

## Apports des bases médico-administratives pour l'épidémiologie et la surveillance : regards croisés France-Québec // Contribution of health administrative databases to epidemiology and surveillance: crossed views from France and Quebec

### Coordination scientifique // Scientific coordination

**Danielle Saint-Laurent**, Institut national de santé publique du Québec ;

**Isabelle Grémy & Hélène Therre**, Institut de veille sanitaire, Saint-Maurice, France

## > SOMMAIRE // Contents

### ÉDITORIAL // Editorial

Les bases de données médico-administratives : un nouveau souffle pour la surveillance en santé publique ?

// Health administrative databases: a new breath for public health surveillance? .....p. 2

**Denis A. Roy & Jean-Claude Desenclos**

### ARTICLE // Article

Le modèle québécois de surveillance des maladies chroniques basé sur l'utilisation des données médico-administratives jumelées

// The Quebec Chronic Disease Surveillance System using health administrative data .....p. 4

**Danielle Saint-Laurent et coll.**

*Institut national de santé publique du Québec (INSPQ), Québec, Canada*

### ARTICLE // Article

Surveillance des maladies chroniques en France : la contribution des bases de données médico-administratives

// Surveillance of chronic diseases in France: the contribution of health administrative databases .....p. 9

**Isabelle Grémy et coll.**

*Institut de veille sanitaire, Saint-Maurice, France*

### ARTICLE // Article

Développement de la surveillance de l'ostéoporose au Québec : utilisation des bases de données médico-administratives

// Development of osteoporosis surveillance in Quebec: use of health administrative databases .....p. 15

**Sonia Jean et coll.**

*Institut national de santé publique du Québec (INSPQ), Québec, Canada*

### ARTICLE // Article

Prévalence de l'hypertension artérielle au Québec : comparaison entre les données médico-administratives et les données d'enquêtes

// Prevalence of arterial hypertension in Quebec: comparison between health administrative and survey data .....p. 23

**Claudia Blais et coll.**

*Institut national de santé publique du Québec (INSPQ), Québec, Canada*

### ARTICLE // Article

L'apport des bases médico-administratives dans la surveillance des maladies respiratoires chroniques en France

// The contribution of health administrative databases in the surveillance of chronic respiratory diseases in France .....p. 30

**Marie-Christine Delmas et coll.**

*Institut de veille sanitaire, Saint-Maurice, France*

### ARTICLE // Article

Développement de la surveillance des maladies respiratoires au Québec à partir des données médico-administratives

// Development of a Quebec respiratory disease surveillance system from health administrative data .....p. 36

**Mariève Doucet et coll.**

*Institut national de santé publique du Québec (INSPQ), Québec, Canada*

(Suite page 2)

La reproduction (totale ou partielle) du BEH est soumise à l'accord préalable de l'InVS. Conformément à l'article L. 122-5 du code de la propriété intellectuelle, les courtes citations ne sont pas soumises à autorisation préalable, sous réserve que soient indiqués clairement le nom de l'auteur et la source, et qu'elles ne portent pas atteinte à l'intégrité et à l'esprit de l'oeuvre. Les atteintes au droit d'auteur attaché au BEH sont passibles d'un contentieux devant la juridiction compétente.

Retrouvez ce numéro ainsi que les archives du Bulletin épidémiologique hebdomadaire sur <http://www.invs.sante.fr/Publications-et-outils/BEH-Bulletin-epidemiologique-hebdomadaire>

**Directrice de la publication** : Dr Françoise Weber, directrice générale de l'InVS  
**Rédactrice en chef** : Judith Benrekassa, InVS, redactionBEH@invs.sante.fr  
**Rédactrice en chef adjointe** : Jocelyne Rajnchapel-Messai  
**Secrétaires de rédaction** : Laetitia Gouffé-Benadiba, Farida Mihoub  
**Comité de rédaction** : Dr Pierre-Yves Bello, Direction générale de la santé; Dr Juliette Bloch, CNSA; Cécile Brouard, InVS; Dr Sandrine Danet, ATIH; Dr Claire Fuhrman, InVS; Dr Bertrand Gagnière, Cire Ouest; Anabelle Gilg Soit Ilg, InVS; Dorothee Grange, ORS Ile-de-France; Dr Rachel Haus-Cheymol, Service de santé des Armées; Dr Nathalie Jourdan-Da Silva, InVS; Agnès Lefranc, InVS; Dr Marie-Eve Raguenaud, Cire Limousin/Poitou-Charentes; Dr Sylvie Rey, Drees; Hélène Therre, InVS; Pr Isabelle Villena, CHU Reims.  
**Institut de veille sanitaire** - Site Internet : <http://www.invs.sante.fr>  
**Préresse** : Jouve  
**ISSN** : 1953-8030

## ARTICLE // Article

Estimation de la fréquence des hémorragies obstétricales en France à partir des données transfusionnelles du PMSI

// Advantages and limitations of using national administrative data on obstetric blood transfusions to estimate the frequency of obstetric hemorrhages in France.....p. 43

**Catherine Quantin et coll.**

*Service de biostatistique et d'informatique médicale, CHU de Dijon, France*

## ARTICLE // Article

Croisement de deux bases de données médico-administratives : méthodologie et étude descriptive pour une application à la surveillance épidémiologique des cancers en France // Crossing two health administrative databases: methodology and descriptive study for application to epidemiological surveillance of cancer in France .....

p. 49

**Yao Kudjawu et coll.**

*Institut de veille sanitaire, Saint-Maurice, France*

## > ÉDITORIAL // Editorial

### LES BASES DE DONNÉES MÉDICO-ADMINISTRATIVES : UN NOUVEAU SOUFFLE POUR LA SURVEILLANCE EN SANTÉ PUBLIQUE ?

// HEALTH ADMINISTRATIVE DATABASES: A NEW BREATH FOR PUBLIC HEALTH SURVEILLANCE?

**Denis A. Roy<sup>1</sup> & Jean-Claude Desenclos<sup>2</sup>**

<sup>1</sup> *Vice-président aux affaires scientifiques, Institut national de santé publique du Québec, Québec, Canada*

<sup>2</sup> *Directeur scientifique adjoint, Institut de veille sanitaire, Saint-Maurice, France*

Ce numéro spécial du BEH vient marquer un pas important dans l'avancement des pratiques de surveillance en santé publique. En s'appuyant sur des travaux de pointe réalisés en France et au Québec dans deux contextes informationnels et socio-sanitaires distincts, ce recueil de textes établit une preuve de concept attendue, particulièrement opportune à ce moment-ci de l'évolution des méthodes et des métiers de la surveillance.

La première avancée a trait à la validité des banques de données médico-administratives (BDMA) à des fins de surveillance. Recueillies pour des fins de gestion et de paiement des prestations médicales, les données médico-administratives apportent-elles un éclairage adéquat et non-biaisé sur les objets de la surveillance ? Les analyses présentées soutiennent globalement une réponse affirmative. La couverture populationnelle des BDMA est quasi-exhaustive. Sans être parfaites, la sensibilité et la spécificité des indicateurs issus des BDMA satisfont, pour des phénomènes de santé bien choisis, les critères scientifiques en usage. Le chaînage des données permet de plus l'adoption d'une perspective longitudinale, qu'autrement seules les coûteuses analyses de cohortes peuvent permettre. L'appariement des BDMA avec d'autres bases, notamment les enquêtes ou les fichiers de données sociales et professionnelles, étendent d'autant plus la portée et la granularité des observations. Les analyses présentées suggèrent cependant que l'utilisation des BDMA pour la surveillance du cancer impose une approche rigoureuse faite d'étapes successives d'analyse, d'appariement et de validation dont la qualité dépend du site

du cancer considéré. Dans ce champ, les registres du cancer constituent l'étalon et des travaux de validation des systèmes multi-sources en référence à ce dernier sont indispensables.

Une deuxième avancée importante réside dans la capacité des BDMA à apporter un éclairage inédit à la politique de santé. Divers exemples viennent confirmer leur contribution informationnelle en appui à la décision des politiques globales de santé. Plusieurs d'entre eux illustrent comment les BDMA procurent une assise empirique pouvant éclairer trois grandes finalités du système de santé : la santé, l'expérience de soins et l'optimisation des ressources consenties. C'est par la prise en compte simultanée et la mise en tension de ces trois finalités qu'on arrivera à accroître la valeur des politiques et programmes de santé. L'intégration d'informations complémentaires est dès lors essentielle, celles sur la santé et ses déterminants, celles sur le rapport des personnes avec le système de soins et celles relatives aux coûts. Les BDMA seules ou appariées entre elles sont en mesure d'éclairer chacune de ces dimensions.

La mise en place de nouveaux mécanismes et processus liant les acteurs de l'information aux acteurs de la décision constitue une troisième avancée, attendue et essentielle. Les échanges continus avec les décideurs, sur la base d'objectifs convenus avec eux, et de données probantes de qualité en lien direct avec l'action des systèmes de santé représentent un levier puissant et une condition nécessaire pour l'amélioration continue des pratiques. Les processus décrits dans ce numéro

confirment que nous sommes en mesure d'engager les acteurs de la surveillance à faire évoluer leur pratique dans cette direction.

Mais pour réaliser le plein potentiel de ce rapprochement de la surveillance avec la prise de décision publique, il nous faudra encore mieux maîtriser les dimensions territoriales et socio-environnementales des analyses de la surveillance. Les facteurs liés aux populations, au temps et au milieu de vie sont les paramètres fondamentaux de l'épidémiologie. Les prochaines étapes du développement des outils de surveillance et de recherche issus des BDMA devront conjuguer ces trois grandes dimensions et mieux les documenter, en lien avec les contextes de la décision publique au palier régional et local. C'est ainsi que nous pourrions éclairer plus justement les acteurs de la scène régionale et locale, en appui à la décentralisation ou à la déconcentration de la décision publique.

Si les possibilités offertes par les BDMA sont prometteuses, leur exploitation est devenue un métier spécifique de l'épidémiologie, que ce soit dans une finalité de recherche, de surveillance, d'évaluation, de pharmaco-épidémiologie... Cette évolution, dont ce numéro du BEH atteste la légitimité et la pertinence, n'en correspond pas moins à un changement de paradigme dont il faut avoir conscience : « *plutôt que collecter des données de novo en fonction des objectifs, la surveillance doit s'adapter aux données disponibles* ». Les données que recèlent les BDMA sont certes exhaustives et riches, mais d'une grande complexité, et surtout ne se prêtent pas à l'usage épidémiologique sans de nombreux traitements et recodages. Elles ne sont pas collectées dans une finalité épidémiologique et sont tributaires, au-delà des phénomènes de santé qu'elles sont censées refléter, de nombreuses conditions liées, notamment, à la production et la gestion des soins. Leur

utilisation en santé publique implique donc, outre des mécanismes transparents d'accès, un partage d'expérience à organiser et un investissement conjoint des organismes de surveillance et de recherche en santé publique. La validation des indicateurs épidémiologiques qui sont produits à partir des BDMA est donc essentielle. Et, malgré leur grande utilité, les BDMA ne permettront pas de répondre à tout dans le champ de la surveillance.

En concluant, il importe de rappeler que l'exhaustivité et la nature des informations contenues dans les BDMA leur donnent un caractère potentiellement identifiant vis-à-vis des citoyens. L'utilisation de ces immenses recueils de données individuelles pour des finalités collectives ne doit donc pas remettre en cause le droit des personnes en matière de vie privée. Un récent rapport français sur l'utilisation des BDMA remis à la Ministre des Affaires sociales et de la Santé<sup>1</sup> propose des principes fondamentaux pour leur utilisation large pour le bien public tout en insistant très justement sur le fait que « *Sans confiance des citoyens dans la protection des données confidentielles, il ne peut exister de système d'information viable* ». Dans la mesure où nos pratiques professionnelles et organisationnelles nous permettront de mériter cette confiance, les BDMA rendront possibles des avancées majeures dans notre domaine. Bonne lecture. ■

#### Référence

[1] Bras PL, Loth A. Rapport sur la gouvernance et l'utilisation des données de santé. Paris: Ministère de la Santé et des Affaires sociales;2013. 128 p. <http://www.drees.sante.gouv.fr/rapport-sur-la-gouvernance-et-l-utilisation-des-donnees-de,11202.html>

#### Citer cet article

Roy DA, Desenclos JC. Les bases de données médico-administratives : un nouveau souffle pour la surveillance en la santé publique ? Bull Epidémiol Hebd. 2013;(Hors-Série):2-3.

## LE MODÈLE QUÉBÉCOIS DE SURVEILLANCE DES MALADIES CHRONIQUES BASÉ SUR L'UTILISATION DES DONNÉES MÉDICO-ADMINISTRATIVES JUMELÉES

// THE QUEBEC CHRONIC DISEASE SURVEILLANCE SYSTEM USING HEALTH ADMINISTRATIVE DATA

Danielle Saint-Laurent<sup>1</sup> (danielle.st.laurent@inspq.qc.ca), Claudia Blais<sup>1,2</sup>, Sonia Jean<sup>1,3</sup>, Caroline Sirois<sup>1,4</sup>,  
Louis Rochette<sup>1</sup>, Valérie Émond<sup>1</sup>

<sup>1</sup> Institut national de santé publique du Québec (INSPQ), Québec, Canada

<sup>2</sup> Faculté de pharmacie, Université Laval, Québec, Canada

<sup>3</sup> Faculté de médecine, Université Laval, Québec, Canada

<sup>4</sup> Université du Québec, Rimouski, Québec, Canada

Soumis le 19.07.2013 // Date of submission: 07.19.2013

### Résumé // Abstract

L'évolution rapide du contexte démographique, politique et environnemental ainsi que les changements continus de la pratique et des connaissances médicales influencent grandement le développement et l'issue des maladies et, par conséquent, la façon d'en assurer la surveillance épidémiologique. Pour répondre aux impératifs contextuels, le Système intégré de surveillance des maladies chroniques du Québec (Sismacq) s'est développé sur la base du jumelage de cinq fichiers médico-administratifs couvrant le fichier des personnes assurées et l'offre de services de santé des Québécois, depuis la consultation médicale jusqu'à l'hospitalisation, en passant par la consommation des médicaments et se terminant par le décès.

Le but de cet article est de décrire le Sismacq. Le potentiel des sources de données jumelées permet non seulement d'établir les prévalences et les incidences des maladies, mais aussi d'étudier des cohortes de personnes atteintes de maladies et d'analyser les issues de santé comme le fardeau des maladies et l'utilisation des ressources. Cette approche intégrée offre également une opportunité de comprendre le cumul des maladies et les comorbidités. Le Sismacq est basé sur trois étapes méthodologiques distinctes que sont 1) l'extraction et le jumelage des données, 2) les analyses et la production des indicateurs et 3) la diffusion.

*The evolving demographic, political and environmental context, as well as the progression of medical practice and findings, greatly influence the development and outcome of diseases and, consequently, their epidemiological surveillance. In order to respond to these contextual imperatives, the Quebec Integrated Chronic Disease Surveillance System (QICDSS) was developed based on the linkage of five health administrative datasets covering the Quebec health services offer, ranging from medical consultation to hospitalization and including prescription drug use and death.*

*The goal of this article is to describe the QICDSS. The linked administrative data bring the potential to develop disease-specific incidence and prevalence measures as well as to study cohorts of persons living with diseases and to analyze health outcomes such as the burden of disease and resource utilization. This integrated approach also offers an opportunity to understand co-occurring diseases and comorbidities. QICDSS relies on three distinct methodological steps: 1) data extraction and linkage; 2) data analysis and production of indicators; and 3) dissemination.*

**Mots-clés :** Système de surveillance, Surveillance, Maladies chroniques, Multimorbidité, Données médico-administratives, Jumelage, Québec

// **Keywords:** Surveillance system, Surveillance, Chronic diseases, Multimorbidity, Health administrative data, Linkage, Quebec

### Introduction

L'évolution rapide du contexte social, politique et environnemental, les changements continus de la pratique et des connaissances médicales et les avancées technologiques influencent le développement et les issues des maladies et, par conséquent, notre façon de concevoir la surveillance. La surveillance des maladies chroniques et son développement s'inscrivent dans ce mouvement de changement et

de transformation. À ce contexte, s'ajoute la nécessité de comprendre les liens des maladies entre elles, tout comme le cumul de ces pathologies qui touchent une grande majorité des personnes atteintes de maladies chroniques. L'ensemble de ces éléments constituent des enjeux importants à considérer dans les systèmes de surveillance que l'on entend déployer pour mesurer l'état de santé de nos populations, et des défis de taille dans leur exploitation.

Au Québec, la surveillance des maladies chroniques est confiée à l'Institut national de santé publique du Québec (INSPQ) par le ministre de la Santé et des Services sociaux. Ce mandat donne la responsabilité à l'Institut de développer la surveillance des maladies chroniques afin d'en dresser le portrait, d'en suivre l'évolution au sein de la population, d'en observer les tendances, de détecter les problèmes émergents et prioritaires et de faire les projections associées à ces maladies.

Pour répondre aux défis de la surveillance des maladies chroniques, le dispositif de surveillance sanitaire mis en place au Québec repose sur des indicateurs calculés à partir des données jumelées provenant de différents fichiers médico-administratifs. Dans un système de santé universel comme celui du Québec, les données médico-administratives constituent une source d'information pertinente pour la surveillance des maladies chroniques et de leurs déterminants. Ce dispositif permet d'avoir une approche populationnelle, couvrant 95% de la population estimée à 8 millions de personnes. Des études ont démontré que les fichiers médico-administratifs ont un niveau élevé de fiabilité et de validité, et leur jumelage offre un potentiel important d'informations, permettant d'établir notamment les prévalences et les incidences des différentes maladies<sup>1,2</sup>. Ils constituent également une réponse adéquate pour étudier les issues de santé comme le fardeau des maladies et l'utilisation des ressources en soins et services de santé. Ils permettent l'accès à des données de façon continue et systématique, ce qui donne la possibilité de constituer des cohortes de personnes atteintes d'affections spécifiques et de les suivre dans le temps. Cependant, les fichiers étant conçus pour répondre à des besoins d'ordre administratif et ayant pour but de gérer des services de santé, leur utilisation à des fins de surveillance épidémiologique nécessite des stratégies méthodologiques adaptées et des efforts scientifiques conséquents.

En 2009, le Système intégré de surveillance des maladies chroniques du Québec (Sismacq) a été créé afin de répondre aux objectifs traditionnels de surveillance par maladie, à ceux liés au cumul des maladies ainsi qu'aux nouveaux impératifs imposés par le contexte évolutif et complexe de l'univers des maladies chroniques. Il est le résultat d'une réflexion sur la faisabilité et la nécessité d'établir un système de surveillance selon un modèle opérationnel novateur, basé sur une approche intégrée des maladies et des sources de données. Fondé sur le jumelage de cinq fichiers médico-administratifs, il permet actuellement de développer la surveillance des maladies suivantes : diabète, maladies cardiovasculaires, maladies respiratoires, ostéoporose, maladies ostéo-articulaires, troubles mentaux, maladie d'Alzheimer et maladies apparentées (figure). La surveillance de certaines de ces maladies est décrite dans trois articles publiés dans ce même numéro<sup>3-5</sup>. Le dispositif vise à prendre en considération l'ensemble du *continuum* de la maladie, de l'état de bonne santé jusqu'à la fin de vie, et à évaluer l'impact de la co-occurrence des maladies et des comorbidités. Les objectifs de la surveillance

des maladies chroniques ont été considérés dans le contexte du développement de ce système. Il s'agit : de mesurer l'ampleur des maladies et leurs déterminants ; d'analyser la mortalité, la survie et les incapacités ; d'évaluer l'utilisation des services de santé ; d'identifier les groupes à risque ; d'analyser les trajectoires de soins. Dans une perspective intégrée, les objectifs de surveillance visent également à : décrire et analyser les comorbidités associées aux maladies ; caractériser les personnes atteintes de plus d'une maladie ; documenter le fardeau pharmacologique et la polypharmacie liés à la multimorbidité.

Les objectifs de cet article sont de décrire le modèle opérationnel du Sismacq et d'en discuter les forces et les faiblesses.

## Le modèle opérationnel du Sismacq

Le modèle opérationnel du Sismacq est constitué de trois étapes distinctes : a) l'extraction et le jumelage des données médico-administratives, b) les analyses et la production des indicateurs et c) la diffusion (figure). Les données proviennent du jumelage de cinq fichiers médico-administratifs, réalisé à partir d'un identifiant unique, le numéro d'assurance maladie du Québec. Les opérations de jumelage sont effectuées à la Régie de l'assurance maladie du Québec (RAMQ), qui administre les systèmes publics d'assurance maladie et de médicaments. Le premier couvre toute la population du Québec, tandis que celui des médicaments couvre l'ensemble de la population de 65 ans et plus, ainsi que la portion de Québécois non couverts par les régimes privés associés à l'emploi, soit environ 33% de la population de moins de 65 ans.

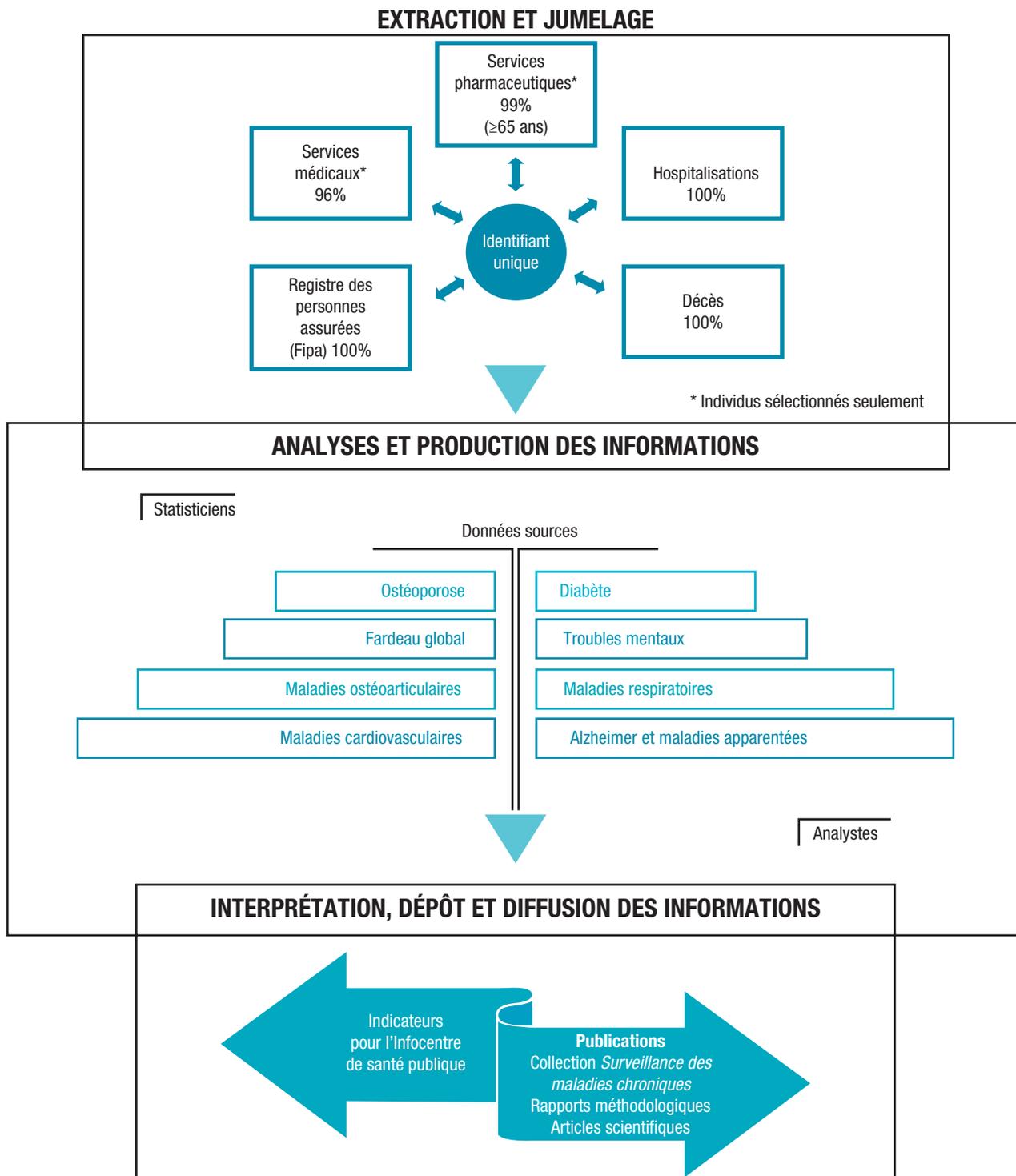
## Les sources et l'extraction des données

Les cinq fichiers utilisés par le Sismacq qui constituent les principales sources d'information sur les utilisateurs du système de santé québécois, les services reçus et les issues de santé sont 1) le fichier d'inscription des personnes assurées (Fipa) ; 2) le fichier des hospitalisations (Med-Écho) ; 3) le fichier des décès du Registre des événements démographiques ; 4) le fichier des services médicaux rémunérés à l'acte et 5) le fichier des services pharmaceutiques pour les personnes âgées de 65 ans et plus. Ces cinq fichiers médico-administratifs sont mis à jour annuellement et la période disponible à la surveillance des maladies chroniques s'étend actuellement du 1<sup>er</sup> janvier 1996 au 31 mars 2012.

