences fondamentales dans l'épidémiologie de l'asthme dans cette région. Cette explication est étayée par le profil apparemment distinct d'une plus forte prévalence dans les zones rurales et urbaines du sud de l'Alberta, ce qui pourrait refléter le rôle des conditions environnementales ou météorologiques dans cette région¹⁶. Il serait justifié de mener des recherches plus approfondies pour expliquer ce profil géographique et de déterminer s'il est répandu ou non dans la population pédiatrique et la population en général.

Nous relevons deux restrictions possibles à notre étude. En premier lieu, nous avons exclu un grand nombre d'enfants ayant immigré en Alberta entre 1998 et 2004. Il est concevable que les profils d'utilisation des services de santé des enfants qui ont emménagé dans cette province diffèrent grandement de ceux des enfants que nous avons inclus dans notre groupe d'étude. En Alberta, les immigrants ont tendance à s'installer dans les zones urbaines, et tout le monde sait que l'immigration a une incidence sur le recours aux services de santé. Les nouveaux arrivants pourraient être plus enclins à fréquenter les établissements de soins de courte durée pour y recevoir des soins primaires. Cette propension aurait pu entraîner des différences systématiques dans l'efficacité des définitions de cas, par exemple en augmentant la sensibilité des définitions E et F dans les zones urbaines. En deuxième lieu, la capacité de déceler un groupe d'asthmatiques statistiquement notable dépend en partie de la prévalence et, plus particulièrement, du nombre de cas. Toutes choses égales par ailleurs, les définitions de cas dont la prévalence globale était élevée étaient plus susceptibles de produire des profils de groupe décelables. Il est possible que les groupes aient été discernés pour certaines définitions, mais pas pour d'autres, simplement parce que le nombre total de cas recensés était différent. Même si cela limitait notre plan d'étude, nous constatons que la plupart des définitions de cas se sont traduites par la détection de groupes d'asthmatiques dans approximativement la même partie de la province. En outre, le seul groupe significatif qui a été découvert à l'aide du

test de la définition de cas constante était le moins nombreux de tous les algorithmes de détermination de cas.

Conclusion

Les données administratives représentent une importante ressource pour la surveillance de la santé publique et de la recherche. Les études de validation qui comparent des définitions de cas sur la base de données administratives avec des évaluations cliniques et des enquêtes sont importantes pour comprendre les forces et les faiblesses de ces données et estimer du mieux possible la prévalence et l'incidence. L'évaluation du risque relatif selon la situation géographique, le temps, l'âge, le sexe, la classe sociale et d'autres mesures est aussi essentielle pour comprendre pleinement l'épidémiologie de la maladie. Nos résultats suggèrent que les propriétés de l'algorithme de détermination des cas ont peu d'incidence sur les comparaisons géographiques relatives de la maladie fondées sur des définitions de cas déterminées à partir des données administratives, même lorsque ces définitions proviennent de données de différentes sources. Toutefois, la prévalence et l'incidence varient considérablement selon la définition et, par conséquent, la surveillance systématique passe par un examen approfondi de l'algorithme exact qui a été utilisé.

Remerciements

Ce projet fut financé par le programme de subventions et de contributions pour la surveillance accrue des maladies chroniques dans le cadre de la Stratégie intégrée sur les modes de vie sains et les maladies chroniques, de l'Agence de la santé publique du Canada.

Références

1. Tu K, Campbell NR, Chen XL, Cauch-Dudek KJ, McAlister FA. Accuracy of administrative databases in identifying patients with hypertension. *Open Med.* 2007;1:E3-5.
2. Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. *J Clin Epidemiol.* 2004;57:131-41.

3. 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; 25:512-6.
4. Robinson JR, Young TK, Roos LL, Gelskey DE. Estimating the burden of disease: comparing administrative data and self-reports. *Med Care*. 1997;35:932-47.
5. Liu L, Reeder B, Shuaib A, Mazagri R. Validity of stroke diagnosis on hospital discharge records in Saskatchewan, Canada: implications for stroke surveillance. *Cerebrovasc Dis*. 1999;9:224-30.
6. Yiannakoulias N, Svenson LW, Hill MD, Schopflocher DP, James RC, Wielgosz AT, Noseworthy TW. Regional comparisons of in-patient and outpatient patterns of cerebrovascular disease diagnosis in the province of Alberta. *Chron Dis Can*. 2003;24:9-16.
7. Yiannakoulias N, Svenson LW, Hill MD, Schopflocher DP, Rowe BH, James RC, Wielgosz AT, Noseworthy TW. Incident cerebrovascular disease in rural and urban Alberta. *Cerebrovasc Dis*. 2004;17:72-8.
8. Tirschwell DL, Longstreth WT Jr. Validating administrative data in stroke research. *Stroke*. 2002; 33:2465-70.
9. Donahue JG, Weiss ST, Goetsch MA, Livingston JM, Greineder DK, Platt R. Assessment of asthma using automated and full-text medical records. *J Asthma*. 1997; 34:273-81.
10. Kokotailo RA, Hill MD. Coding of stroke and stroke risk factors using international classification of diseases, revisions 9 and 10. *Stroke*. 2005; 36:1776-81.
11. Huzel L, Roos LL, Anthonisen NR, Manfreda J. Diagnosing asthma: the fit between survey and administrative database. *Can Respir J*. 2002; 9:407-12.
12. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *J Am Stat Assoc*. 1993;88:9-25.
13. SAS, (Programme informatique), Version 9.1.3, Cary (NC), SAS Institute, 2006.
14. Kulldorff M. A spatial scan statistic. *Commun Stat Theory*. 1997; 26:1481-96.
15. SaTScan (Programme informatique), Version 6.1, Bethesda (MD), Kulldorff M & Information Management Services, Inc, 2006.
16. Verhoef MJ, Rose MS, Ramcharan S. The relationship between chinook conditions and women's physical and mental well-being. *Int J Biometeorol*. 1995;38:148-51.