Le Fipa contient l'information sur les personnes couvertes par les deux régimes d'assurance (maladie et médicaments). On y retrouve essentiellement des informations démographiques et géographiques comme l'âge de la personne assurée, le sexe, le code postal à six positions, la date du décès et l'information sur l'éligibilité à l'assurance maladie du Québec.

Le fichier des hospitalisations contient des informations relatives aux séjours hospitaliers survenus dans les centres hospitaliers québécois dispensant des soins généraux ou spécialisés. Les données concernent les soins physiques et psychiatriques de

**Le modèle opérationnel du Système intégré de surveillance des maladies chroniques du Québec (Sismacq)**



courte durée et les chirurgies d'un jour. Plus précisément, ce fichier comprend principalement les dates d'admission et de sortie, les diagnostics principaux et secondaires et les traitements.

Le fichier des décès recense tous les décès des Québécois. Il comporte la date du décès, sa cause principale et ses causes contributives.

Le fichier des services médicaux rémunérés à l'acte contient les informations relatives au paiement des médecins, des renseignements sur les professionnels

de santé, la date, le lieu, les codes correspondant à l'acte médical posé et au diagnostic présent dans 91% des réclamations.

Finalement, le fichier des services pharmaceutiques inclut les réclamations de médicaments dans le cadre du régime d'assurance médicaments et les informations relatives aux prescriptions, aux services ainsi qu'aux professionnels de santé.

Le jumelage des fichiers se réalise en trois grandes étapes : la sélection à la RAMQ des individus

répondant à des critères diagnostiques, pharmaceutiques, thérapeutiques ou de cause de décès liés aux maladies chroniques ; l'extraction des renseignements des différents fichiers à la RAMQ pour les personnes sélectionnées ; et la transmission des informations demandées de la RAMQ à l'INSPQ. Les critères étant peu spécifiques et couvrant une période de temps importante, 95% de la population totale du Québec est intégrée dans le Sismacq et 99% des personnes de plus de 65 ans. Par ailleurs, il suffit qu'une personne réponde une seule fois à l'un de nos critères, peu importe l'année de référence, et cette personne est intégrée à la base de données.

Une fois la base de donnée constituée, le Sismacq permet la création d'une cohorte populationnelle et longitudinale. Pour tous les individus sélectionnés, il contient des informations sur l'âge, le sexe, le lieu de résidence, la période d'éligibilité, la date de décès s'il y a lieu, ainsi que l'information complète associée à la consommation de soins de santé (consommation médicale, hospitalisation, consommation de médicament, etc.).

## Production d'indicateurs et analyse

L'analyse et la production d'indicateurs représentent le cœur du développement de la surveillance. Étant donné que l'approche retenue est basée sur l'utilisation de données recueillies à des fins de gestion plutôt que de surveillance, il y a d'abord nécessité d'élaborer des critères ou des algorithmes permettant d'obtenir une définition de cas spécifique à chacune des maladies chroniques pour identifier les personnes atteintes. Cette démarche se décline en plusieurs étapes incluant des études de faisabilité, de comparabilité, de validation ainsi que des consultations auprès d'experts cliniques<sup>6</sup>. Deux articles scientifiques dans le présent numéro illustrent ces types d'études<sup>3,4</sup>. Les indicateurs sont ensuite développés afin de couvrir les objectifs de surveillance fixés. Les mesures épidémiologiques de base telles que la prévalence et l'incidence de la maladie sont priorisées, mais les données jumelées permettent aussi d'analyser l'utilisation des services et la mortalité des personnes atteintes de maladies chroniques en les comparant au reste de la population. Cette approche à portée populationnelle est novatrice et comporte des défis scientifiques et méthodologiques qui permettront d'actualiser la surveillance des maladies chroniques en répondant aux besoins grandissants d'information.

## Diffusion des indicateurs, des informations et des analyses

Lorsque les indicateurs sont calculés et validés, ils sont diffusés aux utilisateurs de données que sont les décideurs, les acteurs de surveillance, les professionnels de la santé, les chercheurs, les associations et la population générale. La collection *Surveillance des maladies chroniques* de l'INSPQ permet la diffusion de brèves analyses utiles aux différents

utilisateurs de données<sup>7-9</sup>. Plusieurs travaux de validation et de développement scientifique sont aussi publiés dans des revues scientifiques reconnues par les pairs. L'objectif visé est de transmettre l'ensemble des indicateurs issus du Sismacq et de les rendre disponibles aux niveaux régional et sous-régional via l'Infocentre de santé publique, un portail destiné notamment aux acteurs régionaux de surveillance.

L'interprétation des données issues de la surveillance est aussi une composante essentielle du modèle opérationnel. Elle permet de vulgariser les informations découlant des analyses, de présenter leurs limites, de fournir des pistes d'explication permettant de comprendre les observations faites et, finalement, de faire le lien avec la prise de décision et l'intervention. L'analyse et l'interprétation des données entraînent le développement de différents messages de surveillance. Ces messages sont adaptés selon les publics cibles visés afin d'assurer une bonne compréhension des informations à diffuser.

## Des outils structurants

En plus du modèle opérationnel décrit dans cet article, des outils structurants chapeautent le développement et la mise en œuvre des activités de surveillance reliées au Sismacq. Un cadre de sécurité de l'information spécifique par des procédures détaillées des mesures physiques, administratives et technologiques à mettre en place et à appliquer pour assurer la sécurité des données et la confidentialité de l'information. Le développement d'un cadre de qualité de données viendra définir les dimensions de la qualité nécessaires pour apprécier les sources de données, les variables, la compréhension et l'interprétation des mesures et des résultats produits. Les dimensions couvertes dans le cadre de qualité sont notamment la pertinence, la cohérence, l'utilisabilité et l'objectivité. Un plan de transfert des connaissances sera aussi élaboré en s'inspirant du modèle décrit dans l'article du présent numéro portant sur la surveillance des maladies respiratoires<sup>3</sup>.

## Limites et forces du Sismacq

La principale force et le potentiel du Sismacq résident dans sa dimension populationnelle. Ainsi, les indicateurs produits sont représentatifs de l'ensemble de la population. La perspective longitudinale permet de constituer des cohortes atteintes de différentes maladies et de caractériser les trajectoires de soins. Cependant, l'utilisation de sources de données médico-administratives n'est pas sans limite. En effet, le Sismacq, dans sa conception, n'intègre pas de données cliniques telle que la tension artérielle, ni de résultats de laboratoire comme la valeur de la glycémie ou les informations des dossiers médicaux. Par ailleurs, même si elles représentent la quasi-totalité de la population, seules les personnes utilisant les services du système de santé sont intégrées dans le Sismacq. Les données sur plusieurs déterminants et les habitudes de vie des individus ne font

pas partie des composantes du système. Enfin, les sources de données, parce qu'elles sont populationnelles et couvrent un univers très large, peuvent par leurs caractéristiques laisser place à des erreurs de codage. Toutefois, ces limites sont évaluées par des études de validation qui font partie des démarches de la surveillance des différentes maladies.

Le Québec participe également au développement d'une surveillance pancanadienne des maladies chroniques via le Système canadien de surveillance des maladies chroniques<sup>10</sup>. Cette initiative, coordonnée par l'Agence de la santé publique du Canada, permet de mettre en synergie l'expertise développée dans chacune des provinces et territoires canadiens en matière de surveillance des maladies chroniques. Cette mise en commun apporte également des éléments supplémentaires de validation des données et des approches méthodologiques.

## Conclusion

La mise en place du Sismacq permet une surveillance des maladies chroniques actualisée, qui tient compte des nouvelles réalités associées aux maladies chroniques, à leur co-occurrence et à l'organisation des soins et services pour les personnes atteintes de ces maladies, dans un contexte de vieillissement de la population. Les données du Sismacq viennent ainsi répondre à des besoins grandissants d'information sur les groupes les plus à risque de développer une ou plusieurs maladies chroniques, sur l'utilisation des soins et sur les issues de santé des personnes qui développent ces maladies.

Le Sismacq pourra être bonifié au cours des prochaines années afin de mieux répondre aux besoins d'information des principaux utilisateurs des données de surveillance. Il pourra comprendre un plus grand nombre de fichiers médico-administratifs qui couvriront davantage le *continuum* de soins et services reçus par les Québécois. Par exemple, il est actuellement envisagé d'y intégrer le fichier des services ambulatoires ou les enquêtes de santé portant sur la satisfaction des soins au Sismacq. Bien que le système de surveillance des maladies chroniques inclue la quasi-totalité de la population, il pourrait couvrir l'ensemble de la population. L'exhaustivité permettrait de réaliser des études pour comprendre l'évolution des maladies chroniques à travers le temps et l'espace. ■

## Remerciements

Le financement du projet est assuré par l'Agence de la santé publique du Canada, le ministère de la Santé et des Services sociaux du Québec et l'INSPQ. Un financement ponctuel s'ajoute dans le cadre de projets de recherche liés à la validation des données et au développement méthodologique.

## Références

- [1] Koné Pefoyo AJ, Rivard M, Laurier C. Public health surveillance and role of administrative data. *Rev Epidemiol Santé Publique*. 2009;57(2):99-111.
- [2] Roos LL, Gupta S, Soodeen RA, Jebamani L. Data quality in an information-rich environment: Canada as an example. *Can J Aging*. 2005;24 Suppl 1:153-70.
- [3] Doucet M, Rochette L, Gagné ME, Gamache P, Pouliot S, Émond V. Développement de la surveillance des maladies respiratoires au Québec à partir des données médico-administratives. *Bull Epidemiol Hebd*. 2013;(Hors-série):36-42.
- [4] Jean S, Candas B, Belzile ÉL, Morin S, Bessette L, Brown JP. Développement de la surveillance de l'ostéoporose au Québec : utilisation des banques de données médico-administratives. *Bull Epidemiol Hebd*. 2013;(Hors-série):15-22.
- [5] Blais C, Rochette L, Hamel D. Prévalence de l'hypertension artérielle au Québec : comparaison entre les données médico-administratives et les données d'enquêtes. *Bull Epidemiol Hebd*. 2013;(Hors-série):23-9.
- [6] Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttman A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol*. 2011;64(8):821-9.
- [7] Blais C, Rochette L. Surveillance de l'hypertension au Québec : incidence, prévalence et mortalité. Québec: Institut national de santé publique du Québec;2011. 20 p.
- [8] Lesage A, Émond V. Surveillance des troubles mentaux au Québec : prévalence, mortalité et profil d'utilisation des services. Québec: Institut national de santé publique du Québec;2012. 16 p.
- [9] Pigeon E, Larocque I. Tendances temporelles de la prévalence et de l'incidence du diabète, et mortalité chez les diabétiques au Québec, de 2000-2001 à 2006-2007. Québec: Institut national de santé publique du Québec; 2011. 12 p.
- [10] Robitaille C, Dai S, Waters C, Loukine L, Bancej C, Quach S, et al. Diagnosed hypertension in Canada : incidence, prevalence and associated mortality. *CMAJ*. 2012;184(1):E49-56.

## Citer cet article

Saint-Laurent D, Blais C, Jean S, Sirois C, Rochette L, Émond V. Le modèle québécois de surveillance des maladies chroniques basé sur l'utilisation des données médico-administratives jumelées. *Bull Epidemiol Hebd*. 2013;(Hors-série):4-8.

## SURVEILLANCE DES MALADIES CHRONIQUES EN FRANCE : LA CONTRIBUTION DES BASES DE DONNÉES MÉDICO-ADMINISTRATIVES

// SURVEILLANCE OF CHRONIC DISEASES IN FRANCE: THE CONTRIBUTION OF HEALTH ADMINISTRATIVE DATABASES

Isabelle Grémy (i.gremy@invs.sante.fr), Anne Doussin

Institut de veille sanitaire, Saint-Maurice, France

Soumis le 26.09.2013 // Date of submission: 09.26.2013

### Résumé // Abstract

La surveillance des maladies chroniques en France s'appuie sur différentes sources de données : enquêtes en population générale, registres, causes médicales de décès et bases de données médico-administratives (BDMA). Ces dernières, exhaustives, réactives, bien qu'initialement non conçues pour l'épidémiologie, constituent désormais une source supplémentaire de données.

L'objectif est de décrire les BDMA, de montrer leurs forces, limites et complémentarités par rapport aux autres sources de données, à travers deux exemples : la surveillance du diabète et des cancers.

Les BDMA couvrent la totalité des soins de ville remboursés et des séjours hospitaliers pour la population résidant en France. Elles permettent de repérer pathologies ou processus morbides à travers leurs prises en charge, d'évaluer des actions de prévention ou de dépistage, etc. L'exhaustivité de ces bases autorise des déclinaisons à des niveaux géographiques fins et leur réactivité permet de renseigner les années les plus récentes. Les BDMA peuvent servir de contrôle des biais de mémoire lorsqu'elles sont croisées avec des enquêtes individuelles par questionnaire, ou servir de bases de redressement en comparant répondants et non-répondants (comme dans l'exemple du diabète). Mais, étant basées sur les soins pris en charge, elles évoluent avec les traitements, les politiques de remboursement, etc. Leur performance en tant qu'outil de surveillance épidémiologique doit être régulièrement réinterrogée. Plusieurs méthodes d'évaluation sont possibles en fonction des objectifs et des autres sources de données existantes.

Les BDMA peuvent contribuer à la surveillance des maladies chroniques, ce d'autant plus que leur qualité, tant de codage que de contenu, ne cesse de progresser, que les possibilités de croisement avec d'autres sources de données augmentent et que les conditions de leur accessibilité se clarifient.

*Surveillance of chronic diseases in France is supported by various data sources: surveys in general population, registers, medical causes of death, and health administrative databases (HADB). Such comprehensive and reactive databases, although they were not initially developed for epidemiological purposes, offer an additional data source.*

*The objective is to describe the HADBs, to show their strength, limitations and complementarity compared to other data sources, through two examples: surveillance of diabetes and cancers.*

*HADBs cover all reimbursed health-care in private practice and hospital admissions for population living in France. They enable to identify pathologies or disease processes through their management, and to assess prevention or screening actions, etc. The comprehensiveness of these databases allows their use at fine geographical levels, and their reactivity provides information on the most recent years. HADBs can help control recall bias when crossed with individual questionnaire surveys, or be used for adjustment when comparing respondents and non-respondents (as in the example of diabetes). However, as they are based upon reimbursed health-care, they change with treatments, reimbursement policies, etc. Their effectiveness as a tool for epidemiological surveillance needs to be regularly questioned. Several assessment methods are possible, according to the objectives and the other existing data sources.*

*HADBs can help the surveillance of chronic diseases, all the more as their quality, in terms of encoding and of content is in constant progress. The possibility of crossing these data with other data sources is increasing, and the conditions of their access are clearer.*

---

**Mots-clés :** Maladies chroniques, Surveillance épidémiologique, Bases de données médico-administratives, France  
*// Keywords:* Chronic diseases, Monitoring, Health administrative databases, France

---

## Introduction

La surveillance épidémiologique des maladies chroniques, mission confiée en France à l'Institut de veille sanitaire (InVS), est indispensable pour suivre, et si possible prévoir, les évolutions temporo-spatiales de ces maladies, de leurs déterminants et de leurs complications. Cette surveillance a notamment pour objectifs de décrire ces évolutions, de déceler d'éventuelles inflexions inattendues ainsi que de mesurer l'impact des traitements, des programmes de prévention et de dépistage et des évolutions du contexte social et économique, dans une perspective d'aide à la décision pour le pilotage des politiques de santé publique.

Les bases de données administratives (BDMA) à finalité médico-économique sont désormais mobilisables pour la surveillance en santé publique. Cette évolution significative des sources de données disponibles conduit à repenser la stratégie de surveillance des maladies chroniques. Moins centré sur les méthodes employées, qui sont détaillées dans d'autres articles du même BEH, cet article a surtout pour objectif de présenter l'éventail des possibilités qu'offrent les BDMA, leurs limites, leurs enjeux et leurs perspectives pour la surveillance des maladies chroniques en France.

## Le dispositif de surveillance des maladies chroniques

La surveillance des maladies chroniques a évolué en France de façon considérable au cours des deux dernières décennies. Jusqu'à la fin des années 1990, elle reposait essentiellement sur des enquêtes déclaratives en population générale, les registres et les données de mortalité.

Depuis un peu plus de 10 ans, les BDMA complètent ce dispositif et prennent une place croissante. Ces bases, conçues pour le suivi des dépenses de santé, se révèlent précieuses pour des utilisations de surveillance épidémiologique. En effet, elles sont centralisées, permanentes et réactives, et recueillent des données de recours aux soins sur l'ensemble de la population prise en charge en France.

Le Programme de médicalisation des systèmes d'information (PMSI) hospitalier a été mis en place progressivement à partir des années 1990. Il enregistre actuellement l'ensemble des hospitalisations du secteur public et privé (« Médecine, chirurgie, obstétrique », soins de suite et de réadaptation, etc.). Le PMSI présente des limites du fait de sa finalité médico-économique mais il s'améliore de manière continue, par exemple grâce à une meilleure qualité de codage des diagnostics, la possibilité d'un chaînage des séjours d'un même patient grâce à un numéro d'anonymat irréversible unique depuis 2001, et une définition du diagnostic principal basée maintenant sur le diagnostic ayant motivé l'hospitalisation, déterminé à la sortie du patient, depuis 2009.

Le système national d'informations inter-régimes de l'assurance maladie (Sniiram) a été créé par la loi en décembre 1998. Cet entrepôt de données individuelles est alimenté en continu par les données servant au remboursement des soins de ville. Il a concerné progressivement tous les régimes d'assurance maladie et couvre actuellement la quasi-totalité de la population vivant en France (plus de 65 millions de personnes). Sont renseignées les prestations présentées au remboursement (date, actes, médicaments, dispositifs médicaux...), des informations sur les prescripteurs et les professionnels de santé ayant délivré les soins, des informations sur le patient concerné, dont la notion de bénéficiaire de la Couverture maladie universelle complémentaire (CMUc), d'affection de longue durée (ALD), de maladie professionnelle et d'invalidité, pour lesquelles les pathologies sont codées selon la Classification internationale des maladies (CIM-10). Toutefois, le motif des actes et les résultats d'examens ne sont pas renseignés. La durée de conservation des données a augmenté. Elle est aujourd'hui de trois ans plus l'année en cours et de 10 ans en « archives ».

Le PMSI est « inclus » dans le Sniiram et, grâce à l'utilisation du même numéro d'anonymat irréversible, soins de ville et hospitalisations d'un même patient peuvent être chaînés.

Depuis leur création, Sniiram et PMSI sont donc en perpétuelle évolution ; leur couverture tend vers l'exhaustivité, en termes de population mais aussi de soins, et les données disponibles s'enrichissent en permanence. De nombreuses utilisations, seules ou combinées, de ces BDMA sont devenues possibles pour la surveillance épidémiologique, notamment des maladies chroniques.

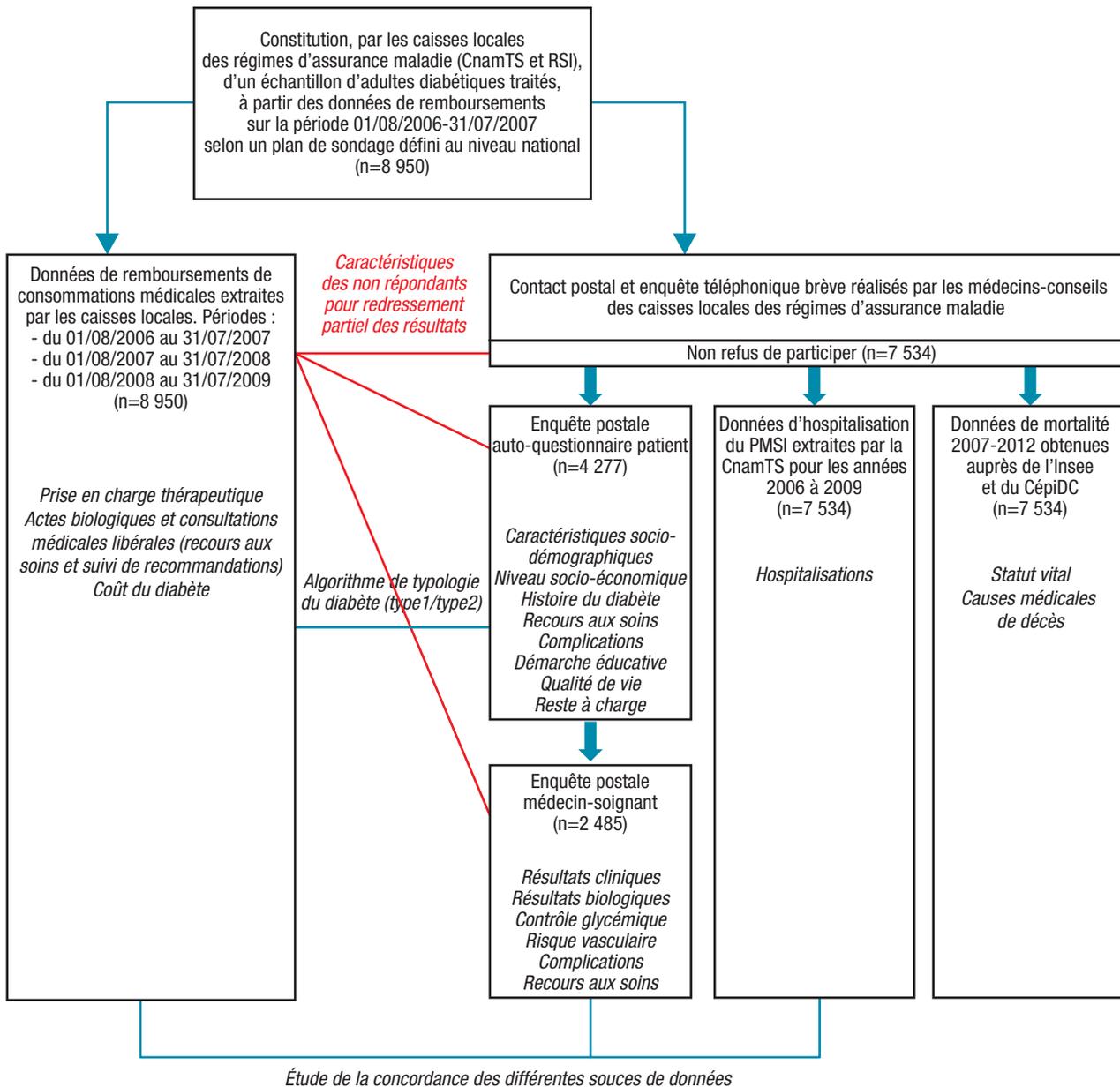
Deux exemples concrets peuvent illustrer ces différentes utilisations : la surveillance du diabète et celle des cancers.

## Surveillance du diabète

La surveillance du diabète repose, entre autres, sur le dispositif Entred (Échantillon national témoin représentatif des personnes diabétiques) qui associe, pour un échantillon de diabétiques traités par antidiabétiques, enquête par auto-questionnaire et données des BDMA (soins ambulatoires ou de ville et hospitalisations *via* le PMSI) (figure 1). Les objectifs sont de décrire, notamment, l'état de santé des diabétiques traités, leur prise en charge ainsi que les déterminants et les complications associés au diabète. Cette étude, réalisée à deux reprises selon une méthode similaire, a permis d'établir un état des lieux de l'état de santé de la population diabétique et de sa prise en charge médicale en début et fin de la précédente décennie. Elle est complétée par des études de mortalité à 5 et 10 ans. Ce dispositif montre par exemple une amélioration sensible, qui reste cependant insuffisante au regard des recommandations, de la prise en charge des personnes diabétiques<sup>1</sup> (figure 2), et met en évidence les inégalités sociales en termes d'état de santé et de qualité des soins. Limite inévitable d'un tirage au sort à partir des BDMA, et donc lié au recours au système

Figure 1

Sources de données et principales analyses dans l'enquête Entred-Métropole, adultes, France 2007-2010



CnamTS : Caisse nationale d'assurance maladie des travailleurs salariés ; RSI : régime social des indépendants ; PMSI: Programme de médicalisation des systèmes d'information ; CépiDC : Centre d'épidémiologie sur les causes médicales de décès (Inserm).

de soins, ce dispositif ne fournit aucun renseignement sur les personnes diabétiques non diagnostiquées et/ou non traitées par antidiabétiques.

**Surveillance des cancers**

La surveillance des cancers est assurée par le dispositif des registres. *Nec plus ultra* de la qualité des données (exhaustivité, recherche active des cas, codage standardisé, programme de formation continue pour les codeurs, panel d'experts pour le codage des cas difficiles, retour possible aux dossiers médicaux, etc.), ce dispositif présente cependant deux écueils majeurs, inhérents à la qualité recherchée :

- une couverture géographique relativement faible (moins de 20% de la population, variable selon le cancer considéré) ;

- un délai de trois ans minimum entre la collecte des données et leur publication.

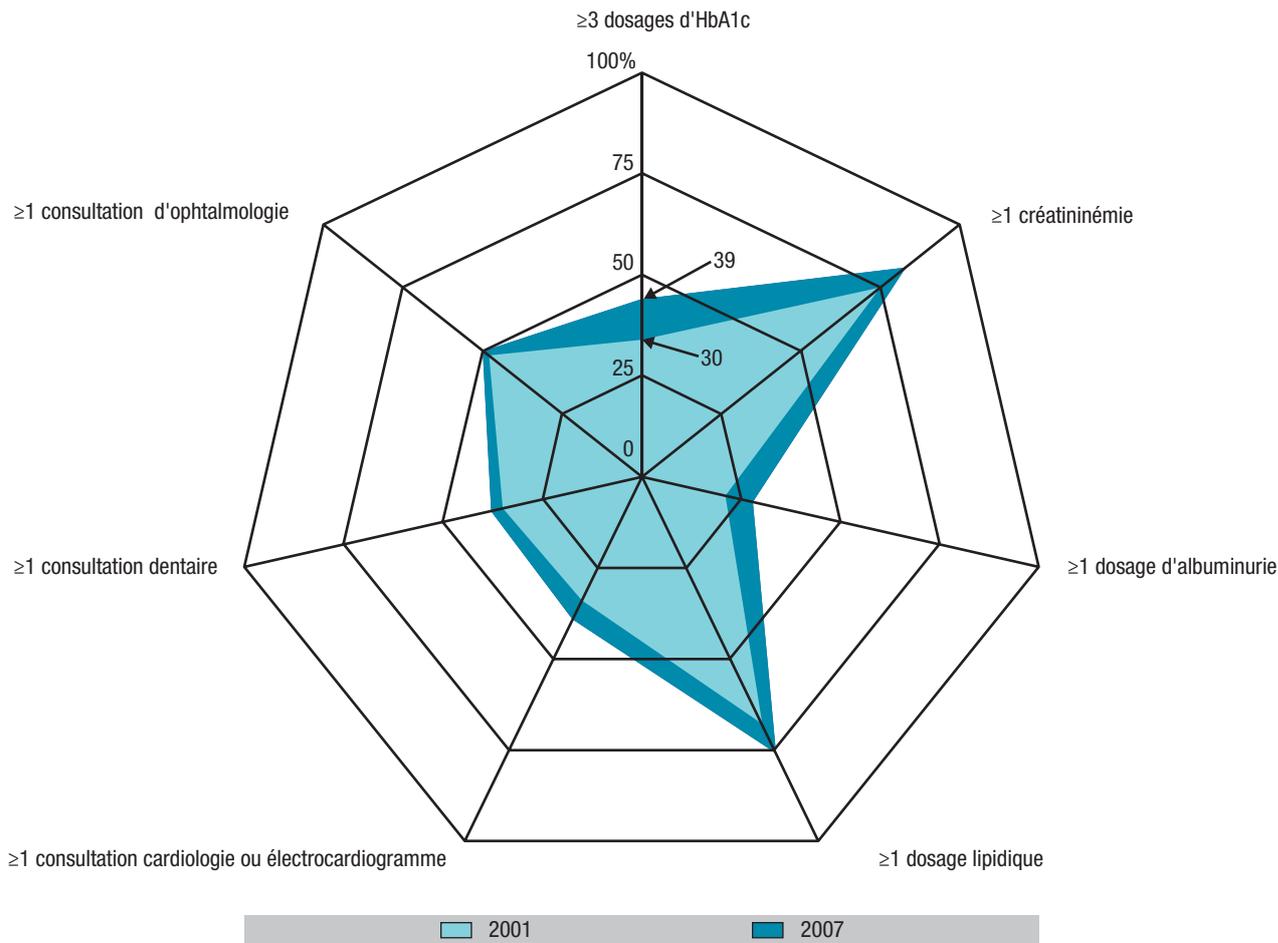
Dans ces conditions, les BDMA peuvent contribuer à la surveillance des cancers pour fournir :

- des informations sur les tendances récentes de l'incidence nationale, c'est-à-dire sur les trois dernières années ;
- des estimations de l'incidence de cancers au niveau régional et départemental dans des zones non couvertes par les registres.

Concernant les tendances récentes de l'incidence nationale, les courbes d'évolution des nouvelles prises en charge pour ALD cancer montrent, pour la plupart des localisations de cancers, des évolutions stables et parallèles aux courbes d'incidence

Figure 2

Évolution entre 2001 et 2007 du suivi des actes recommandés annuellement dans le diabète de type 2



Lecture : 39% des personnes diabétiques enquêtées dans Entred 2007 avaient bénéficié de 3 dosages d'HbA1c dans l'année, contre 30% dans Entred 2001, témoignant d'une amélioration du suivi de l'hémoglobine glyquée chez les personnes diabétiques.

issues des registres. Sur les années les plus récentes, pour lesquelles les données de registres ne sont pas encore disponibles, on suppose que cette relation entre les deux bases de données se maintient, et ce d'autant plus que les courbes obtenues corroborent des connaissances et hypothèses épidémiologiques fiables : par exemple, l'inflexion des ALD notée pour le cancer de la prostate était déjà observée avec la stabilisation de l'incidence des données issues des registres, et elle est cohérente avec les observations européennes de diminution de l'incidence.

Pour les estimations d'incidence infra-régionale ou départementale, utiliser les BDMA (PMSI ou ALD)<sup>2</sup> nécessite de vérifier au préalable qu'elles fournissent de bons proxys de l'incidence et, si oui, de déterminer laquelle des deux, PMSI ou ALD, produit les estimations les plus fiables. La performance de chacune doit être impérativement évaluée par comparaison, pour les départements couverts par les registres, entre les incidences produites à partir des BDMA et celles des registres, qui servent de *gold standard* : une méthode statistique *ad hoc* compare les rapports I/PMSI ou I/ALD pour chaque

département de la zone registre (I est l'incidence départementale issue du registre et PMSI ou ALD est « l'incidence administrative » issue du PMSI ou des ALD). De plus, un algorithme de critères, non présenté ici, a également été mis en œuvre pour permettre d'identifier la plus performante des deux bases.

Cette étude de validation montre que PMSI ou ALD peuvent fournir séparément des estimations d'incidence départementale ou régionale pour 14 des localisations de cancers étudiées (11 pour les ALD). En revanche, pour plusieurs localisations (thyroïde, mélanome, leucémies, système nerveux central) ces bases, en l'état actuel, ne permettent pas des estimations valides. Croiser au niveau individuel PMSI, ALD et/ou d'autres données du Sniiram pourrait permettre d'identifier, par des algorithmes plus fins, les types de cancers pour lesquels les bases ALD ou PMSI séparées se sont révélées peu performantes. L'article de Y. Kudjawa et coll. dans ce même numéro illustre cette possibilité.

## Forces et limites des bases médico-administratives

Si les BDMA ne répondent pas à tous les besoins de surveillance, elles offrent des possibilités nouvelles, comme le montrent les deux exemples ci-dessus et l'ensemble des articles de ce numéro. En effet, Sniiram et PMSI sont, avec les certificats de décès, les principales sources de données médicalisées nationales françaises exhaustives, permanentes et disponibles à un niveau géographique fin (la commune de résidence du patient). Réunies au sein d'un même système d'information, appelé ici « Sniiram-PMSI », elles couvrent quasiment l'ensemble des soins pris en charge par l'assurance maladie de la population résidant en France. Le processus d'amélioration continue, en termes de contenu (nouvelles variables, nouvelles définitions, qualité du codage, etc.) et de couverture populationnelle (quasi-totalité des régimes), les rend utilisables à des fins de surveillance, notamment des maladies chroniques. Les évolutions récentes du cadre réglementaire (arrêté du 19 juillet 2013), qui donnent un accès élargi au Sniiram aux agences sanitaires françaises, et en particulier à l'InVS, vont permettre de développer cette utilisation.

Les BDMA permettent de repérer certaines pathologies, événements de santé ou processus morbides. Elles permettent de distinguer les cas incidents des cas prévalents : grâce à la date des soins et à l'historique disponibles, il est possible de repérer la première occurrence d'un recours aux soins qui définit un cas incident, toute occurrence subséquente pour une même pathologie étant considérée comme prévalente. Le Sniiram-PMSI permet aussi d'apprécier la prise en charge et d'évaluer certaines actions de prévention ou de dépistage comme la couverture vaccinale, les mammographies ou le dosage des PSA.

Un autre intérêt majeur, bien illustré par Entred, mais aussi dans des études de cohortes telles que Constances ([www.constances.fr](http://www.constances.fr)) ou Coset ([www.coset.fr](http://www.coset.fr)), dans les enquêtes transversales santé et protection sociale (ESPS : [www.irdes.fr/EspaceRecherche/Enquetes/ESPS/index.html](http://www.irdes.fr/EspaceRecherche/Enquetes/ESPS/index.html)) ou dans le projet Hygie ([www.irdes.fr/EspaceRecherche/Partenariats/Hygie/index.htm](http://www.irdes.fr/EspaceRecherche/Partenariats/Hygie/index.htm)), réside dans la possibilité de croiser ces bases, au niveau individuel, avec d'autres sources de données, et de créer ainsi de nouveaux systèmes d'information « multisources ». Le croisement des BDMA avec des données d'enquête permet de pallier l'absence de données sociodémographiques, professionnelles et sur les habitudes de vie dans les BDMA ou, à l'inverse, de contrôler certaines données déclaratives (biais de mémoire), d'assurer un suivi passif, y compris des non-répondants et des perdus de vue, et éventuellement d'opérer des redressements. L'appariement avec d'autres sources de données administratives, comme celles de la Caisse nationale d'assurance vieillesse (Cnav), est une autre façon d'accéder à des informations sociales, économiques et professionnelles.

Toutefois, ces bases, en particulier celles de l'assurance maladie, sont complexes, comportent de très nombreuses variables et sont lourdes à utiliser. Par construction, le Sniiram-PMSI ne concerne que les patients pris en charge et les soins remboursés et présentés au remboursement. On n'y trouve ni motif de recours, ni résultat d'examen. Limite ou avantage, il est impossible de remonter au dossier du patient en raison de l'anonymisation irréversible. Ces BDMA évoluent en permanence, notamment en fonction des évolutions de la prise en charge et des politiques de remboursement. La « qualité » des informations est liée à leur finalité *princeps* : par exemple, l'inscription en ALD est « professionnel et patient dépendante » ; elle peut dépendre, entre autres, de l'existence et de la qualité d'une couverture complémentaire de patient ou de la sévérité et du stade de la maladie.

Enfin, dans une perspective d'analyse et de surveillance épidémiologiques, le repérage des patients atteints d'une pathologie peut s'avérer difficile, voire impossible, en particulier pour les pathologies peu ou pas hospitalisées ou qui ne font pas l'objet d'une prise en charge en ALD. L'élaboration d'algorithmes, plus ou moins complexes, avec les cliniciens et les producteurs de données, combinant la consommation de médicaments, la réalisation d'actes médicaux spécifiques et les hospitalisations, peut permettre d'améliorer ce repérage.

Pour ces différentes raisons, les indicateurs d'incidence et de prévalence « administratives » issus de ces bases et leurs évolutions ne peuvent être directement assimilés à des indicateurs épidémiologiques. D'un point de vue épidémiologique, les qualités métrologiques de ces BDMA doivent être évaluées. En fonction des objectifs, plusieurs types d'études sont possibles :

- par comparaison au cas par cas, pour estimer la sensibilité et la valeur prédictive positive, soit par retour au dossier médical au niveau local et sur des effectifs restreints (pour le PMSI, le retour au dossier est possible au niveau d'un établissement et sous réserve de l'obtention des autorisations requises), soit par croisement avec des données d'enquêtes ou de cohortes ;
- par comparaison avec des données agrégées mais reposant sur des données dont la qualité est éprouvée (comme c'est le cas dans l'exemple ci-dessus sur les cancers).

Une veille prospective des évolutions des prises en charge et des bases elles-mêmes est indispensable et peut conduire, en cas de modification importante, à réinterroger la validité des algorithmes utilisés.

Malgré les avantages des BDMA, la surveillance des maladies chroniques ne peut reposer uniquement sur elles. Elle nécessite le recours à d'autres sources de données : enquêtes, examens de santé, etc., notamment pour l'étude des déterminants et habitudes de vie, des stades précoces des pathologies, des recours différentiels au système de soins ou encore pour des populations spécifiques peu ou pas

repérables dans les BDMA (personnes sans domicile par exemple).

## Enjeux et perspectives

Du fait du vieillissement de la population, au-delà de la surveillance épidémiologique thématique des maladies chroniques et afin de guider les politiques de prévention et de prise en charge, une surveillance appropriée de la co-occurrence des maladies, de leurs déterminants et de leurs conséquences en termes de handicap et de dépendance reste à définir et à développer. Dans cette perspective, il paraît indispensable qu'un lien puisse être fait entre les BDMA et les sources de données médico-sociales dans le domaine du handicap et de la dépendance (maisons départementales des personnes handicapées, allocation personnalisée d'autonomie...).

Le contexte économique actuel, particulièrement défavorable, requiert plus que jamais de mettre en œuvre une surveillance efficace, peu coûteuse mais réactive, dans laquelle l'évaluation de l'impact du contexte économique sur la santé, ainsi que la mesure et le suivi des inégalités de santé, en particulier socio-économiques, représentent des priorités. Les BDMA sont pour cela des sources de données particulièrement intéressantes, à condition de pouvoir disposer de variables socio-économiques et professionnelles, soit par le recours à des variables proxys telles que les indices de désavantage social, qui approchent le statut socio-économique individuel sur la base d'une combinaison validée d'indicateurs économiques et sociaux de la zone de résidence, soit par l'appariement avec des données administratives comme celles de la Caisse nationale d'assurance vieillesse (Cnav).

Le pilotage régional de la politique de santé nécessite de disposer d'informations à un niveau géographique infra-régional pertinent. Les informations sont disponibles au niveau communal dans le Sniiram-PMSI, mais une géolocalisation plus fine pourrait permettre des études contextuelles et environnementales au niveau géographique le plus adapté.

Le projet Amphi (Analyse de la mortalité post-hospitalisation), à la recherche d'indicateurs de mortalité représentatifs de la qualité des soins, a montré que l'appariement de la base de données PMSI-MCO à celle des causes médicales de décès *via* le Sniiram était possible. Cet appariement peut et doit maintenant sortir du stade de l'expérimentation. En effet, outre les appariements déjà évoqués, réunir au sein d'un même système d'information Sniiram-PMSI et causes médicales de décès permettrait d'étudier la mortalité par cause pour certaines pathologies identifiables selon des critères de prise en charge et ferait de ce système un outil d'une très grande puissance.

À plus long terme, la mise en place du dossier médical personnel (DMP) et du cadre d'interopérabilité des

systèmes d'informations en santé pourrait conduire à une structuration et une standardisation des informations de santé enregistrées par les professionnels pour la prise en charge. Elle pourrait ainsi permettre une utilisation de ces informations pour la santé publique, en particulier la surveillance. Compte tenu de l'importance des BDMA, mais aussi de leurs limites, il est impératif que ces données produites pour le soin puissent, à terme, être appariées aux données du Sniiram et du PMSI pour tout ou partie de la population. Cela permettrait par exemple de pouvoir renseigner motifs de consultation et résultats d'examen biologiques, mais aussi habitudes de vie et autres déterminants.

La France dispose de bases de données administratives médicalisées exhaustives nationales enrichies en continu. La possibilité de les appairer à d'autres bases de données d'envergure nationale représente un atout majeur face aux enjeux de santé publique actuels, notamment concernant les maladies chroniques.

Des évolutions du cadre législatif et réglementaire sont indispensables pour rendre ces appariements possibles dans des délais raisonnables et en garantissant aux citoyens la sécurité et la confidentialité des données personnelles de santé ainsi que leurs utilisations dans l'intérêt général.

Ces enjeux et perspectives font l'objet de réflexions stratégiques au niveau national, notamment au Haut Conseil de la santé publique<sup>3</sup> et dans un très récent rapport remis à la Ministre de la santé<sup>4</sup>. ■

## Références

[1] Pornet C, Bourdel-Marchasson I, Lecomte P, Eschwège E, Romon I, Fosse S, *et al*, for the Entred Scientific Committee. Trends in the quality of care for elderly people with type 2 diabetes: the need for improvements in safety and quality (the 2001 and 2007 Entred studies). *Diabetes Metab.* 2011;37(2):152-61.

[2] Colonna M, Mitton N, Remontet L, Belot A, Bossard N, Grosclaude P, *et al*. Méthode d'estimation de l'incidence régionale des cancers à partir des données d'incidence des registres, des données de mortalité par cancer et des bases de données médico-administratives. *Bull Epidémiol Hebd.* 2013;(43-44-45): 566-74.

[3] Haut Conseil de la santé publique. Pour une meilleure utilisation des bases de données administratives et médico-administratives nationales pour la santé publique et la recherche. Paris:HCSP; 2012. 56 p. <http://www.hcsp.fr/Explore.cgi/avisrapportsdomaine?clefr=258>

[4] Bras PL, Loth A. Rapport sur la gouvernance et l'utilisation des données de santé. Paris:Ministère de la Santé et des Affaires sociales;2013. 128 p. <http://www.drees.sante.gouv.fr/rapport-sur-la-gouvernance-et-l-utilisation-des-donnees-de,11202.html>

## Citer cet article

Grémy I, Doussin A. Surveillance des maladies chroniques en France : la contribution des bases de données médico-administratives. *Bull Epidémiol Hebd.* 2013;(Hors-série):9-14.

## DÉVELOPPEMENT DE LA SURVEILLANCE DE L'OSTÉOPOROSE AU QUÉBEC : UTILISATION DES BASES DE DONNÉES MÉDICO-ADMINISTRATIVES

// DEVELOPMENT OF OSTEOPOROSIS SURVEILLANCE IN QUEBEC: USE OF HEALTH ADMINISTRATIVE DATABASES

Sonia Jean <sup>1,2</sup> (sonia.jean@inspq.qc.ca), Bernard Candas <sup>1,2</sup>, Étienne L. Belzile <sup>2,3</sup>, Suzanne Morin <sup>4</sup>, Louis Bessette <sup>2,3</sup>, Jacques P. Brown <sup>2,3</sup>

<sup>1</sup> Institut national de santé publique du Québec (INSPQ), Québec, Canada

<sup>2</sup> Université Laval, Québec, Canada

<sup>3</sup> Centre de recherche du Centre hospitalier universitaire de Québec, Québec, Canada

<sup>4</sup> Université McGill, Montréal, Canada

Soumis le 26/06/2013 // Date of submission: 06/26/2013

### Résumé // Abstract

Une étude de cohorte rétrospective utilisant le Fichier des services médicaux rémunérés à l'acte (FSMRA) du Québec a été réalisée afin de développer et valider un algorithme d'identification des cas incidents de fractures de fragilisation (FF). L'utilisation de l'algorithme validé et le jumelage des données médico-administratives a permis de quantifier l'incidence des FF et d'évaluer leur impact sur l'utilisation des ressources médicales.

La validité de l'algorithme varie selon le site de la fracture et est élevée pour la majorité des sites. Basé sur l'algorithme validé, on observe que les fractures les plus communes sont celles associées à l'ostéoporose. Les fractures de la hanche et du bassin sont plutôt rares avant 70 ans, mais augmentent substantiellement ensuite. Une proportion importante des ressources médicales est consacrée à celles-ci en raison d'une hospitalisation fréquente. Concernant les fractures périphériques, elles surviennent fréquemment dès 50 ans et les soins cliniques liés à leurs traitements sont non négligeables.

En conclusion, le FSMRA permet l'identification des FF incidentes avec une valeur populationnelle unique pour leur surveillance. La méthodologie développée et le jumelage de données médico-administratives démontrent qu'il est possible d'élargir la surveillance des FF et de réaliser un portrait global et détaillé de leur fardeau pour le système de santé.

*A retrospective cohort study using physician-billing claims database (PBCD) was performed to develop and validate an algorithm for fragility fracture (FF) cases identification. The use of the validated algorithm and linkage of health administrative data allowed to quantify FF incidence and to evaluate their impact on health resources utilization.*

*The algorithm validity varies across the fracture site and is high for the majority. According to the validated algorithm, we observe that the most common fractures are those associated with osteoporosis. Hip and pelvic fractures are rare before 70 years of age but increase substantially thereafter. They require considerable medical resources since a large proportion of cases are hospitalized. Peripheral fractures occur frequently after the age of 50 years and clinical care related to their treatment are significant.*

*In conclusion, PBCD allows identification of incident FF cases with a unique value for their monitoring at population level. The methodology developed and the linkage of administrative healthcare databases show that is possible to broaden surveillance of FF and achieve a comprehensive and detailed picture of their burden for the health care system.*

**Mots-clés :** Ostéoporose, Fractures, Données médico-administratives, Validation, Utilisation des ressources médicales, Québec

// **Keywords:** Osteoporosis, Fractures, Health administrative data, Validation, Health resources utilization, Quebec

### Introduction

Les fractures de fragilisation (FF) représentent un problème majeur de santé publique, affectant chaque année un nombre croissant d'individus. Le fardeau des FF est important en raison de la mortalité, morbidité et des coûts de soins de santé qui y sont reliés<sup>1-4</sup>. Elles se manifestent essentiellement à partir de 50 ans et, passé cet âge, elles représentent

81% de l'ensemble des fractures<sup>5</sup>. Malgré leurs conséquences, 80% d'entre elles ne bénéficieront pas d'une prise en charge adéquate<sup>5</sup>.

Au Québec, les contraintes budgétaires actuelles et l'importance de ce problème nécessitent la mise en place d'une surveillance des FF. On doit donc réaliser des études épidémiologiques populationnelles afin d'évaluer l'ampleur et le fardeau des FF, et d'estimer

leur impact sur le système de santé. Les données médico-administratives (DMA) peuvent être utiles pour la réalisation de telles études. À ce jour, les fichiers des hospitalisations (FH) ont été largement utilisés, mais leur validité est limitée à des problèmes de santé nécessitant une hospitalisation, comme les fractures de la hanche<sup>6,7</sup>, qui représentent une petite proportion de toutes les FF. Puisque les fractures surviennent généralement dans un contexte aigu et douloureux, elles nécessitent plusieurs contacts avec le système de santé<sup>7</sup>. Elles devraient donc être facilement identifiables en utilisant le Fichier des services médicaux rémunérés à l'acte (FSMRA) de la Régie de l'assurance maladie du Québec (RAMQ).

À ce jour, aucune étude n'a eu recours aux données de ce fichier pour examiner l'impact des FF sur la santé des populations et le système de santé. Ainsi, les objectifs de cette étude sont 1) d'évaluer la fiabilité des données recueillies dans le FSMRA pour l'identification des cas incidents de FF et 2) d'établir un portrait populationnel des FF chez les femmes de 50 ans et plus et de quantifier leur impact en termes d'utilisation des ressources médicales. La réalisation de cette étude permettra, d'abord, d'évaluer la pertinence de la méthodologie proposée et, par la suite, les informations issues de ces indicateurs aideront à cerner l'ampleur et le fardeau que représentent les fractures sur les ressources médicales et sur les budgets du système de santé. Elles seront utiles aux décideurs pour la planification, l'allocation des ressources et l'organisation des soins, informations essentielles pour maintenir et accroître l'efficacité du système de santé. Elles permettront de suivre l'évolution de ce problème et, éventuellement, seront utiles pour la réalisation d'analyses médico-économiques ayant pour objet d'évaluer l'efficacité d'interventions thérapeutiques ou de santé visant une gestion optimale de cette problématique.

## Matériel et méthodes

### Sources de données et méthodologie de l'étude

Au Québec, le système public d'assurance maladie est administré par la RAMQ, qui gère le FSMRA contenant les informations relatives au remboursement des médecins : numéro d'identification du médecin, numéro d'assurance maladie du patient, date, endroit et code d'acte correspondant à l'acte médical posé et optionnellement, dans 91% des réclamations, le code du diagnostic selon la classification internationale des maladies (CIM-9) est inscrit. La RAMQ gère également le FH, contenant les caractéristiques démographiques des patients, les dates d'admission et de sortie, les diagnostics principaux et secondaires (CIM-10 depuis le 1<sup>er</sup> avril 2006), le traitement et la destination à la sortie pour chaque admission dans les hôpitaux généraux et de soins aigus.

Dans cette étude de cohorte populationnelle rétrospective chez les femmes de 50 ans et plus provenant de trois régions socio-sanitaires, une analyse combinant les informations enregistrées au FSMRA a

permis de développer et valider un algorithme d'identification des cas incidents de fractures. À partir de l'algorithme, toutes les variables relatives aux patientes ayant subi une fracture ont été identifiées et utilisées en lien avec les données démographiques pour estimer les taux d'incidence. Finalement, pour établir l'utilisation des ressources médicales directement liées à chaque fracture dans l'année suivante, les informations de consommation de soins enregistrées dans le FSMRA et le FH ont été utilisées.

Les sites anatomiques de fractures étudiés sont : vertèbre, sacrum, coccyx, bassin, hanche, fémur, épaule, humérus, coude, avant-bras, poignet, genou, tibia, péroné, pied ou cheville. Les fractures du crâne, de la main, des doigts et des orteils ont été exclues puisqu'elles ne sont pas associées à l'ostéoporose. Dans cette étude, l'ensemble des fractures à traumatisme faible et élevé survenant à ces sites anatomiques, dont la majorité sont des FF (plus de 81% de l'ensemble des fractures<sup>5</sup>), ont été considérées tel que proposé dans la littérature<sup>8</sup>.

Ce projet a été approuvé par le Comité d'éthique à la recherche du Centre hospitalier universitaire de Québec (CHUQ) et le jumelage des données a été approuvé par la Commission d'accès à l'information du Québec.

### Développement et validation de l'algorithme

Basée sur le FSMRA, une étude rétrospective chez les femmes de 50 ans et plus provenant de trois régions socio-sanitaires a été réalisée entre le 1<sup>er</sup> septembre 2003 et le 31 mars 2006. Pour développer l'algorithme, tous les actes médicaux potentiellement associés à une FF enregistrés au FSMRA au cours de la période ont été identifiés (la liste des codes d'acte retenus est disponible par ailleurs<sup>9</sup>). Deux groupes de codes d'actes ont été sélectionnés : a) codes d'actes spécifiques au traitement d'une fracture (réductions ouvertes et fermées, (figure 1)) et b) codes d'actes non spécifiques au traitement d'une fracture (immobilisation, consultation médicale, visite de suivi, etc.) mais qui sont accompagnés d'un code diagnostique CIM-9 de fracture et réclamés par un médecin orthopédiste (MO), un médecin urgentiste ou un médecin généraliste (figure 1).

À partir des actes sélectionnés, trois algorithmes d'identification des cas incidents de fractures ont été développés. Une fracture est présente selon :

#### Algorithme 1

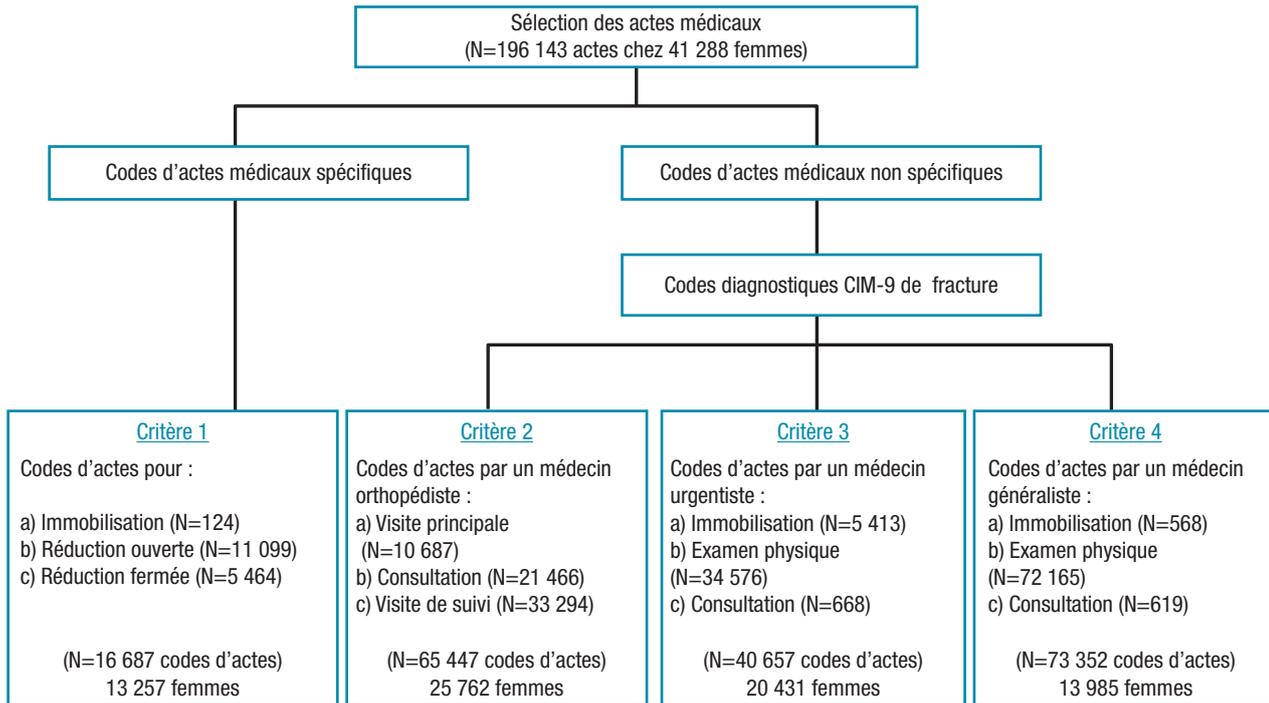
Si on observe au moins un code d'acte associé à : a) une procédure médicale spécifique au traitement d'une fracture (réduction ouverte, réduction fermée, immobilisation) ou b) une visite principale avec un MO avec un code diagnostique de fracture ou c) une consultation avec un MO avec un code diagnostique de fracture.

#### Algorithme 2

Il exclut les cas de l'algorithme 1 identifiés par seulement un code d'acte d'une visite principale avec un

Figure 1

**Sélection des services médicaux pour l'identification des fractures d'après le Fichier des services médicaux rémunérés à l'acte (FSMRA), Québec**



MO. L'ajout d'au moins un autre acte à la visite principale est exigé afin de conserver uniquement les cas ayant un suivi clinique.

**Algorithme 3**

Il exclut les cas de fractures de l'algorithme 2 identifiés par seulement un code d'acte d'une consultation avec un MO. L'ajout d'au moins un autre acte à la consultation est exigé afin de conserver uniquement les cas ayant un suivi clinique.

L'acte permettant l'identification de la fracture a été désigné comme étant l'acte index. Le site de fracture est identifié grâce au code de l'acte index si ce code est spécifique au traitement d'une fracture ou, s'il ne l'est pas, grâce au code diagnostique. Pour établir la séquence temporelle des soins associés à chaque fracture, l'algorithme identifie tous les autres actes médicaux (visite aux urgences, visite de suivi, etc.) liés au même site anatomique. La date de fracture correspond à la première date dans cette séquence. Finalement, une période de 6 mois entre deux séquences liées à un même site anatomique doit être observée afin de minimiser les biais de classification du suivi d'une ancienne fracture comme étant une nouvelle fracture.

Pour évaluer la performance des algorithmes, on a mesuré leurs valeurs prédictives positives (VPP) et leurs sensibilités (Sn). La VPP correspond à la capacité de l'algorithme à identifier des cas de fractures cliniquement confirmés. Pour estimer la VPP, un échantillon aléatoire de 1 506 fractures identifiées par l'algorithme 1 a été sélectionné et la revue des dossiers médicaux hospitaliers par un archiviste a permis de confirmer l'existence de cette fracture.

Au Québec, tous les patients ayant subi une fracture sont référés pour une consultation dans une clinique d'orthopédie (salle de plâtre) localisée dans les établissements hospitaliers ; par conséquent, le dossier hospitalier est considéré comme étant le *gold standard*.

La Sn de l'algorithme correspond à sa capacité à identifier tous les cas de fracture attendus dans la population. Pour estimer la Sn, l'information clinique d'une cohorte de 1 914 cas incidents de fracture représentative des fractures attendues dans la population, cliniquement confirmés et recrutés prospectivement dans les cliniques d'orthopédie (salles de plâtre) dans le cadre d'un projet décrit par ailleurs<sup>10</sup>, a été utilisée ; puis une vérification de l'identification de ces cas par l'algorithme a été effectuée.

**Estimation de l'incidence et utilisation des ressources médicales**

À partir de l'algorithme sélectionné, toutes les patientes ayant subi une fracture entre le 1<sup>er</sup> janvier 2004 et le 31 décembre 2005 ont été identifiées<sup>9</sup>. Cette cohorte représentative de l'ensemble des fractures attendues dans cette population a été utilisée en lien avec les données démographiques pour estimer les taux d'incidence. Pour évaluer l'utilisation des ressources médicales associée aux traitements de ces fractures, les services médicaux enregistrés dans le FSMRA ont été utilisés pour décrire les traitements chirurgicaux associés aux fractures ainsi que le nombre moyen de visites avec un MO et d'autres cliniciens<sup>10</sup>. Finalement, le FH a été utilisé pour estimer la proportion des fractures nécessitant une hospitalisation et évaluer les durées de séjour (DS) hospitalier.

## Résultats

Au total, 196 143 actes chez 41 288 femmes ont été sélectionnés dans le FSMRA selon les critères utilisés pour développer les algorithmes (figure 1). Les trois algorithmes ont identifié 24 001, 21 093 et 18 927 fractures, respectivement.

Les VPP des algorithmes varient selon le site de la fracture (tableau 1) et, pour la plupart des sites, les VPP de l'algorithme 3 sont supérieures et varient de 0,81 à 0,96. Les VPP les plus élevées sont observées pour les fractures du poignet (0,96) et de la hanche/fémur (0,93). Parmi les fractures identifiées par l'algorithme 3 et confirmées dans les dossiers hospitaliers, le site de la fracture est concordant dans 95% des cas, et la date de fracture identifiée par l'algorithme est dans un intervalle de deux jours de la date réelle pour 88% des cas. Parmi les cas ayant une réduction ouverte, la chirurgie est documentée dans le dossier hospitalier de 95% des cas. Les Sn varient également selon le site de la fracture (tableau 2). Les restrictions appliquées aux algorithmes 2 et 3 se traduisent par une diminution de la Sn qui demeure tout de même élevée ( $\geq 70\%$ ) pour la majorité des sites anatomiques. Comme pour les VPP, les Sn les plus élevées sont observées pour les fractures de la hanche/fémur et du poignet. La Sn des trois algorithmes est très faible pour le site vertèbres, coccyx, sacrum. Basée sur les ratios de vraisemblance, une comparaison des algorithmes montre que l'algorithme 3 est supérieur pour prédire l'existence d'une fracture (résultats publiés par ailleurs<sup>9</sup>).

L'algorithme 3 a identifié un total de 18 927 fractures (17 661 femmes). Pour quantifier l'incidence des

fractures et leur impact sur l'utilisation des ressources médicales, les fractures des vertèbres, coccyx, sacrum et du genou ont été exclues compte tenu de la faible validité de l'algorithme pour ces sites. Donc, un total de 17 471 fractures (16 376 femmes) a été retenu pour l'étude du fardeau.

Le tableau 3 présente le nombre de fractures, l'âge médian ainsi que les taux bruts et standardisés selon l'âge. L'âge médian des fractures de la hanche/fémur et du bassin est d'environ dix ans supérieur à l'âge médian des fractures périphériques (FP) (bras et jambes). Les fractures de la hanche/fémur représentent 30% des fractures alors que les FP en représentent 70%, dont les plus fréquentes sont celles du poignet (20%), de l'épaule/humérus (17,1%) et de la cheville (11,9%). Les taux d'incidence des fractures du bras augmentent linéairement avec l'âge et ceux de la jambe demeurent constants (figure 2). Les fractures de la hanche/fémur et du bassin sont relativement rares avant l'âge de 70 ans, mais leurs incidences augmentent exponentiellement par la suite.

Pour traiter ces fractures, 100 923 codes d'actes et plus de 8 500 hospitalisations ont été identifiées dans le FSMRA et le FH, respectivement. Le tableau 4 présente la proportion de fractures ayant un code d'acte associé à une chirurgie, le nombre moyen de codes d'actes associés à une consultation médicale avec un MO ou tout autre médecin, ainsi que la proportion de fractures nécessitant une hospitalisation, avec la durée de séjour moyenne. Comme attendu, les fractures de la hanche/fémur consomment une grande part des ressources médicales puisque la majorité d'entre elles nécessitent une chirurgie (91,1%) et une

Tableau 1

### Valeur prédictive positive (VPP) des algorithmes selon le site de fracture, Québec

Site (N)*	VPP [IC95%]		
	Algorithme 1 (N=1 506)	Algorithme 2 (N=1 341)	Algorithme 3 (N=1 211)
Vertèbres, coccyx, sacrum (25)	0,76 [0,59-0,93]	0,83 [0,67-0,98]	0,82 [0,64-1,00]
Bassin (30)	0,63 [0,46-0,81]	0,73 [0,56-0,90]	0,81 [0,64-0,98]
Hanche, fémur (368)	0,83 [0,79-0,87]	0,89 [0,86-0,92]	0,93 [0,90-0,96]
Épaule, humérus (238)	0,81 [0,76-0,86]	0,85 [0,81-0,90]	0,89 [0,84-0,93]
Avant-bras, coude (147)	0,86 [0,80-0,91]	0,89 [0,84-0,92]	0,90 [0,85-0,96]
Poignet (249)	0,93 [0,90-0,96]	0,95 [0,92-0,98]	0,96 [0,93-0,98]
Genou (76)	0,24 [0,14-0,33]	0,29 [0,17-0,40]	0,63 [0,49-0,76]
Tibia, péroné (83)	0,75 [0,64-0,83]	0,85 [0,76-0,93]	0,89 [0,81-0,97]
Pied (86)	0,85 [0,77-0,93]	0,86 [0,79-0,94]	0,89 [0,81-0,96]
Cheville (318)	0,73 [0,66-0,80]	0,80 [0,73-0,87]	0,88 [0,82-0,94]
Non déterminé (N=54)	0,58 [0,45-0,71]	-	-

\* N Algorithme 1

Tableau 2

## Sensibilité (Sn) des algorithmes selon le site de fracture, Québec

Site (N)	Sn [IC95%]		
	Algorithme 1	Algorithme 2	Algorithme 3
Vertèbres, coccyx, sacrum (10)	0,50 [0,19-0,81]	0,40 [0,10-0,70]	0,40 [0,10-0,70]
Bassin (22)	0,82 [0,66-0,98]	0,82 [0,66-0,98]	0,73 [0,54-0,91]
Hanche, fémur (146)	0,99 [0,97-1,0]	0,97 [0,94-1,0]	0,95 [0,91-0,98]
Épaule, humérus (271)	0,93 [0,90-0,96]	0,89 [0,85-0,93]	0,79 [0,74-0,80]
Avant-bras, coude (178)	0,94 [0,91-0,98]	0,93 [0,89-0,97]	0,88 [0,83-0,93]
Poignet (700)	0,95 [0,94-0,97]	0,94 [0,92-0,96]	0,90 [0,88-0,92]
Genou (0)	-	-	-
Tibia, péroné (146)	0,91 [0,87-0,96]	0,87 [0,82-0,92]	0,84 [0,78-0,90]
Pied (86)	0,84 [0,76-0,92]	0,76 [0,67-0,85]	0,70 [0,60-0,79]
Cheville (318)	0,94 [0,91-0,97]	0,91 [0,87-0,94]	0,88 [0,85-0,92]

Tableau 3

## Incidence et âge médian selon le site de fracture (algorithme 3), Québec

Site de la fracture	Algorithme 3			
	Nombre de fractures	Âge	Taux pour 100 000 personne-années	
	N (%)	Médiane (Q1-Q3)	Taux bruts [IC95%]	Taux standardisés selon l'âge* [IC95%]
<b>Total</b>	17 471 (100,0%)	73 (61-82)	1 561,1 [1 538,1-1 584,5]	1 426,6 [1 404,9-1 448,6]
<b>Site</b>				
Bassin	336 (1,9%)	79 (71-86)	30,0 [26,9-33,4]	25,2 [22,5-28,1]
Hanche, fémur	5 166 (29,6%)	82 (75-88)	461,6 [449,1-474,4]	360,4 [350,2-370,6]
Épaule, humérus	2 988 (17,1%)	72 (61-80)	267,0 [257,5-276,8]	251,6 [242,4-260,9]
Avant-bras, coude	1 708 (9,8%)	68 (59-79)	152,6 [145,5-160,0]	147,3 [140,2-154,5]
Poignet	3 492 (20,0%)	69 (59-79)	312,0 [301,8-322,6]	301,4 [291,2-311,7]
Tibia, péroné	875 (5,0%)	64 (57-76)	78,2 [73,1-83,5]	76,6 [71,5-81,1]
Pied	823 (4,7%)	62 (56-72)	73,5 [68,6-78,7]	74,9 [69,7-80,1]
Cheville	2 083 (11,9%)	63 (57-73)	186,1 [178,2-194,3]	189,5 [181,3-197,8]

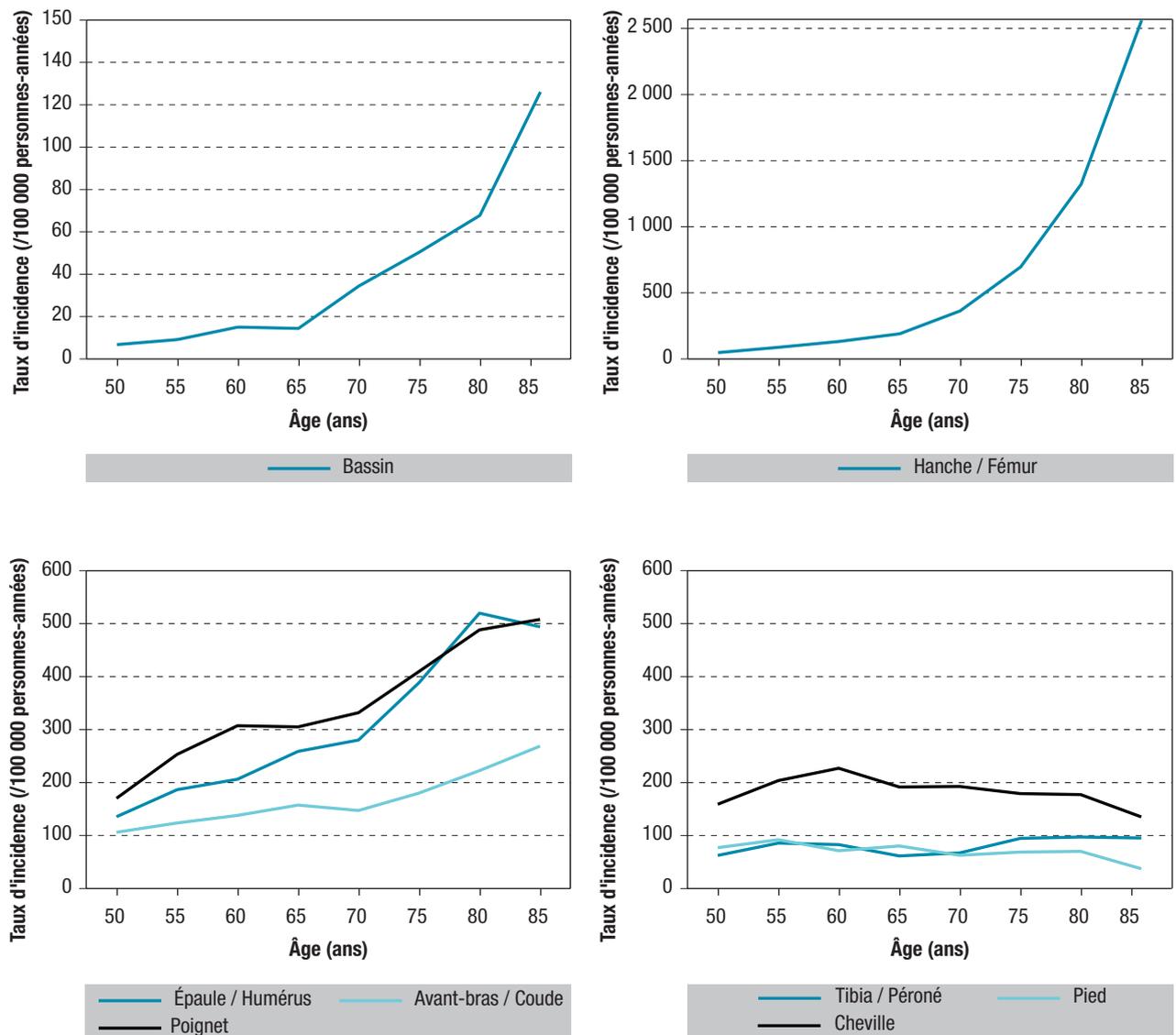
\* Standardisation directe selon la structure de la population canadienne de l'année 1991.

hospitalisation (94,5%) avec une durée de séjour moyenne de 39,2 jours (tableau 4). Elles nécessitent également un nombre élevé de consultations médicales avec une moyenne de 2 consultations avec un MO et 6,9 consultations avec d'autres médecins. Malgré la faible proportion des fractures du bassin traitées chirurgicalement (7,2%), une bonne

proportion nécessite une hospitalisation (67,4%), probablement associée à l'âge avancé des patients. Comme pour les fractures de la hanche, elles utilisent une proportion importante des ressources médicales liées aux consultations (moyenne de 2 consultations avec MO et 6,3 consultations avec d'autres médecins) et à l'hospitalisation (durée de séjour moyenne

Figure 2

Taux d'incidence des fractures selon l'âge et le site anatomique, Québec, 2004-2005



de 32,4 jours). Finalement, certaines FP requièrent une proportion importante des ressources médicales liées aux traitements chirurgicaux et aux hospitalisations (fractures du tibia/péroné : 40,1% en chirurgie et 54,2% en hospitalisation). Ces fractures nécessitent en moyenne 2 consultations avec un MO et le nombre de consultations avec d'autres médecins varie selon le site de la fracture. Les durées d'hospitalisation pour le traitement des FP sont plus courtes que celles observées pour les fractures de hanche/fémur et du bassin, mais tout de même considérables pour certains sites (tibia/péroné 26,3 jours). Ces résultats sont décrits dans une autre publication<sup>11</sup>.

**Discussion**

Dans cette étude, l'utilisation du FSMRA a permis d'élaborer et de valider un algorithme pour l'identification des fractures. À l'aide de l'algorithme et du jumelage des données médico-administratives, l'incidence des fractures et leur impact sur l'utilisation des ressources médicales ont été évalués. Selon les

résultats de ces analyses, les FF sont nombreuses et, en termes d'utilisation des ressources médicales, lourdes de conséquences. Outre les fractures de la hanche, les FP surviennent également fréquemment et l'utilisation des ressources médicales associée à celles-ci est non négligeable, tant sur le plan des soins cliniques liés aux traitements chirurgicaux et aux consultations médicales, que sur le plan des hospitalisations. Dans cette étude, les résultats de l'analyse des taux d'incidence sont compatibles avec ceux de la littérature récente et procurent ainsi une validation indirecte de la méthodologie utilisée. On retrouve peu d'études ayant décrit l'utilisation des ressources associées aux traitements des FF et la plupart se concentrent sur les fractures de la hanche. De plus, leur comparaison est souvent limitée par des différences dans les systèmes de santé, la population et la période étudiées ainsi que dans l'information utilisée dans l'analyse. Cependant, les résultats observés pour la fracture de la hanche sont en adéquation avec ceux d'autres études canadiennes et, tel que rapporté par d'autres études, les variations

Tableau 4

## Utilisation des ressources médicales selon le site de fracture, Québec

Site (N)	Chirurgie <sup>a</sup> en %	Nombre moyen de consultations médicales <sup>b</sup>		Hospitalisation en %	Durée moyenne d'hospitalisation en jours (médiane)
		Orthopédiste (clinique d'orthopédie)	Autres médecins (nombre de visites durant l'hospitalisation)		
Bassin (291)	7,2	1,8	6,3 (4,7)	67,4	32,4 (26)
Hanche, fémur (4 536)	91,1	2,0	6,9 (5,4)	94,5	39,2 (31)
Épaule, humérus (2 603)	16,9	2,4	2,5 (1,2)	30,7	23,5 (8)
Avant-bras, coude (1 498)	13,0	2,2	1,4 (0,2)	18,8	12,7 (3)
Poignet (3 157)	7,9	2,4	1,5 (0,3)	33,3	8,0 (1)
Tibia, péroné (638)	40,1	2,6	3,3 (2,2)	54,2	26,3 (9)
Pied (769)	2,6	1,7	1,3 (0,2)	8,2	16,8 (5)
Cheville (1 835)	32,9	2,6	2,4 (1,1)	41,7	16,6 (3)

<sup>a</sup> Proportion de fractures ayant une réclamation (code d'acte) associée à une chirurgie dans le Fichier des services médicaux rémunérés à l'acte (FSMRA).

<sup>b</sup> Nombre moyen de réclamations (code d'acte) associées à une consultation médicale avec un médecin orthopédiste ou tout autre médecin.

observées selon le site de la fracture traduisent des différences dans leur sévérité.

Certaines limites de la méthodologie utilisée ont été identifiées. La principale porte sur l'incapacité de l'algorithme à identifier de façon adéquate les fractures vertébrales compte tenu de la faible sensibilité à ce site. D'autres auteurs<sup>12</sup> confirment que ces fractures sont en général sous-déclarées dans les données administratives puisqu'elles sont souvent asymptomatiques et reçoivent peu d'attention clinique. D'autre part, aucune procédure n'a été utilisée pour éliminer les fractures traumatiques et pathologiques. Néanmoins, selon une étude récente, les FF représentent plus de 81% de l'ensemble des fractures<sup>5</sup>. De plus, dans les études populationnelles, il est proposé de ne pas exclure les fractures traumatiques puisqu'elles sont également associées à une diminution de la densité minérale osseuse et sont associées à un risque accru de fracture subséquente<sup>8</sup>. De même, il est recommandé de ne pas exclure les fractures pathologiques, car elles représentent une faible proportion et leur exclusion pourrait conduire à une sous-estimation du fardeau associé à l'ostéoporose. L'absence de facteurs de risque clinique et la possibilité d'erreurs de codage sont également des limites inhérentes à l'utilisation des données administratives.

Malgré ces limites, nos résultats ont des implications importantes pour la surveillance épidémiologique des fractures. Le développement d'un algorithme valide nous a permis d'obtenir des estimations fiables de leur incidence. Ces estimations sont représentatives, extrapolables à la population et permettent d'éliminer les biais de sélection usuels. Grâce à l'algorithme, il est désormais possible d'identifier l'ensemble des fractures, incluant celles qui ne conduisent pas nécessairement à une hospitalisation, offrant ainsi une meilleure compréhension de leur fardeau. L'algorithme développé permet non seulement d'effectuer la

surveillance des fractures de la hanche, mais permet d'élargir celles-ci à l'ensemble des sites de fractures associées à l'ostéoporose. De plus, la méthodologie développée grâce à l'utilisation de l'algorithme validé et le jumelage des données médico-administratives démontre qu'il est possible de réaliser un portrait global et détaillé du fardeau clinique des FF.

## Conclusion

En conclusion, cette étude montre qu'il est possible d'utiliser le FSMRA pour effectuer le repérage des cas incidents de fractures dans la population. La méthodologie utilisée, qui repose sur l'utilisation d'un algorithme validé et le jumelage des données médico-administratives, permet non seulement d'obtenir des indicateurs sur la prévalence, l'incidence et la mortalité, mais offre également l'opportunité d'élargir la surveillance à d'autres indicateurs tels que ceux concernant l'utilisation des services de santé.

Finalement, comme le suggère l'interprétation des indicateurs issus de cette analyse, les FF sont nombreuses et, en termes d'utilisation des ressources médicales, lourdes de conséquences. Outre les fractures de la hanche, les autres FF sont également fréquentes et sont, elles aussi, un fardeau important pesant sur les individus et le système de santé. Dans le contexte d'une population vieillissante, cette problématique entraînera certainement une charge importante sur le système de santé. Ainsi, les données de base que nous avons générées dans cette analyse permettront de suivre l'évolution de ce problème. L'information issue de ces indicateurs aidera également à comprendre le fardeau que représentent les FF sur les ressources médicales et budgétaires du système de santé, permettant ainsi de soutenir la planification et l'organisation des soins afin de maintenir et d'accroître l'efficacité du système de santé. ■

## Remerciements

Les auteurs remercient Karine Picard (archiviste médicale) pour sa contribution à la revue des dossiers hospitaliers. Cette étude a bénéficié d'une subvention sans restriction de Servier Canada Inc, ainsi que du support des investigateurs du projet ROCQ (*Recognizing Osteoporosis and its Consequences in Quebec*) et leurs partenaires financiers (Merck Canada, Sanofi-Aventis Canada Inc., Warner Chilcott, Amgen Canada Inc., Eli Lilly Canada Inc. et Novartis Pharma Canada Inc.).

## Références

- [1] Cooper C. The crippling consequences of fractures and their impact on quality of life. *Am J Med.* 1997;103(2A):12S-17S; discussion 17S-19S.
- [2] Ioannidis G, Papaioannou A, Hopman WM, Akhtar-Danesh N, Anastassiades T, Pickard L, *et al.* Relation between fractures and mortality: results from the Canadian Multicentre Osteoporosis Study. *CMAJ.* 2009;181(5):265-71.
- [3] Melton LJ 3<sup>rd</sup>. Adverse outcomes of osteoporotic fractures in the general population. *J Bone Miner Res.* 2003;18(6):1139-41.
- [4] Bessette L, Jean S, Lapointe-Garant MP, Belzile EL, Davison KS, Ste-Marie LG, *et al.* Direct medical costs attributable to peripheral fractures in Canadian post-menopausal women. *Osteoporos Int.* 2012;23(6):1757-68.
- [5] Bessette L, Ste-Marie LG, Jean S, Davison KS, Beaulieu M, Baranci M, *et al.* The care gap in diagnosis and treatment of women with a fragility fracture. *Osteoporos Int.* 2008;19(1):79-86.
- [6] Levy AR, Mayo NE, Grimard G. Rates of transcervical and pertrochanteric hip fractures in the province of Quebec, Canada, 1981-1992. *Am J Epidemiol.* 1995;142(4):428-36.

[7] Ray WA, Griffin MR, Fought RL, Adams ML. Identification of fractures from computerized Medicare files. *J Clin Epidemiol.* 1992;45(7):703-14.

[8] Mackey DC, Lui LY, Cawthon PM, Bauer DC, Nevitt MC, Cauley JA, *et al.* ; Study of Osteoporotic Fractures (SOF) and Osteoporotic Fractures in Men Study (MrOS) Research Groups. High-trauma fractures and low bone mineral density in older women and men. *JAMA.* 2007;298(20):2381-8.

[9] Jean S, Candas B, Belzile E, Morin S, Bessette L, Dodin S, *et al.* Algorithms can be used to identify fragility fracture cases in physician-claims databases. *Osteoporos Int.* 2012;23(2):483-501.

[10] Bessette L, Ste-Marie LG, Jean S, Davison KS, Beaulieu M, Baranci M, *et al.* Recognizing osteoporosis and its consequences in Quebec (ROCQ): background, rationale, and methods of an anti-fracture patient health-management programme. *Contemp Clin Trials.* 2008;29(2):194-210.

[11] Jean S, Bessette L, Belzile EL, Davison KS, Candas B, Morin S, *et al.* Direct medical resource utilization with osteoporosis-related nonvertebral fractures in postmenopausal women. *J Bone Miner Res.* 2013;28(2):360-71.

[12] Curtis JR, Mudano AS, Solomon DH, Xi J, Melton ME, Saag KG. Identification and validation of vertebral compression fractures using administrative claims data. *Med Care.* 2009;47(1):69-72.

## Citer cet article

Jean S, Candas B, Belzile EL, Morin S, Bessette L, Brown JP. Développement de la surveillance de l'ostéoporose au Québec : utilisation des bases de données médico-administratives. *Bull Epidémiol Hebd.* 2013;(Hors-série):15-22.

## PRÉVALENCE DE L'HYPERTENSION ARTÉRIELLE AU QUÉBEC : COMPARAISON ENTRE LES DONNÉES MÉDICO-ADMINISTRATIVES ET LES DONNÉES D'ENQUÊTES

// PREVALENCE OF ARTERIAL HYPERTENSION IN QUEBEC: COMPARISON BETWEEN HEALTH ADMINISTRATIVE AND SURVEY DATA

Claudia Blais<sup>1,2</sup> (claudia.blais@inspq.qc.ca), Louis Rochette<sup>1</sup>, Denis Hamel<sup>1</sup>

<sup>1</sup> Institut national de santé publique du Québec (INSPQ), Québec, Canada

<sup>2</sup> Faculté de pharmacie, Université Laval, Québec, Canada

Soumis le 14/06/2013 // Date of submission: 06/14/2013

### Résumé // Abstract

**Introduction** – L'hypertension artérielle est un facteur de risque majeur des maladies cardiovasculaires et primordial pour la santé publique. Plusieurs sources de données permettent, au Québec, d'évaluer sa prévalence et l'objectif de cette étude est de les comparer.

**Méthodes** – L'évolution de l'hypertension entre 2000-2001 et 2010-2011 est analysée avec les données médico-administratives jumelées du Système intégré de surveillance des maladies chroniques du Québec (Sismacq) et les données auto-déclarées de l'Enquête sur la santé dans les collectivités canadiennes (ESCC), chez les Québécois âgés de 20 à 79 ans. La prévalence en 2009 est comparée avec les données de l'Enquête canadienne avec mesures de la santé (ECMS).

**Résultats** – En 2009, la prévalence avec les données médico-administratives, estimée à 20,6% [IC95%:20,5-20,6] était similaire à celle mesurée ou médicamentée, soit 22,6% [IC95%:20,3-24,8] mais légèrement supérieure à celle auto-déclarée, 18,0% [IC95%:16,9-19,1]. La plus grande différence se trouve parmi les personnes âgées de 65 à 79 ans, particulièrement les hommes. Autant les données du Sismacq que de l'ESCC indiquent une augmentation de la prévalence entre 2000-2001 et 2010-2011.

**Conclusion** – La prévalence de l'hypertension estimée avec les données médico-administratives est similaire à celle obtenue avec des données d'enquêtes combinant mesures de tension artérielle avec utilisation de médicaments.

**Introduction** – Hypertension is a major risk factor for cardiovascular diseases and essential for public health. Multiple data sources allow the evaluation of its prevalence and the objective is to compare them.

**Methods** – Evolution of hypertension between 2000-2001 and 2010-2011 is analyzed using the linked health administrative data of the Quebec Integrated Chronic Disease Surveillance System (QICDSS) and the self-reported data from the Canadian Community Health Survey (CCHS), among Quebecers aged 20-79 years. The prevalence in 2009 is then compared with the data from the Canadian Health Measures Survey (CHMS).

**Results** – In 2009, the prevalence obtained with health administrative data, 20.6% [CI95%: 20.5-20.6] is similar to the one measured or medicated, 22.6% [CI95%:20.3-24.8] but slightly higher than the self-reported, 18.0% [CI95%:16.9-19.1]. The largest difference is found for people aged 65-79 years, especially men. Both QICDSS and CCHS's data show an increase in the prevalence of hypertension between 2000-2001 and 2010-2011.

**Conclusion** – The prevalence of hypertension estimated with health administrative data is similar to the one obtained with survey data that combined measures of blood pressure and the use of drugs.

**Mots-clés** : Hypertension artérielle, Surveillance, Tendances, Données médico-administratives, Données d'enquêtes, Québec

// **Keywords**: Hypertension, Surveillance, Trends, Health administrative data, Survey data, Quebec

### Introduction

Parmi les facteurs de risque des maladies cardiovasculaires, l'hypertension artérielle (HTA) est le plus important car 13% des décès globaux lui sont attribués<sup>1</sup>. Le poids de ce facteur de risque étant majeur pour la santé publique, il est essentiel que des études populationnelles soient effectuées pour en mesurer la prévalence et son évolution dans le temps. Institué par la Loi sur

les services de santé et les services sociaux adoptée en décembre 1971, le système québécois de santé et de services sociaux a pour but le maintien, l'amélioration et la restauration de la santé et du bien-être de la population en rendant accessible un ensemble de services de santé et de services sociaux. L'universalité, l'équité et l'administration publique sont au cœur des principes fondamentaux et, ainsi, les services de santé et les services sociaux sont accessibles à tous sans

discrimination. Ainsi, au Québec, l'administration du système universel de santé s'effectue, entre autres, au moyen de banques de données médico-administratives de santé qui sont continuellement mises à jour. Ces données sont entreposées à la Régie de l'assurance maladie du Québec (RAMQ) et peuvent ensuite être jumelées. Les enquêtes de santé populationnelles sont, quant à elles, effectuées sur une base régulière et fournissent d'autres types de données. Les données recueillies dans ces enquêtes peuvent être auto-déclarées ou mesurées.

Le principal objectif de cette étude est de comparer la prévalence de l'HTA au Québec, estimée à partir de trois sources de données : 1) les données médico-administratives jumelées, 2) les données auto-déclarées issues de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) et 3) les données mesurées issues de l'Enquête canadienne sur les mesures de la santé (ECMS).

## Méthodes

### Sources de données

#### Données médico-administratives

La méthodologie utilisée pour effectuer la surveillance de l'HTA se base sur le « *Canadian Chronic Disease Surveillance System* », un réseau de collaboration entre les systèmes de surveillance provinciaux et territoriaux, appuyé par l'Agence de la santé publique du Canada (ASPC)<sup>2</sup>. Le Système intégré de surveillance des maladies chroniques du Québec (Sismacq) de l'Institut national de santé publique du Québec (INSPQ) est utilisé pour déterminer si une personne a eu un diagnostic d'HTA. Le Sismacq est composé de cinq fichiers médico-administratifs jumelés couvrant actuellement la période du 1<sup>er</sup> janvier 1996 au 31 mars 2011 ; il est mis à jour annuellement. La description de ces cinq fichiers médico-administratifs jumelés est publiée dans l'article de D. Saint-Laurent et coll. dans ce même numéro. En résumé, il y a : 1) le fichier d'inscription des personnes assurées (Fipa), 2) le fichier des hospitalisations, 3) le fichier des décès, 4) le fichier des services médicaux rémunérés à l'acte et 5) le fichier des services pharmaceutiques. Le Fipa contient de l'information sur les personnes couvertes par les deux régimes d'assurance (maladie et médicaments) et on y retrouve essentiellement des informations démographiques et géographiques. Aucun de ces fichiers ne contient de l'information socioéconomique, mais la présence du code postal dans le Fipa permet de faire le lien avec un indice de défavorisation matérielle et sociale développé à l'aide des données de recensement. Les informations cliniques, codifiées avec la Classification internationale des maladies (CIM), se retrouvent principalement dans les fichiers des services médicaux rémunérés à l'acte et dans celui des hospitalisations. Ces deux banques de données sont principalement utilisées pour estimer la prévalence de l'HTA. Le premier fichier contient principalement un code d'acte correspondant à l'acte médical posé que le médecin facture à la RAMQ et, optionnellement, dans

environ 91% des réclamations, le code du diagnostic, tandis que le second fichier contient les dates d'admission et de sortie des patients, les diagnostics principaux et secondaires, les traitements et la destination à la sortie pour chaque admission dans les hôpitaux généraux et de soins aigus. Soulignons que, depuis le 1<sup>er</sup> avril 2006, le fichier des hospitalisations utilise la 10<sup>e</sup> révision de la CIM.

#### Données d'enquêtes réalisées par Statistique Canada

##### Enquête sur la santé dans les collectivités canadiennes (ESCC)

L'ESCC est une enquête annuelle depuis 2007 qui vise à recueillir des renseignements sur l'état de santé, l'utilisation des services et les déterminants de la santé de la population canadienne âgée de 12 ans et plus<sup>3</sup>. Avant 2007, la collecte de données était bisannuelle. Des données de prévalence d'HTA auto-déclarée sont disponibles pour les périodes de 2000-2001 jusqu'à 2011. Les habitants des réserves et autres populations autochtones, les personnes vivant en institution, les membres à temps plein des Forces armées canadiennes et les habitants de régions de faible densité de population sont exclus. Les méthodes de collectes des données sont des entretiens téléphoniques ou en face-à-face au domicile des répondants.

##### Enquête canadienne sur les mesures de la santé (ECMS)

L'ECMS est une enquête qui a pour but de recueillir de l'information sur la santé au moyen d'une entrevue faite au domicile des participants et par des mesures physiques prises dans une clinique mobile<sup>4</sup>. Elle vise les personnes âgées entre 6 et 79 ans, vivant dans des ménages privés au Canada. Les répondants de quatre sites représentatifs de la province de Québec ont été interviewés pour chacun des deux cycles de l'enquête (2007-2009 et 2009-2011). Pour chaque répondant, la tension artérielle du bras droit est mesurée six fois à une minute d'intervalle, en position assise, suivant une période de cinq minutes de repos, selon le nouveau protocole de l'ECMS<sup>5</sup>. Ces mesures sont prises avec un oscillomètre automatisé et la moyenne des cinq dernières est retenue<sup>6,7</sup>.

## Définitions

#### Données médico-administratives

Pour être considéré hypertendu, un individu âgé de 20 ans et plus devait être éligible à l'assurance maladie du Québec, et avoir : a)  $\geq 2$  diagnostics d'HTA inscrits au fichier des services médicaux rémunérés à l'acte à l'intérieur d'une période de 730 jours OU, b) un diagnostic (principal ou secondaire) d'HTA enregistré au fichier des hospitalisations<sup>2</sup>. Les codes diagnostiques suivants sont inclus : 401 à 405 pour la CIM-9 et I10 à I13 et I15 pour la CIM-10. Afin d'exclure les cas d'HTA gestationnelle, les diagnostics d'HTA pour les femmes âgées entre 20 et 54 ans, enregistrés dans une fenêtre de 120 jours avant ou 180 jours suivant toute hospitalisation associée à la grossesse et à l'accouchement, ne sont pas retenus. Pour ces cas, les codes utilisés sont : 641 à 676 et V27 pour la CIM-9 et O10-O19, O21-O95, O98, O99 et Z37 en CIM-10. Cette

définition a été validée dans trois provinces canadiennes avec une sensibilité de 73-75%, une spécificité de 94-95%, des valeurs prédictives positive et négative de 81-87% et 88-92%, respectivement<sup>8,9</sup>.

### Données d'enquêtes

Plusieurs questions en lien avec l'HTA sont posées au participant lors de l'ESCC mais seule la question suivante, présente à tous les cycles, est retenue dans nos analyses: « *Faites-vous de l'hypertension ?* » (*Rappelez-vous, nous nous intéressons aux problèmes de santé diagnostiqués par un professionnel de la santé qui durent ou qui devraient durer 6 mois ou plus*). Environ 1% des Québécois ne déclarent pas leur HTA lorsqu'ils prennent un anti-hypertenseur. Ce biais est donc jugé négligeable. Pour l'ECMS, l'HTA est définie par une moyenne élevée des tensions artérielles systoliques ou diastoliques mesurées ( $\geq 140/90$  mm Hg) ou la notion de traitement antihypertenseur au cours du dernier mois (mesurée/médicaments)<sup>6</sup>.

### Analyses statistiques

Pour les trois sources de données, seule la population âgée de 20 à 79 ans au Québec est considérée, excluant les femmes enceintes. La prévalence calculée avec les données médico-administratives est obtenue en divisant le nombre total de personnes

diagnostiquées hypertendues par les estimations de la population québécoise assurée. Pour les données d'enquêtes, les proportions pondérées d'hypertendus sont obtenues en utilisant les poids d'enquête calculés afin d'assurer la représentativité de la population visée. Des prévalences brutes sont présentées pour chaque combinaison de groupe d'âge et de sexe. Pour l'ensemble de la population, des prévalences ajustées suivant la méthode de standardisation directe sur la structure d'âge de la population québécoise âgée de 20 à 79 ans en 2001 sont également produites pour suivre l'évolution des prévalences dans le temps. Toutes les estimations sont accompagnées de leurs intervalles de confiance (IC) à un niveau de 95%. Lorsque les IC ne se chevauchent pas pour deux estimations issues de sources différentes, on peut conclure à une différence significative, bien que ce test soit considéré conservateur (le seuil réel étant inférieur à 5%). Pour les données d'enquêtes, des tests de Chi<sup>2</sup> ont été effectués afin de comparer les prévalences entre les deux sexes pour un groupe d'âge donné. Les analyses statistiques ont été effectuées avec SAS<sup>®</sup> Enterprise Guide version 5.1 (Cary, NC, USA).

### Résultats

Le tableau résume les caractéristiques des trois sources d'information utilisées pour la mesure de la prévalence de l'HTA au Québec.

Tableau

#### Caractéristiques des trois sources de données pour la surveillance de l'hypertension artérielle au Québec

	Sismacq	ESCC	ECMS
<b>Type de données</b>	Données médico-administratives	Données d'enquête	Données d'enquête
<b>Années disponibles</b>	1996-1997 à 2010-2011	2000-2001 à 2011	2007-2009 à 2009-2011
<b>Âge disponible</b>	Tous	$\geq 12$ ans vivant en ménage privé	6 à 79 ans vivant en ménage privé
<b>Nombre de personnes assurées ou interrogées</b>	7 429 388 en 2000-2001 (20-79 ans : 5 377 877)  7 931 533 en 2010-2011 (20-79 ans : 5 824 242)	22 012 répondants québécois en 2000-2001 (20-79 ans : 18 869)  12 007 répondants québécois en 2011 (20-79 ans : 9 883)	1 545 répondants québécois en 2007-2009 (20-79 ans : 942)  1 374 répondants québécois en 2009-2011 (20-79 ans : 768)
<b>Définition de l'hypertension (excluant les femmes enceintes)</b>	$\geq 2$ diagnostics au fichier des services médicaux à l'intérieur de 730 jours OU un diagnostic enregistré au fichier des hospitalisations	Auto-déclarée ; « <i>Faites-vous de l'hypertension ?</i> »	Mesurée ou auto-déclaration de médicaments : tension artérielle mesurée $\geq 140/90$ mm Hg OU utilisation d'antihypertenseurs dans le dernier mois
<b>Exclusion</b>	<ul style="list-style-type: none"> <li>Femmes enceintes</li> <li>Personnes vivant en institution (ménages collectifs)</li> <li>Numéro d'assurance maladie inexistant ou invalide</li> <li>Inadmissibilité à l'assurance maladie de la personne assurée</li> <li>Carte d'assurance maladie expirée ou retournée (cette exception ne s'applique pas pour les hommes de 20-29 ans et les femmes de 20-25 ans)</li> </ul>	<ul style="list-style-type: none"> <li>Femmes enceintes</li> <li>Personnes vivant en institution (ménages collectifs)</li> <li>Réserves et autres peuplements autochtones</li> <li>Membres à temps plein des Forces armées canadiennes</li> <li>Régions de faible densité de population (Nunavik et des Terres-Cries-de-la-Baie-James)</li> </ul>	<ul style="list-style-type: none"> <li>Femmes enceintes</li> <li>Personnes vivant en institution (ménages collectifs)</li> <li>Réserves et autres peuplements autochtones</li> <li>Membres à temps plein des Forces armées canadiennes</li> <li>Régions de faible densité de population (Nunavik et des Terres-Cries-de-la-Baie-James)</li> </ul>

Sismacq : Système intégré de surveillance des maladies chroniques du Québec.

ESCC : Enquête sur la santé dans les collectivités canadiennes.

ECMS : Enquête canadienne sur les mesures de la santé.

## Données médico-administratives

En 2010-2011, le nombre de Québécois âgés entre 20 et 79 ans diagnostiqués hypertendus se chiffre à 1 195 207, ce qui représente une prévalence brute de 20,5% [IC95%:20,5-20,6] et une prévalence ajustée pour l'âge de 18,5% [IC95%:18,4-18,5]. La figure 1 montre que, chez les Québécois d'âge compris entre 20 et 44 ans, la prévalence est toujours légèrement plus élevée chez les hommes que chez les femmes. Pour les personnes entre 45 et 64 ans, la prévalence est plus élevée chez les femmes jusqu'à l'année 2006-2007, les hommes présentant ensuite une prévalence plus élevée. Cependant, chez les personnes âgées de 65 à 79 ans, la prévalence est beaucoup plus élevée chez les femmes, et ce, durant toute la période étudiée ( $p < 0,05$ ).

## Données d'enquêtes auto-déclarées

En 2011, dans le cadre de l'ESCC, le nombre de Québécois âgés entre 20 et 79 ans qui déclarent avoir eu un diagnostic d'HTA est estimé à 1 006 690, ce qui représente une prévalence brute de 17,5% [IC95%:16,5-18,6] et une prévalence ajustée pour

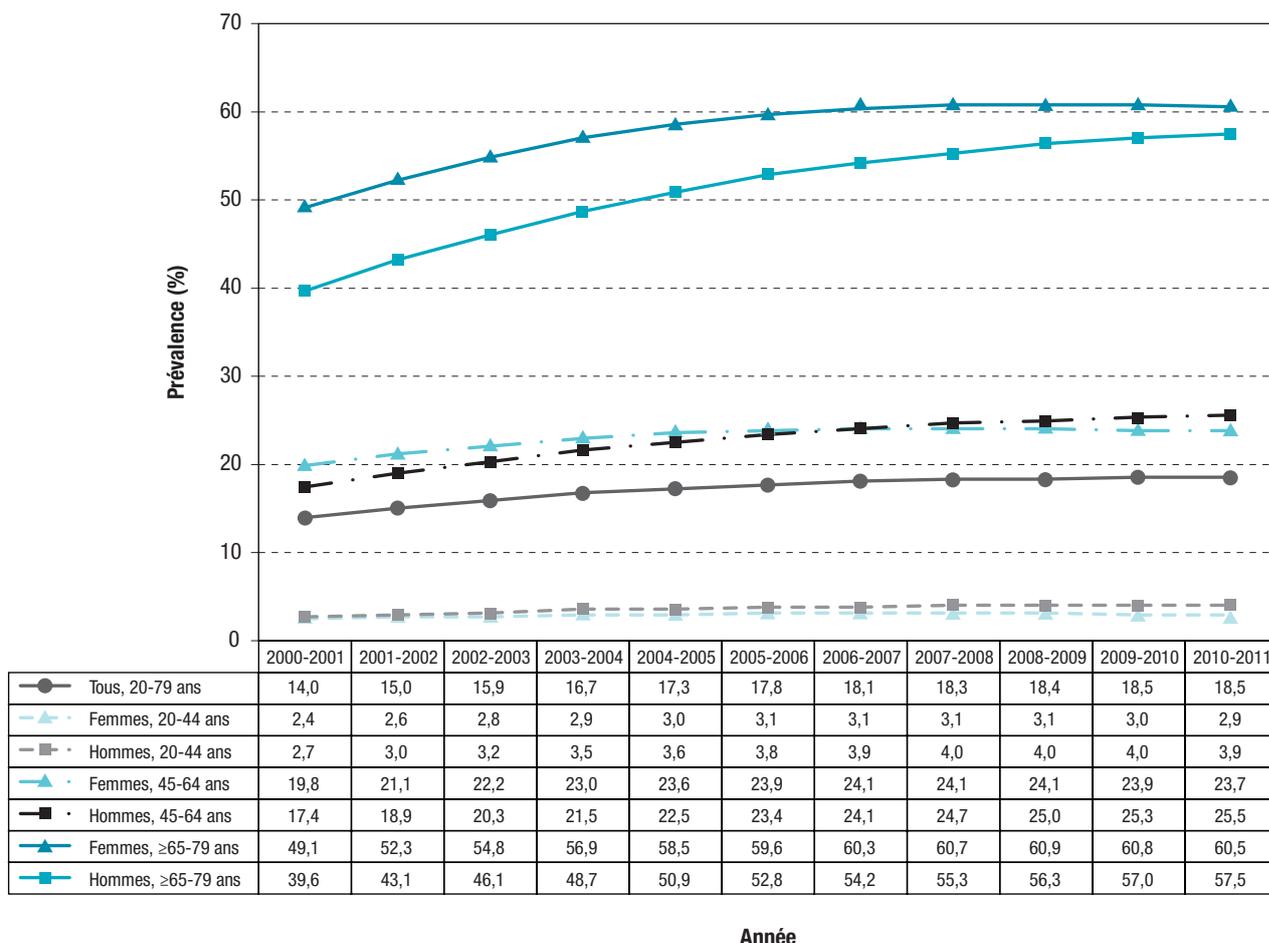
l'âge de 15,7% [IC95%:14,8-16,6]. La figure 2 montre que la différence entre les hommes et les femmes, selon les groupes d'âge, est presque similaire à celle observée avec les données médico-administratives, et ce, tout au long de la période. Les hommes et les femmes âgés entre 20 et 44 ans semblent présenter une prévalence similaire de l'HTA. La prévalence de l'HTA auto-déclarée chez les femmes âgées entre 45 et 64 ans tend à être supérieure à celle des hommes de leur âge jusqu'à l'année 2007-2008, où l'on observe une tendance à l'inverse. Chez les personnes âgées de 65 à 79 ans, les femmes affichent une prévalence supérieure à celle des hommes, à l'exception de la dernière année ( $p = 0,0861$ ). Tout comme avec les données médico-administratives, il y a également une augmentation relative de la prévalence ajustée de l'HTA auto-déclarée de 16,3% entre 2000-2001 et 2011 ( $p < 0,05$ ).

## Comparaison de la prévalence de l'HTA en 2009

Le nombre de Québécois hypertendus âgés de 20 à 79 ans est estimé à 1 275 290 selon les deux cycles combinés de l'ECMS. Tel qu'illustré à la figure 3, en 2009, la prévalence de l'HTA dans l'ECMS est

Figure 1

### Évolution de la prévalence standardisée\* de l'hypertension diagnostiquée par catégorie d'âge et de sexe pour la population âgée de 20 à 79 ans au Québec, de 2000-2001 à 2010-2011



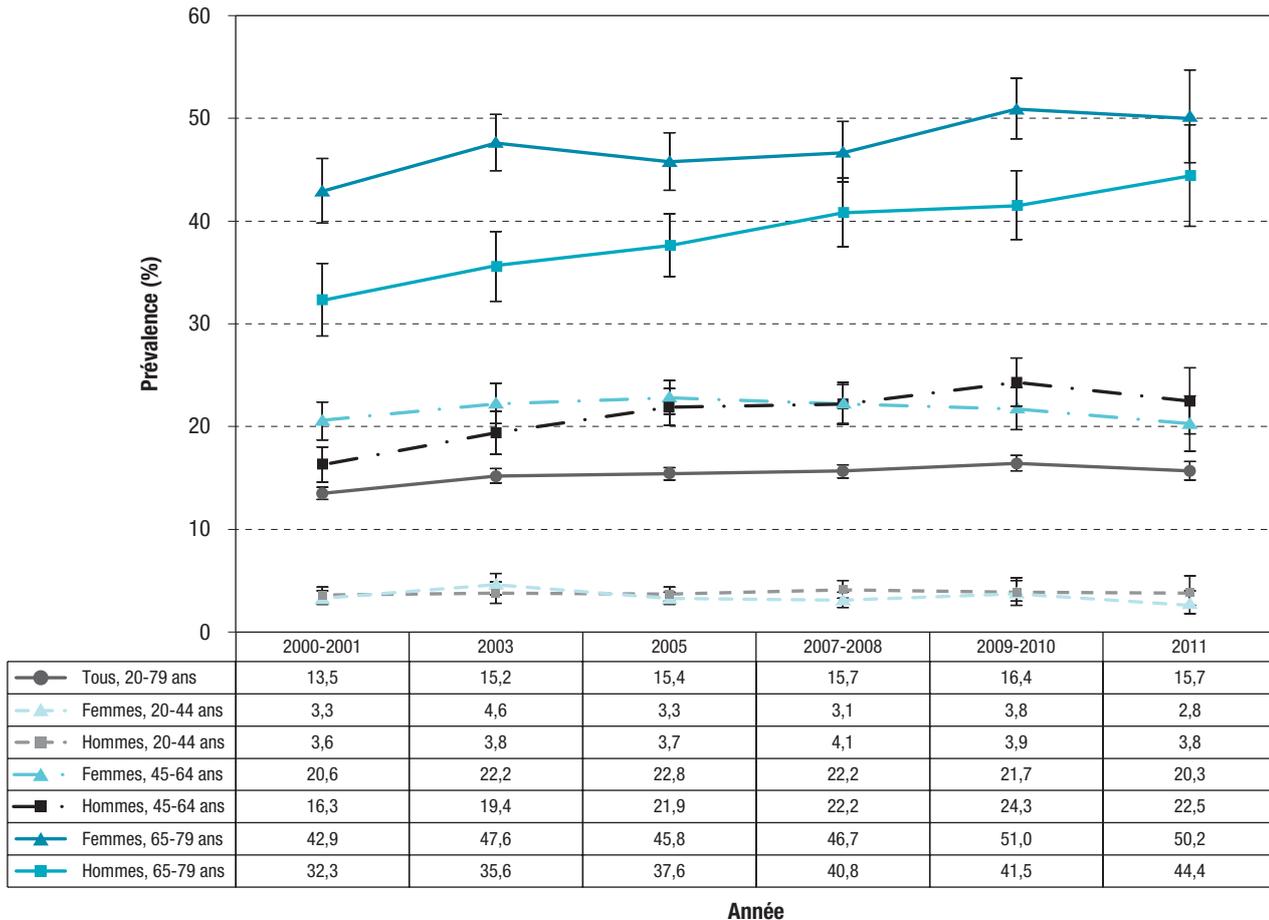
\* Prévalence ajustée à partir de la structure d'âge de la population âgée de 20 à 79 ans du Québec en 2001.

Note : les intervalles de confiance pour les données médico-administratives ne sont pas présentés en raison de leur faible amplitude.

Source : Système intégré de surveillance des maladies chroniques du Québec (Sismacq) de l'Institut national de santé publique du Québec (INSPQ).

Figure 2

**Évolution de la prévalence standardisée\* de l'hypertension auto-déclarée par catégorie d'âge et de sexe pour la population âgée de 20 à 79 ans au Québec, de 2000-2001 à 2011**



\* Prévalence ajustée à partir de la structure d'âge de la population âgée de 20 à 79 ans du Québec en 2001.

Source : Les cycles 1.1 à 2011 de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) de Statistique Canada.

de 22,6% [IC95%:20,3-24,8]. Elle n'est pas statistiquement différente de celle calculée avec les données médico-administratives qui se chiffre à 20,6% [IC95%:20,5-20,6], alors que la prévalence de l'HTA auto-déclarée (ESCC) est significativement plus faible, soit 18,0% [IC95%:16,9-19,1], comparativement à celles estimées à partir des deux autres sources de données (p<0,05). Notons que la plus grande différence entre les sources de données est observée chez les personnes âgées de 65 à 79 ans et plus particulièrement les hommes, avec des estimations variant de 42,7% pour les données auto-déclarées à 61,1% pour les données combinant mesures et médicaments.

**Discussion**

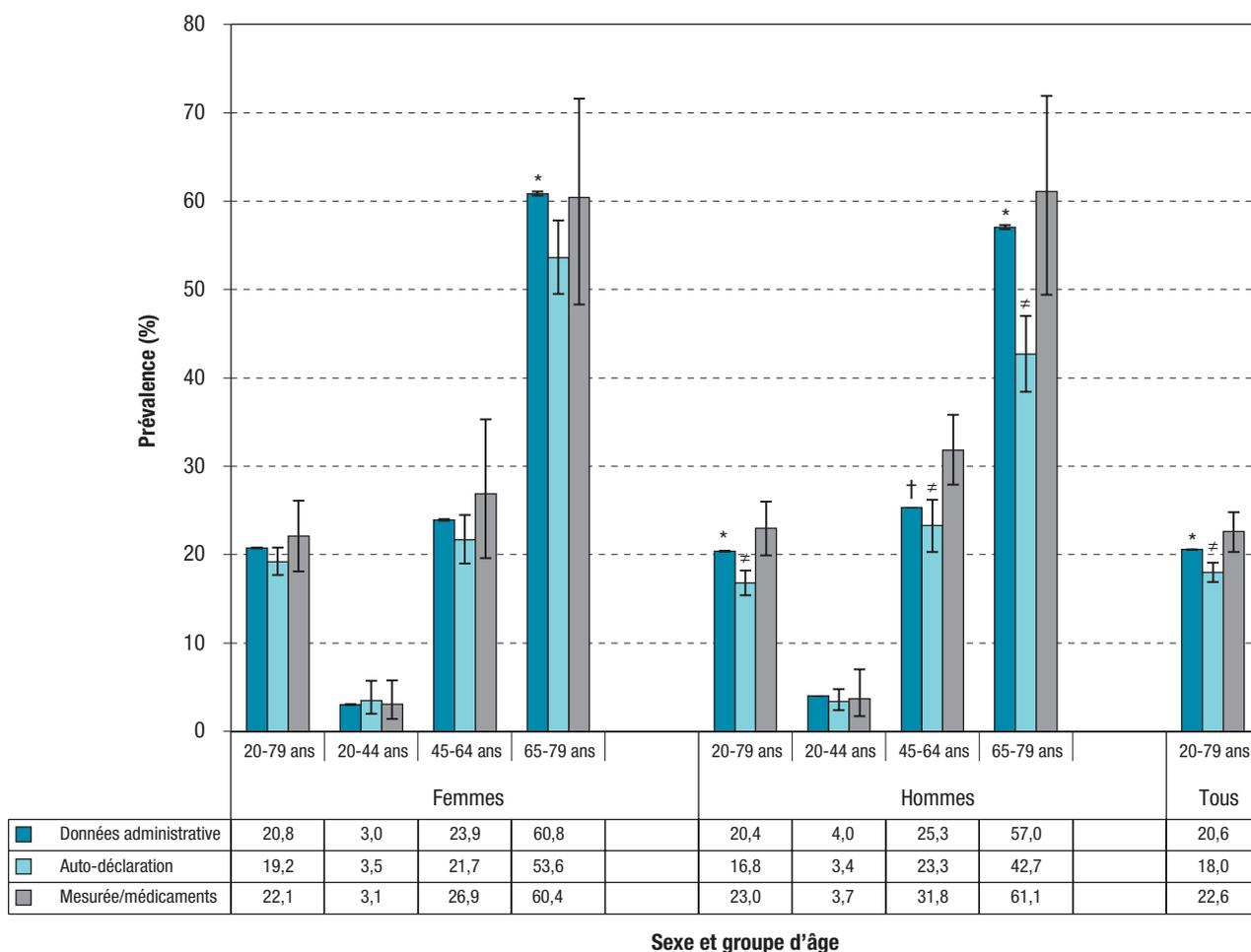
En 2009, chez les Québécois âgés de 20 à 79 ans, la prévalence de l'HTA calculée à partir des données médico-administratives est comparable à celle calculée avec les données d'enquêtes mesurées, tandis que les données auto-déclarées révèlent une légère sous-estimation. La plus grande différence entre les sources de données est observée chez les personnes âgées entre 65 et 79 ans,

plus particulièrement les hommes. Selon les plus récentes données médico-administratives, l'HTA a été diagnostiquée chez 1 195 207 Québécois âgés entre 20 et 79 ans en 2010-2011 (20,5%). La prévalence de l'HTA augmente avec les années, qu'elle soit estimée avec les données médico-administratives ou avec les données auto-déclarées de l'ESCC. En outre, en 2009, selon ces deux sources de données, la différence observée entre les hommes et les femmes est particulièrement marquée chez les personnes âgées entre 65 et 79 ans, catégorisant alors les femmes de cet âge comme un groupe à risque élevé, alors que cela n'est plus vrai avec les données mesurées de l'ECMS.

La prévalence relativement plus faible de l'HTA auto-déclarée, lorsque comparée aux prévalences obtenues par les deux autres sources de données, n'est pas surprenante et peut être expliquée par le fait que l'HTA est souvent asymptomatique, donc fréquemment inconnue par les individus participant à des enquêtes dont les données sont auto-déclarées. De plus, certaines personnes omettent de déclarer leur HTA lorsque cette dernière est traitée et contrôlée. À l'inverse, lorsque les données sont mesurées comme dans l'ECMS, certaines personnes peuvent

Figure 3

**Prévalence de l'hypertension basée sur les données médico-administratives, les données d'enquêtes, hypertension auto-déclarée et tension artérielle mesurée/consommation de médicaments, selon l'âge et le sexe pour la population âgée de 20 à 79 ans au Québec, 2009**



\* Statistiquement différent entre l'hypertension des données médico-administratives de l'hypertension auto-déclarée de l'ESCC, p<0,05.  
 † Statistiquement différent entre l'hypertension des données médico-administratives de l'hypertension mesurée/médicaments de l'ECMS, p<0,05.  
 ‡ Statistiquement différent entre l'hypertension auto-déclarée de l'ESCC et l'hypertension mesurée/médicaments de l'ECMS, p<0,05.  
 Source : Système intégré de surveillance des maladies chroniques du Québec (Sismacq) de l'Institut national de santé publique du Québec (INSPQ) (données médico-administratives de l'année 2009-2010), Enquête sur la santé dans les collectivités canadiennes (ESCC) de Statistique Canada, année 2009 seulement, et Enquête canadienne sur les mesures de la santé (ECMS) de Statistique Canada, cycles 1 et 2 combinés (de 2007 à 2011).

découvrir leur HTA. L'augmentation de la prévalence de l'HTA observée dans cette étude peut être expliquée en partie par le fait que davantage de personnes sont dépistées et diagnostiquées par les médecins et, conséquemment, plus de personnes sont conscientes de leur condition. La diminution de la mortalité parmi les personnes hypertendues est une autre partie de l'explication<sup>10</sup>. Bien que les femmes de 65 à 79 ans soient plus hypertendues en 2009 selon les données médico-administratives et auto-déclarées, cette tendance tend à s'estomper dans l'année qui suit et ceci serait en partie attribuable à une meilleure prise en charge de l'HTA chez les hommes. Ces derniers voient ainsi leur prévalence rejoindre celle des femmes. Par ailleurs, il est bien démontré que les hommes ont moins conscience de leur HTA. Cela contribuerait possiblement à la sous-estimation, avec

les données médico-administratives et d'enquêtes auto-déclarées par rapport aux données d'enquêtes mesurées. En effet, selon les données mesurées de l'ECMS, ces derniers présentent plus d'HTA en comparaison des femmes. Enfin, en guise de comparaison, notons que la prévalence de l'HTA au Québec telle que mesurée en 2009 (22,6%) est plus faible que celle mesurée en France en 2006-2007 (31,0%) chez les 18 à 74 ans<sup>11</sup> ou aux États-Unis en 2007-2008 (30,7%) chez les personnes âgées de 18 ans et plus<sup>12</sup>. Une des principales explications pour cette plus faible prévalence au Québec serait les efforts canadiens pour prévenir et contrôler l'HTA. En effet, le Canada occupe la place de chef de file mondial pour la prévention, le traitement et le contrôle de l'HTA, suite à la mise en œuvre d'un plan stratégique national de lutte contre l'HTA au milieu des années 1990<sup>10</sup>.

## Limites

La prévalence de l'HTA peut être sous-estimée, car les personnes résidant en centre d'accueil ou dans d'autres types d'institutions de santé ne font pas partie de l'échantillon des trois sources de données. Les données auto-déclarées des deux enquêtes peuvent être sujettes à une mauvaise compréhension et au biais de rappel. De plus, la taille d'échantillon limitée des enquêtes, plus particulièrement l'ECMS, génère des estimations de faible précision et les tests statistiques décèlent ainsi moins souvent des différences significatives, spécialement pour les comparaisons entre les groupes d'âges et les sexes. Cependant, obtenir de plus grands échantillons avec des données mesurées s'avère plus coûteux que les données médico-administratives qui utilisent des données qui sont continuellement mises à jour et qui couvrent presque la population entière. Dans l'ECMS, même si la tension artérielle est mesurée six fois selon un protocole standard, ces mesures sont effectuées au cours d'une seule visite, ce qui est différent de la définition clinique de l'HTA qui implique des valeurs élevées mesurées en plusieurs circonstances. Cette différence pourrait possiblement surestimer la prévalence de l'HTA obtenue avec l'ECMS. Finalement, les données médico-administratives ne comprennent que les personnes qui sont en contact avec le système de santé, c'est-à-dire celles ayant consulté un médecin ou ayant été hospitalisées. De plus, en présence de multiples comorbidités, telles que par exemple le diabète, l'HTA pourrait ne pas figurer dans le fichier des services médicaux. Comme 75% des diabétiques sont également hypertendus, cette situation peut être très fréquente<sup>13</sup>.

## Conclusion

Les trois sources de données étudiées révèlent une prévalence similaire de l'HTA au Québec en 2009, malgré une légère sous-estimation de celle obtenue avec les données auto-déclarées. La présente étude démontre bien l'augmentation de la prévalence de l'HTA diagnostiquée ou déclarée depuis 2000-2001, et ce, autant avec les données médico-administratives qu'avec les données d'enquêtes. Néanmoins, les données médico-administratives du Québec devraient être davantage utilisées pour la surveillance des diverses maladies cardiovasculaires, comme les cardiopathies ischémiques ou les maladies vasculaires cérébrales. En effet, ces données sont accessibles et constamment mises à jour et ce, pour tous les Québécois, peu importe leur âge et leur statut socioéconomique. De surcroît, les données médico-administratives jumelées permettent la détermination de l'incidence, la mortalité et la consommation de médicaments. Les résultats de la présente étude sont importants pour la planification des soins de

santé et la santé publique, et rappellent le besoin de réduire le lourd fardeau de l'HTA chez les Québécois, et spécialement chez les personnes âgées. ■

## Remerciements

Les auteurs remercient Carole Blanchet, de l'Institut national de santé publique du Québec, pour la révision de l'article.

## Références

- [1] World Health Organization. Global health risks: Mortality and burden of disease attributable to selected major risks. Geneva: WHO;2009. 70 p.
- [2] Robitaille C, Dai S, Waters C, Loukine L, Bancej C, Quach S, *et al.* Diagnosed hypertension in Canada: incidence, prevalence and associated mortality. *CMAJ*. 2012;184(1):E49-E56.
- [3] Statistique Canada, 2013. Enquête sur la santé dans les collectivités canadiennes (ESCC) [Internet]. Gouvernement du Canada, Ottawa; 2013. [http://www23.statcan.gc.ca/imdb/p2SV\\_f.pl?Function=getSurvey&SDDS=3226&Item\\_Id=144171&lang=fr](http://www23.statcan.gc.ca/imdb/p2SV_f.pl?Function=getSurvey&SDDS=3226&Item_Id=144171&lang=fr)
- [4] Giroux S. Canadian Health Measures Survey: sampling strategy overview. *Health Rep*. 2007;18 Suppl:31-6.
- [5] Campbell NR, Joffres MR, McKay DW. Hypertension surveillance in Canada: minimum standards for assessing blood pressure in surveys. *Can J Public Health*. 2005;96(3):217-20.
- [6] Wilkins K, Campbell NR, Joffres MR, McAlister FA, Nichol M, Quach S, *et al.* Blood pressure in Canadian adults. *Health Rep*. 2010;21(1):37-46.
- [7] Bryan S, Saint-Pierre LM, Campbell N, Clarke J, Tremblay MS. Resting blood pressure and heart rate measurement in the Canadian Health Measures Survey, cycle 1. *Health Rep*. 2010;21(1):71-8.
- [8] Tu K, Campbell NRC, Chen ZL, Cauch-Dudek KJ, McAlister FA. Accuracy of administrative databases in identifying patients with hypertension. *Open Med*. 2007;1(1):E18-E26.
- [9] Quan H, Khan N, Hemmelgarn BR, Tu K, Chen G, Campbell N, *et al.* Validation of a case definition to define hypertension using administrative data. *Hypertension*. 2009;54(6):1423-8.
- [10] Campbell NR, Chen G. Canadian efforts to prevent and control hypertension. *Can J Cardiol*. 2010;26 Suppl C:14C-7C.
- [11] Godet-Thobie H, Vernay M, Noukpoape A, Salanave B, Malon A, Castetbon K, *et al.* Niveau tensionnel moyen et prévalence de l'hypertension artérielle chez les adultes de 18 à 74 ans, ENNS 2006-2007. *Bull Epidémiol Hebd*. 2008;(49-50):478-83. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=1728](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=1728)
- [12] Crim MT, Yoon SS, Ortiz E, Wall HK, Schober S, Gillespie C, *et al.* National surveillance definitions for hypertension prevalence and control among adults. *Circ Cardiovasc Qual Outcomes*. 2012;5(3):343-51.
- [13] Gee ME, Janssen I, Pickett W, McAlister FA, Bancej CM, Joffres M, *et al.* Prevalence, awareness, treatment, and control of hypertension among Canadian adults with diabetes, 2007 to 2009. *Can J Cardiol*. 2012;28(3):367-74.

## Citer cet article

Blais C, Rochette L, Hamel D. Prévalence de l'hypertension artérielle au Québec : comparaison entre les données médico-administratives et les données d'enquêtes. *Bull Epidémiol Hebd*. 2013;(Hors-série):23-9.

## L'APPORT DES BASES MÉDICO-ADMINISTRATIVES DANS LA SURVEILLANCE DES MALADIES RESPIRATOIRES CHRONIQUES EN FRANCE

// THE CONTRIBUTION OF HEALTH ADMINISTRATIVE DATABASES IN THE SURVEILLANCE OF CHRONIC RESPIRATORY DISEASES IN FRANCE

Marie-Christine Delmas (mc.delmas@invs.sante.fr), Marjorie Boussac-Zarebska, Marie Houot, Delphine Lauzeille, Yuriko Iwatsubo, Claire Fuhrman

Institut de veille sanitaire, Saint-Maurice, France

Soumis le 11/06/2013 // Date of submission: 06/11/2013

### Résumé // Abstract

Les bases de données médico-administratives (BMA) sont un des outils de la surveillance épidémiologique des maladies respiratoires chroniques en France. Les données sur la morbidité hospitalière sont fournies par le Programme de médicalisation des systèmes d'information (PMSI). L'utilisation des données de remboursement de l'assurance maladie est plus récente. L'objet de cet article est d'illustrer, à partir de quelques exemples d'exploitation de ces données, leur apport dans la surveillance des maladies respiratoires chroniques en France.

L'analyse de ces données montre le poids croissant de la bronchopneumopathie chronique obstructive (BPCO) en France. Les taux d'hospitalisation pour BPCO ont augmenté entre 1998 et 2010, avec en parallèle une diminution de la létalité intra-hospitalière des patients hospitalisés pour BPCO. Les données de remboursement de l'assurance maladie montrent également une augmentation entre 2006 et 2010 de la prévalence du traitement par oxygénothérapie de longue durée. Les données du PMSI permettent de documenter des disparités géographiques marquées dans les taux d'hospitalisation ou dans les taux de réadmission pour maladie respiratoire chronique. La surveillance des risques professionnels des maladies respiratoires chroniques implique de disposer, en plus, des données professionnelles des personnes. C'est le cas des données du Régime social des indépendants (RSI) à partir desquelles une étude de faisabilité a été conduite, visant à repérer les personnes asthmatiques à partir des consommations médicamenteuses et ainsi comparer les risques d'asthme par secteurs d'activité.

Les différents exemples présentés dans cet article témoignent de la richesse des BMA et de leur apport dans la surveillance des maladies respiratoires chroniques.

*Health administrative databases are one of the tools for epidemiological surveillance of chronic respiratory diseases (CRD) in France. Hospital morbidity data are provided by the Hospital Discharge Data System (PMSI). The use of data from health insurance reimbursements is more recent. The purpose of this article is to illustrate, using a few examples, these data contribution to the monitoring of CRDs in France.*

*The analysis of these data shows the growing importance of chronic obstructive pulmonary disease (COPD) in France. Hospitalization rates for COPD increased between 1998 and 2010, while intra-hospital mortality of patients hospitalized for COPD decreased. Data from health insurance reimbursements also show an increase in the prevalence of long term oxygen therapy between 2006 and 2010. The PMSI data contribute to document geographic disparities in rates of hospitalization or readmission for CRDs. Surveillance of occupational hazards for CRD involves the availability of personal occupational data. This is the case of the data from the social security scheme for self-employed workers, from which a feasibility study was conducted to identify asthma patients based on their drug consumption and compare the risk of asthma by activity sectors.*

*The examples discussed in the article show the richness of health administrative databases, and their contribution to the monitoring of CRDs.*

**Mots-clés :** Maladies respiratoires chroniques, Surveillance, Bases médico-administratives, France

// **Keywords:** Chronic respiratory diseases, Monitoring, Health administrative databases, France

### Introduction

Un système de surveillance épidémiologique est un outil permanent de mesure et de suivi de l'état de santé des populations. La surveillance des maladies respiratoires chroniques en France porte principalement

sur deux pathologies qui représentent un lourd poids en termes de santé publique : l'asthme et la bronchopneumopathie chronique obstructive (BPCO). Ces pathologies comportent des déterminants multiples, parmi lesquels les déterminants professionnels qui méritent une surveillance épidémiologique spécifique.

Le syndrome d'apnées du sommeil (SAS), pathologie fréquente et consommatrice de ressources tant pour son diagnostic que pour son traitement, a également été intégré au dispositif de surveillance qui est construit autour de différentes sources de données. Les bases médico-administratives (BMA), qui sont des bases qui ont été créées à des fins de gestion budgétaire, sont une des sources de données de cette surveillance. Elles comprennent les données sur les hospitalisations produites dans le cadre du Programme de médicalisation des systèmes d'information (PMSI) ainsi que les données de remboursement de l'assurance maladie. L'objet du présent article est de fournir des exemples sur l'utilisation de ces BMA dans la surveillance épidémiologique des maladies respiratoires chroniques en France et d'en discuter les limites.

## PMSI et surveillance des maladies respiratoires chroniques

Mis en place dans les années 1990 avec pour finalité l'analyse médico-économique de l'activité hospitalière, le PMSI consiste en un enregistrement de données médico-administratives sur les séjours effectués dans les établissements hospitaliers publics et privés français. Les données administratives incluent le sexe, la date de naissance et le code postal de résidence de la personne, les dates et modes d'entrée et de sortie ainsi que l'identification de l'établissement et le type d'unités médicales fréquentées pendant le séjour à l'hôpital. Les informations médicales comportent le diagnostic principal (DP), les diagnostics reliés (DR) et associés significatifs (DAS) ainsi que les actes médicaux. Les diagnostics sont codés avec la 10<sup>e</sup> révision de la Classification internationale des maladies (CIM-10). Jusqu'en 2008, le DP représentait le motif de soins ayant mobilisé l'essentiel de l'effort médical et soignant ; depuis 2009, il s'agit du diagnostic ayant motivé l'hospitalisation. Depuis 2002, un code anonyme unique (à l'exception des jumeaux du même sexe) est généré pour chaque patient. Il

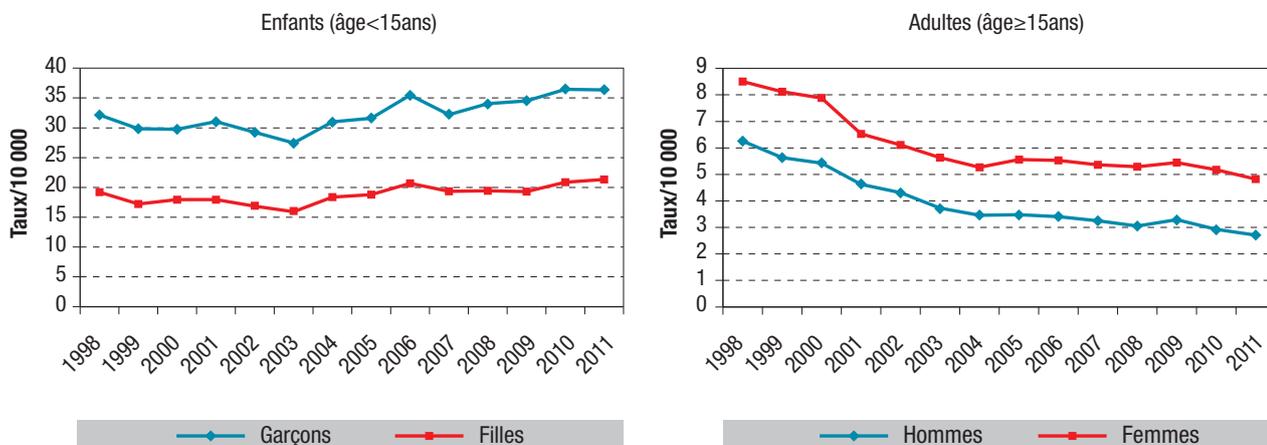
permet de chaîner ses séjours dans tous les établissements publics ou privés, et ce tant que le patient est assuré par le même ouvrant droit. Des informations détaillées sur les données disponibles, les nomenclatures et les règles de codage sont disponibles sur le site de l'Agence technique de l'information sur l'hospitalisation (Atih) (<http://www.atih.sante.fr>).

Les hospitalisations pour asthme sont considérées comme étant en grande partie évitables grâce à une prise en charge adaptée, notamment la prise d'un traitement de fond (corticoïdes inhalés) pour les asthmes persistants. De même, une prise en charge optimale dès les premiers signes d'aggravation pourrait permettre d'éviter une partie des hospitalisations pour exacerbation de BPCO. Les données du PMSI permettent de décrire l'évolution des taux d'hospitalisation ainsi que l'évolution de données indirectes de sévérité : durée de séjour, nombre d'unités fréquentées, hospitalisation en unité de soins intensifs, ventilation assistée et létalité intra-hospitalière. Elles comprennent cependant un certain nombre de limites. Certaines hospitalisations pour exacerbations d'asthme ou de BPCO peuvent ne pas être reconnues comme telles. Un biais de classification est également possible, que ce soit chez le jeune enfant entre asthme et bronchiolite ou chez l'adulte entre asthme et BPCO. D'autre part, concernant la BPCO pour laquelle les comorbidités sont fréquentes et le sous-diagnostic important, le codage de cette pathologie peut ne pas apparaître en DP ou même en DAS. Les passages aux urgences qui ne sont pas suivis d'une hospitalisation, ce qui est fréquent dans le cas de l'asthme, ne sont pas intégrés dans le PMSI. Enfin, les admissions programmées pour évaluation ou investigation ne peuvent pas être clairement identifiées.

Les données du PMSI montrent que les taux annuels d'hospitalisation pour asthme (asthme (J45) ou asthme aigu grave (J46) en DP) ont diminué entre 1998 et 2011 chez l'adulte mais pas chez l'enfant (figure 1). Pour l'étude des exacerbations de BPCO,

Figure 1

### Taux annuels\* d'hospitalisation pour asthme, France métropolitaine, 1998-2011



\* Taux standardisés sur l'âge, population française 2006 (Insee). Source : PMSI-ATIH.

les séjours de moins de 2 jours dont l'issue n'est pas le décès sont exclus afin d'éviter de prendre en compte les hospitalisations programmées<sup>1</sup>. L'évolution entre 1998 et 2011 des taux d'hospitalisation standardisés sur l'âge montre un poids croissant de la BPCO en termes de morbidité hospitalière, chez les hommes (+2,4% par an) comme chez les femmes (+5,8% par an) (figure 2). L'analyse en termes de patients montre des tendances similaires. On observe également une diminution du ratio hommes/femmes (de 2,6 à 1,9), une diminution de la létalité intra-hospitalière (de 7,6% à 6,4%) alors que l'âge moyen des personnes hospitalisées augmente (de 72,0 à 73,4 ans).

Les données du PMSI montrent également des disparités régionales marquées dans la morbidité hospitalière liée aux maladies respiratoires chroniques. Concernant l'asthme de l'enfant, les taux d'hospitalisations les plus élevés sont observés dans les départements d'outre-mer, dans le quadrant Nord de la France, ainsi que dans le sud de la façade atlantique<sup>2</sup>. C'est également dans la région Nord-Pas-de-Calais que la proportion d'enfants réadmis au cours de la même année calendaire est la plus élevée.

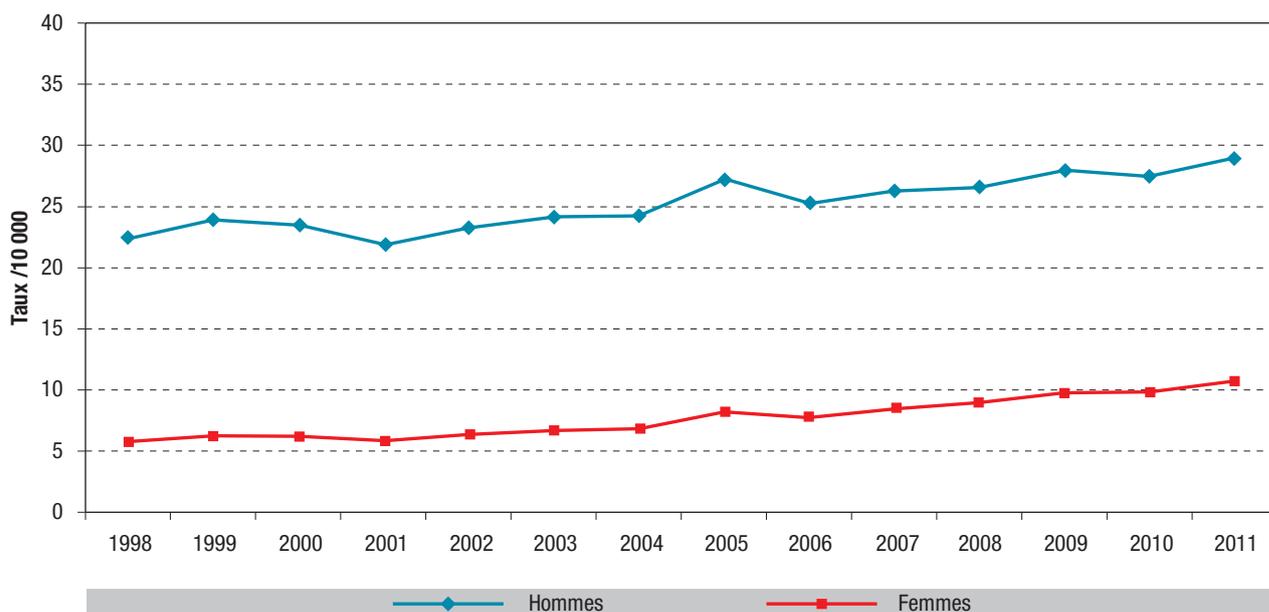
## Données de l'assurance maladie et surveillance des maladies respiratoires chroniques

### Le Sniiram

Le système national d'information inter-régimes de l'assurance maladie (Sniiram) regroupe l'ensemble des données de remboursement de l'assurance maladie. Deux entrepôts de données individuelles sont disponibles : le Datamart de consommation inter-régimes (DCIR) et l'Échantillon généraliste de bénéficiaires (EGB). Le DCIR couvre, à ce jour, la quasi-totalité des régimes d'assurance maladie, notamment le régime général, le régime agricole (qui couvre les exploitants et salariés agricoles) et le régime des indépendants (qui couvre les artisans, commerçants, industriels et professions libérales), ainsi que la plupart des régimes spéciaux, par exemple, le régime de la Société nationale des chemins de fer français (SNCF). Seules les informations concernant les bénéficiaires ayant consommé des soins sont disponibles et les données sont conservées trois années plus l'année en cours. L'EGB a été créé en 2005 à partir du Sniiram. Cet échantillon

Figure 2

### Taux annuels\* d'hospitalisation pour exacerbation de bronchopneumopathie chronique obstructive\*\* chez les adultes âgés de 25 ans ou plus, France métropolitaine, 1998-2011



\* Taux standardisés sur l'âge, population française 2006 (Insee). Source : PMSI-ATIH.

\*\*Exacerbation de BPCO (les séjours de moins de 2 jours dont l'issue n'est pas le décès sont exclus) :

- Diagnostic principal (DP) de maladie pulmonaire obstructive chronique avec infection aiguë des voies respiratoires (J44.0) ou avec épisode aigu non précisé (J44.1)
- ou
- DP d'insuffisance respiratoire aiguë (J96.0) avec un diagnostic associé significatif (DAS) de maladie pulmonaire obstructive chronique (J44) ou d'emphysème (J43)
- ou
- DP d'infection respiratoire basse (J10-J18, J20-J22) avec un DAS de maladie pulmonaire obstructive chronique (J44) ou d'emphysème (J43)
- ou
- DP de maladie pulmonaire obstructive chronique (J44) ou d'emphysème (J43) avec un diagnostic associé d'infection respiratoire basse (J10-J18, J20-J22) ou d'insuffisance respiratoire aiguë (J96.0).

permanent résulte d'un sondage au 1/97<sup>e</sup> des bénéficiaires de l'assurance maladie, qu'ils aient ou non reçu des remboursements de soins. Jusqu'en 2010, il incluait uniquement les bénéficiaires du régime général hors sections locales mutualistes (organismes mutualistes gérant la part obligatoire d'assurance maladie pour leurs adhérents, notamment les étudiants et fonctionnaires), soit environ 75% de la population résidant en France. Les bénéficiaires du Régime social des indépendants (RSI) et de la Mutualité sociale agricole (MSA), qui représentent chacun environ 5% de la population française, y sont inclus depuis 2011. À terme, l'EGB comportera jusqu'à 20 ans d'historique.

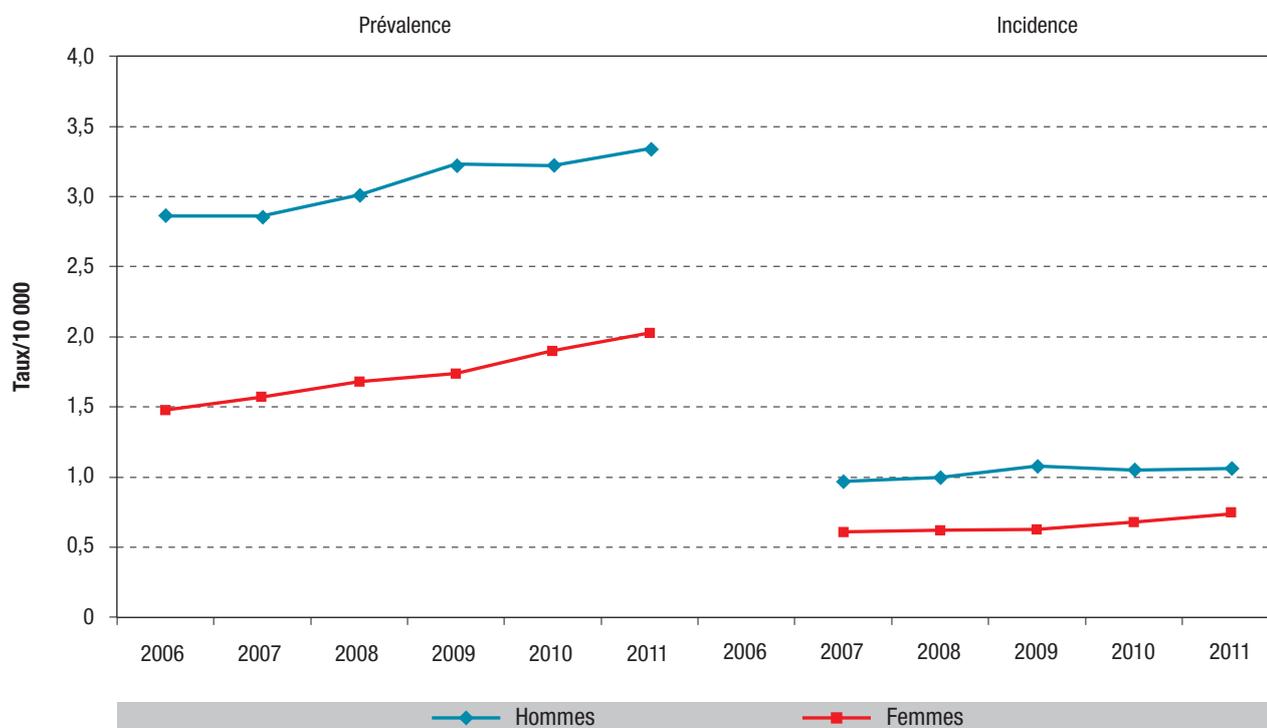
Les données de remboursement disponibles dans le DCIR et dans l'EGB concernent les soins de ville (médicaments, actes médicaux, actes de biologie, dispositifs médicaux, prestations de transports...). Les données sociodémographiques comportent le sexe, l'âge, la commune de résidence, la caisse d'affiliation, l'affiliation à la couverture maladie universelle (indicateur de faible niveau socioéconomique pour les personnes de moins de 60 ans) et la date de décès le cas échéant. Un diagnostic médical (codé selon la CIM-10) est disponible pour les personnes prises en charge à 100% par l'assurance maladie dans le cadre d'une affection longue durée (ALD) ou d'une maladie professionnelle. Toutefois, cette information est difficile à interpréter. En effet, pour certaines maladies comme l'asthme, la demande

de prise en charge à 100% au titre de l'insuffisance respiratoire grave n'est pas systématique. Pour la BPCO au cours de laquelle les comorbidités sont fréquentes, la demande n'est pas toujours effectuée si la personne est déjà prise en charge pour une autre maladie (insuffisance cardiaque par exemple). Les données de l'EGB et du DCIR sont complétées par les données d'hospitalisations (PMSI) fournies par l'ATIH et, depuis 2009, par les données d'activité externe des hôpitaux (consultations et soins effectués à l'hôpital en dehors d'une hospitalisation ou d'un passage aux urgences).

Les données de l'EGB sont utilisées pour estimer la prévalence et l'incidence des mises sous oxygénothérapie de longue durée (OLD) et du traitement du SAS par pression positive continue (PPC)<sup>3,4</sup>. En ce qui concerne l'OLD, la prévalence a augmenté entre 2006 et 2011 (de 2,9% par an chez les hommes et de 6,7% par an chez les femmes) (figure 3). L'incidence des mises sous OLD est restée stable chez les hommes alors qu'elle a augmenté de 4,7% par an chez les femmes. La médiane de survie était de 27 mois. On a observé également une nette augmentation de la prévalence du traitement par PPC (+15% par an entre 2006 et 2011). La principale limite de cette source de données est le manque d'information sur les diagnostics médicaux ainsi que sur la gravité. Ainsi, moins de la moitié des personnes sous OLD étaient prises en charge en ALD pour insuffisance respiratoire grave. Concernant la gravité, aucune information sur le degré

Figure 3

**Prévalence et incidence\* annuelle des mises sous oxygénothérapie de longue durée chez les adultes (≥20 ans). Échantillon généraliste de bénéficiaires (régime général hors sections locales mutualistes), France, 2006-2011**



\* Taux standardisés sur l'âge, population française 2006 (Insee). Source : EGB, Caisse nationale de l'assurance maladie des travailleurs salariés (CnamTS).

d'hypoxie à la mise sous OLD n'est disponible. De plus, les données de consommation de soins ne sont pas disponibles pour les personnes hospitalisées en long séjour ou résidant en établissement d'hébergement pour personnes âgées dépendantes (Ehpad).

Les enquêtes déclaratives en population générale ou chez les patients consultants montrent que le contrôle de l'asthme est souvent insuffisant. Les données de l'EGB permettent de suivre les tendances des remboursements de médicaments de la classe ATC R03 (médicaments pour les syndromes obstructifs des voies aériennes) et fournissent ainsi des indications complémentaires sur la qualité de la prise en charge des maladies respiratoires chroniques obstructives. Elles montrent une augmentation, entre 2006 et 2011, des taux de consommateurs de 2-mimétiques inhalés de courte durée d'action (2-CD) et de corticoïdes inhalés (CI) (figure 4). Toutefois, alors que les personnes ayant une forte consommation de 2-CD devraient bénéficier d'un traitement par CI afin de contrôler les symptômes, la proportion de forts consommateurs de 2-CD ( $\geq 4$  boîtes dans l'année) qui n'ont pas consommé ou qui ont eu moins de 3 boîtes de CI dans l'année, est restée stable, aux alentours de 32%.

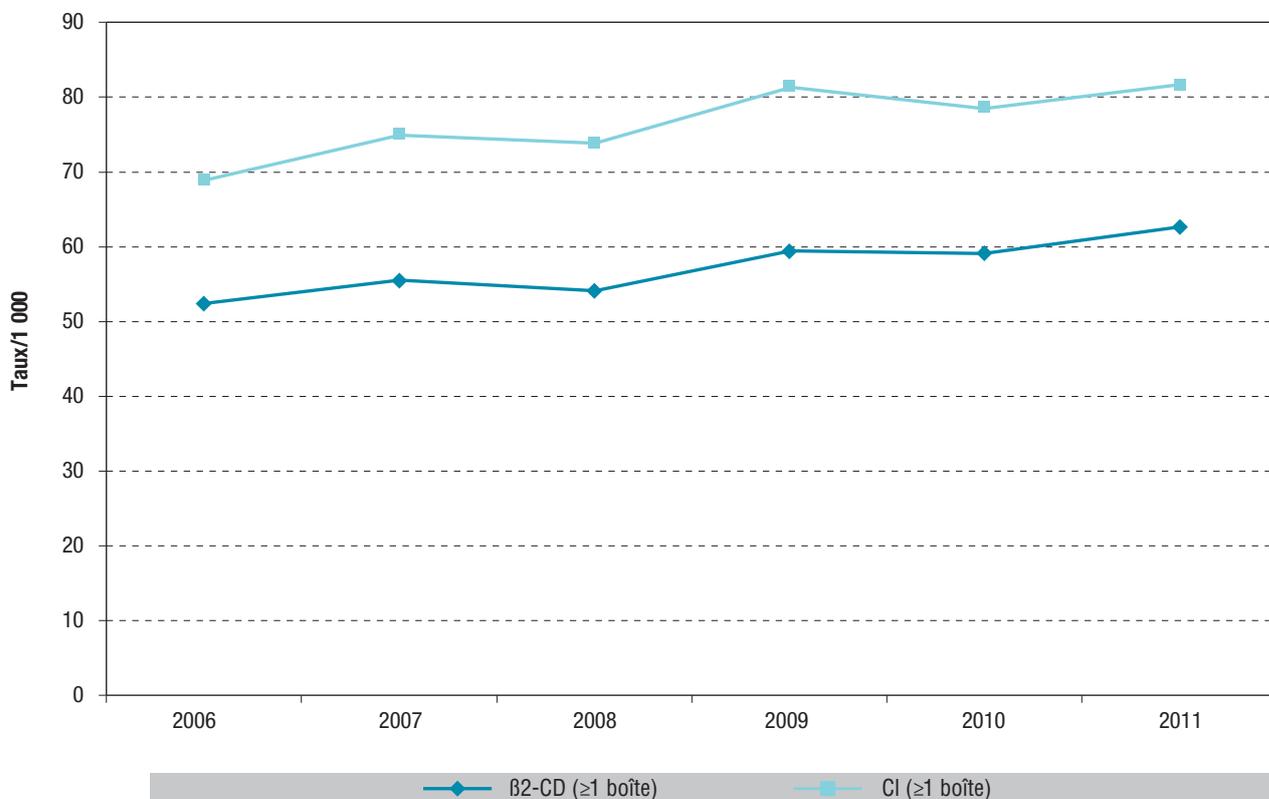
### Les données du Régime social des indépendants (RSI)

Si l'on s'intéresse à la surveillance des risques professionnels, les données du Sniiram qui donnent une

indication sur l'état de santé ne sont pas suffisantes. En effet, il faut également disposer de données professionnelles. C'est dans ce cadre que la faisabilité de repérage des sujets asthmatiques, à partir des données de remboursement de médicaments « anti-asthmatiques », a été étudiée chez des travailleurs affiliés au RSI qui dispose de ces deux types de données<sup>5</sup>. L'objectif de cette étude pilote était de décrire l'asthme en fonction des secteurs d'activité économique dans cette population. Pour cela, un modèle prédictif de l'asthme actuel a été élaboré à partir des données de remboursement de médicaments « anti-asthmatiques » et sur la consultation d'un pneumologue, en prenant comme référence la déclaration des sujets en réponse à un auto-questionnaire. Ce modèle a permis de classer les sujets en asthmatique ou non-asthmatique avec une sensibilité de 80% et une spécificité de 86%. La prévalence de l'asthme actuel a été calculée et les risques étudiés en fonction des secteurs d'activité, tels que définis par la nomenclature de l'Insee. La prévalence de l'asthme ainsi obtenue est égale à 3,4% (proche de celle obtenue par autoquestionnaire de 3,9%). Dans cette population, les secteurs d'activité les plus à risque sont la boulangerie-pâtisserie (OR=2,8 ; [IC95%:2,1-3,7]), la fabrication d'appareils médicochirurgicaux (OR=1,9 [1,0-3,7]), le transport de voyageurs par taxi (OR=1,5 [1,0-2,2]) et la restauration de type traditionnel (OR=1,3 [1,0-1,7]). Ce modèle

Figure 4

#### Taux de consommateurs de 2-mimétiques inhalés de courte durée d'action (2-CD) et de corticoïdes inhalés (CI). Données de l'Échantillon généraliste de bénéficiaires, France, 2006-2011



Source : EGB, Caisse nationale de l'assurance maladie des travailleurs salariés (CnamTS).

devra faire l'objet d'une validation sur un autre échantillon avant de pouvoir être utilisé dans d'autres populations de travailleurs. Si le résultat est concluant, il sera alors envisageable de mettre en place un système de surveillance de l'asthme actuel chez des travailleurs actifs à partir des données de remboursement de médicaments, sans avoir à interroger les sujets, sachant cependant que cette méthode ne pourrait s'appliquer qu'aux personnes asthmatiques traitées.

## Discussion

Les exemples présentés dans cet article témoignent de la richesse des BMA et de leur apport dans la surveillance épidémiologique des maladies respiratoires chroniques. La question des performances du PMSI dans l'identification des hospitalisations pour exacerbation d'asthme ou de BPCO se pose bien évidemment. Le retour au dossier médical pour un échantillon représentatif de cas repérés dans le PMSI permettrait de calculer une valeur prédictive positive (VPP). En revanche, pour une maladie aussi fortement sous-diagnostiquée que la BPCO, l'utilisation du dossier médical pour repérer les faux négatifs n'est pas pertinente. Par ailleurs, la procédure de validation nécessiterait d'être répétée régulièrement. Malgré leurs limites, les données du PMSI fournissent des indications sur le poids de la morbidité hospitalière liée à l'asthme et à la BPCO et sur ses tendances spatio-temporelles.

Le manque d'informations sur les diagnostics médicaux est une des principales limites à l'interprétation des données portant sur les remboursements de soins par l'assurance maladie, les seuls diagnostics disponibles étant ceux figurant, le cas échéant, dans le PMSI ou au titre d'une ALD. Or les médicaments donnés dans le traitement de l'asthme ne sont pas spécifiques. Ils sont en effet prescrits dans d'autres maladies respiratoires chroniques (BPCO notamment), mais aussi dans des pathologies aiguës. L'étude pilote réalisée chez les travailleurs indépendants a cherché à estimer la prévalence de l'asthme actuel par la construction d'un algorithme qui utilise les données de remboursement de médicaments, l'objectif étant de comparer le risque d'asthme actuel selon les secteurs d'activité. En utilisant le modèle prédictif, on retrouve un risque élevé dans l'industrie alimentaire, ce qui était attendu, mais également dans des secteurs non à risque *a priori*, comme le transport de voyageurs par taxi. Ce résultat devra être confirmé dans une population de taille beaucoup plus importante. Ce travail est prévu avec les données de l'ensemble des actifs affiliés au RSI. Toutefois, il est important de rappeler que l'analyse du risque d'asthme par secteur d'activité réalisée ici avec le modèle prédictif repose sur l'hypothèse d'un recours aux soins et de prescriptions similaires selon les catégories professionnelles et que la méthode

choisie ne s'applique qu'aux personnes asthmatiques traitées. Il convient également de rappeler que les données du Sniiram ne comportent pas de données professionnelles. Pour une surveillance des risques professionnels, il sera donc nécessaire de relier les données du Sniiram à des données professionnelles. Les données de remboursement de médicaments fournissent également des indicateurs indirects sur le contrôle de l'asthme. Avec pour l'instant un très faible recul (6 ans de données), les données de l'EGB ne montrent pas d'éléments en faveur d'une amélioration de la prise en charge de l'asthme. En revanche, elles mettent en évidence, au travers de l'augmentation des prises en charge par OLD, une augmentation de la morbidité respiratoire sévère.

En conclusion, en fournissant des indicateurs sur la morbidité respiratoire, les BMA sont des outils précieux pour la surveillance épidémiologique des maladies respiratoires chroniques. Toutefois, seule l'interprétation conjointe des données issues de différentes sources (enquête déclarative, enquête avec examen de santé, hospitalisations, causes médicales de décès...) permet d'appréhender les différentes facettes de la morbidité liée aux maladies respiratoires chroniques. Le couplage de données d'enquête avec les données de l'assurance maladie ouvre des perspectives intéressantes, par exemple pour étudier les recours aux soins des personnes atteintes de BPCO mais non diagnostiquées, ou pour étudier la prévalence de pathologies respiratoires selon des caractéristiques professionnelles. ■

## Références

- [1] Fuhrman C, Roche N, Vergnenègre A, Zureik M, Chouaid C, Delmas MC. Hospital admissions related to acute exacerbations of chronic obstructive pulmonary disease in France, 1998-2007. *Respir Med.* 2011;105:595-601.
- [2] Delmas MC, Marguet C, Nicolau J, Raheison C, Fuhrman C. Les hospitalisations pour asthme chez l'enfant en France, 2002-2010. *Arch Pediatr.* 2013;20:739-47.
- [3] Fuhrman C, Boussac-Zarebska M, Roche N, Delmas MC. Oxygénothérapie de longue durée en France, 2006-2011. *Revue des Maladies Respiratoires* (sous presse).
- [4] Fuhrman C, Nguyễn XL, Fleury B, Boussac-Zarebska M, Druet C, Delmas MC. Le syndrome d'apnées du sommeil en France: un syndrome fréquent et sous-diagnostiqué. *Bull Epidémiol Hebd.* 2012;(44-45):510-4. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=11093](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=11093)
- [5] Iwatsubo Y, Lauzeille D, Houot M, Mevel M, Chabault E, Delabre L, *et al.* Surveillance épidémiologique de l'asthme au sein de la population des artisans et commerçants affiliés au Régime social des indépendants (RSI). Résultats de l'étude pilote. *Saint-Maurice: Institut de veille sanitaire; 2012.* 127 p. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=10936](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=10936)

## Citer cet article

Delmas MC, Boussac-Zarebska M, Houot M, Lauzeille D, Iwatsubo Y, Fuhrman C. L'apport des bases médico-administratives dans la surveillance des maladies chroniques en France. *Bull Epidémiol Hebd.* 2013;(Hors-série):30-5.

## DÉVELOPPEMENT DE LA SURVEILLANCE DES MALADIES RESPIRATOIRES AU QUÉBEC À PARTIR DES DONNÉES MÉDICO-ADMINISTRATIVES

// DEVELOPMENT OF A QUEBEC RESPIRATORY DISEASE SURVEILLANCE SYSTEM FROM HEALTH ADMINISTRATIVE DATA

Mariève Doucet<sup>1,2</sup> (marieve.doucet@inspq.qc.ca), Louis Rochette<sup>1</sup>, Myriam Elizabeth Gagné<sup>1</sup>, Philippe Gamache<sup>1</sup>, Sophie Pouliot<sup>1</sup>, Valérie Émond<sup>1</sup>

<sup>1</sup> Institut national de santé publique du Québec (INSPQ), Québec, Canada

<sup>2</sup> Faculté de médecine, Université Laval, Québec, Canada

Soumis le 19/07/2013 // Date of submission: 07/19/2013

### Résumé // Abstract

Au Québec, la surveillance de l'asthme et de la broncho-pneumopathie chronique obstructive (BPCO) constitue une priorité de santé publique. S'appuyant sur un cadre conceptuel, le volet « maladies respiratoires » du Système intégré de surveillance des maladies chroniques du Québec (Sismacq) est construit à partir de données issues du jumelage de fichiers médico-administratifs. L'objectif de cet article est de présenter les huit étapes de développement de ce volet.

Ces étapes comprennent : 1) l'assurance de la qualité des données par des études de faisabilité ; 2) la définition de l'objectif de surveillance : la communication des messages de surveillance pour l'action en santé publique ; 3) la détermination des publics cibles : les décideurs et acteurs de surveillance principalement ; 4) le développement des premiers messages de surveillance à communiquer sur la prévalence et l'incidence de l'asthme et de la BPCO ; 5) la diffusion de ces messages auprès des publics cibles, notamment par voie électronique ; 6) la mise sur pied d'un comité d'usagers pour faire la promotion des messages auprès de publics cibles ; 7) la mise en application ainsi que 8) l'évaluation du système. Toutes ces étapes posent plusieurs défis pour les années à venir.

*In Quebec, asthma and chronic obstructive pulmonary disease (COPD) surveillance has been highlighted as a public health priority. Based on a conceptual framework, the "respiratory disease" component of the Quebec Integrated Chronic Disease Surveillance System is built from data extracted from linked health administrative datasets. The objective of this article is to present the eight stages of development of this component.*

*These stages are: 1) ensuring data quality by conducting feasibility studies; 2) defining the purpose of the surveillance system, which is the communication of surveillance products for public health actions; 3) defining target audiences, namely decision-makers and surveillance practitioners; 4) developing surveillance products, such as asthma and COPD incidence and prevalence rates and trends; 5) selecting communication media and channels, mainly electronic; 6) setting up a stakeholder committee to market surveillance products; 7) implementing the system, and 8) evaluating it. In the next years, passing through these stages will require us to rise to many challenges.*

**Mots-clés :** Surveillance, Maladies respiratoires, Asthme, Broncho-pneumopathie chronique obstructive, Données médico-administratives, Québec

// **Keywords:** Surveillance, Respiratory diseases, Asthma, Chronic obstructive pulmonary disease, Health administrative data, Quebec

### Introduction

Au Canada, l'asthme et la broncho-pneumopathie chronique obstructive (BPCO) sont les maladies respiratoires chroniques ayant les prévalences les plus élevées<sup>1</sup>. Au Québec, la surveillance de ces deux pathologies a été identifiée comme une priorité de santé publique. De fait, l'Institut national de santé publique du Québec (INSPQ) s'est vu confier, par le ministère de la Santé et des Services sociaux (MSSS) du Québec, le mandat de développer le volet « maladies respiratoires » du Système intégré de surveillance des maladies chroniques du Québec (Sismacq), un système basé sur le jumelage des données médico-administratives (DMA).

### Le cadre théorique de surveillance utilisé

Le choix du cadre théorique est important afin de faire une surveillance populationnelle adéquate de l'asthme et de la BPCO. Comme il n'existe aucun cadre spécifique à la surveillance des maladies respiratoires, nous avons choisi un cadre général de surveillance que nous avons adapté à ces conditions et au contexte québécois. Ce choix reposait sur les critères suivants : le cadre devait 1) avoir été développé spécifiquement pour la surveillance en santé publique, 2) être adaptable à un système de surveillance utilisant les DMA. Le cadre théorique américain de Remington et Nelson, qui répond à ces deux critères, est utilisé pour

la surveillance menée aux *Centers for Disease Control and Prevention* (Atlanta, États-Unis)<sup>2</sup>. Il est présenté en huit étapes (figure 1). Les six premières étapes ont permis la conceptualisation du volet « maladies respiratoires » du Sismacq ; les deux dernières concernent sa mise en application et son évaluation.

L'objectif de cet article est de présenter chacune des huit étapes permettant le développement du volet « maladies respiratoires » du Sismacq et d'en identifier les défis et enjeux.

## Étapes de développement du volet « maladies respiratoires » du Sismacq

### Assurer la qualité des données de surveillance (étape 1)

Assurer la qualité des données est une étape cruciale dans le développement d'un système de surveillance basé sur l'utilisation des DMA. Au Canada, plusieurs provinces utilisent déjà les DMA aux fins de la

surveillance des maladies respiratoires<sup>3</sup>. Au Québec, ces sources de données sont : le fichier des personnes assurées, le fichier des services médicaux rémunérés à l'acte, le fichier des services pharmaceutiques, le fichier des hospitalisations et le fichier des décès. Ces différents fichiers sont décrits dans l'article de D. Saint-Laurent et coll. publié dans ce même numéro.

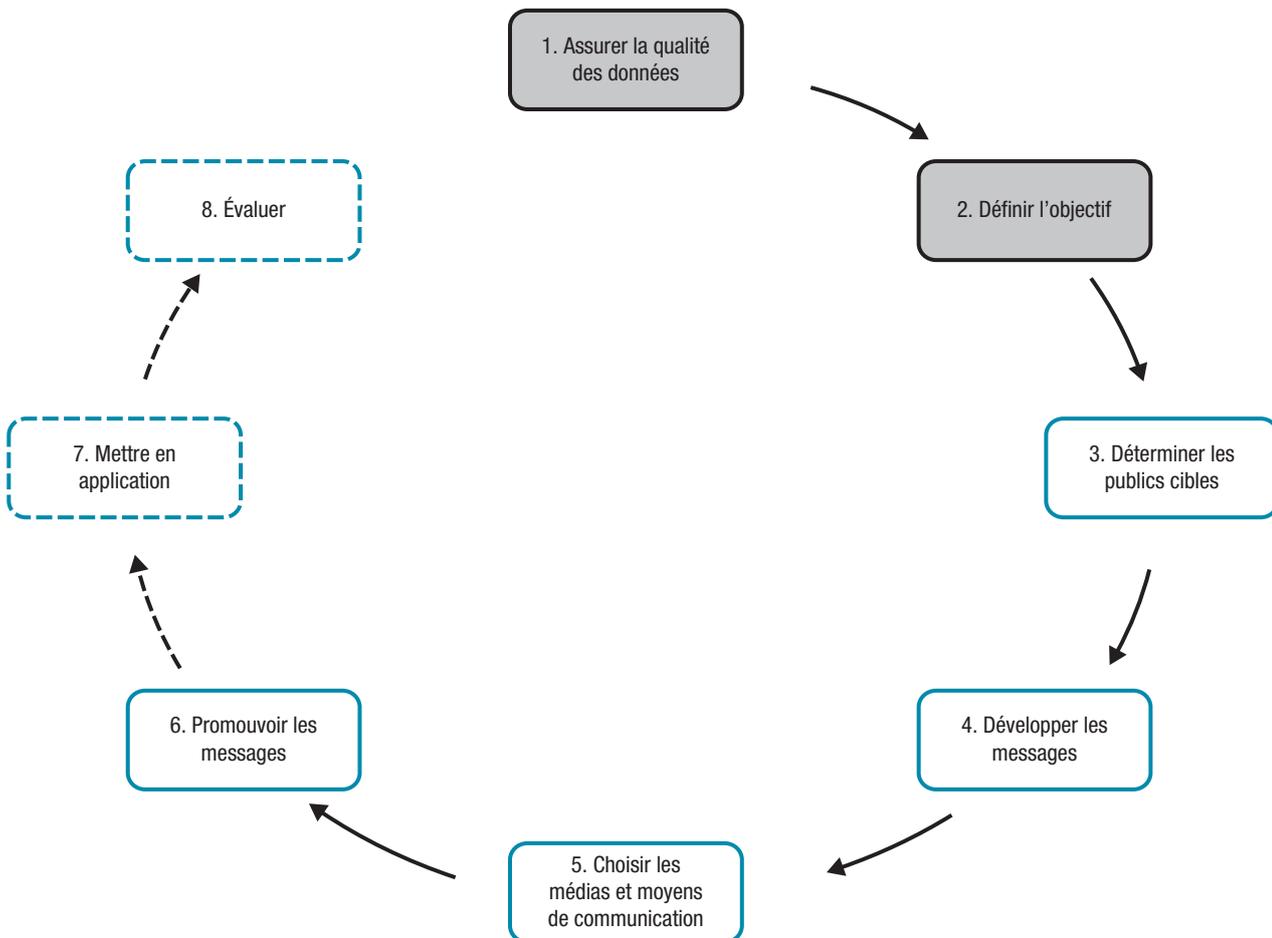
La nature administrative de ces fichiers rend nécessaire la conduite d'études de faisabilité, études qui permettent de démontrer que les DMA peuvent être utilisées aux fins de la surveillance.

### Études de faisabilité pour l'asthme et la BPCO

Pour évaluer la faisabilité d'une surveillance de l'asthme et de la BPCO au Québec à partir des DMA jumelées du Sismacq, une revue de littérature des études de validation de définitions de cas doit d'abord être faite. Cette revue permet de répertorier les définitions de cas qui ont été validées ailleurs et qui doivent maintenant être testées dans les DMA du Québec. Les données obtenues lors de ce test doivent par la suite être comparées

Figure 1

### Étapes de développement du volet « maladies respiratoires » du Système intégré de surveillance des maladies chroniques du Québec (Sismacq)



Ces étapes sont adaptées du cadre théorique *Framework for communicating surveillance information* de Remington, 2010 [2].

Étapes 1 et 2 : déjà réalisées.

Étapes 3 à 6 : en cours de réalisation.

Étapes 7 et 8 : seront réalisées dans le futur.

avec d'autres sources d'information (données d'enquête, d'études cliniques, comité d'experts), après quoi une définition de cas peut être choisie.

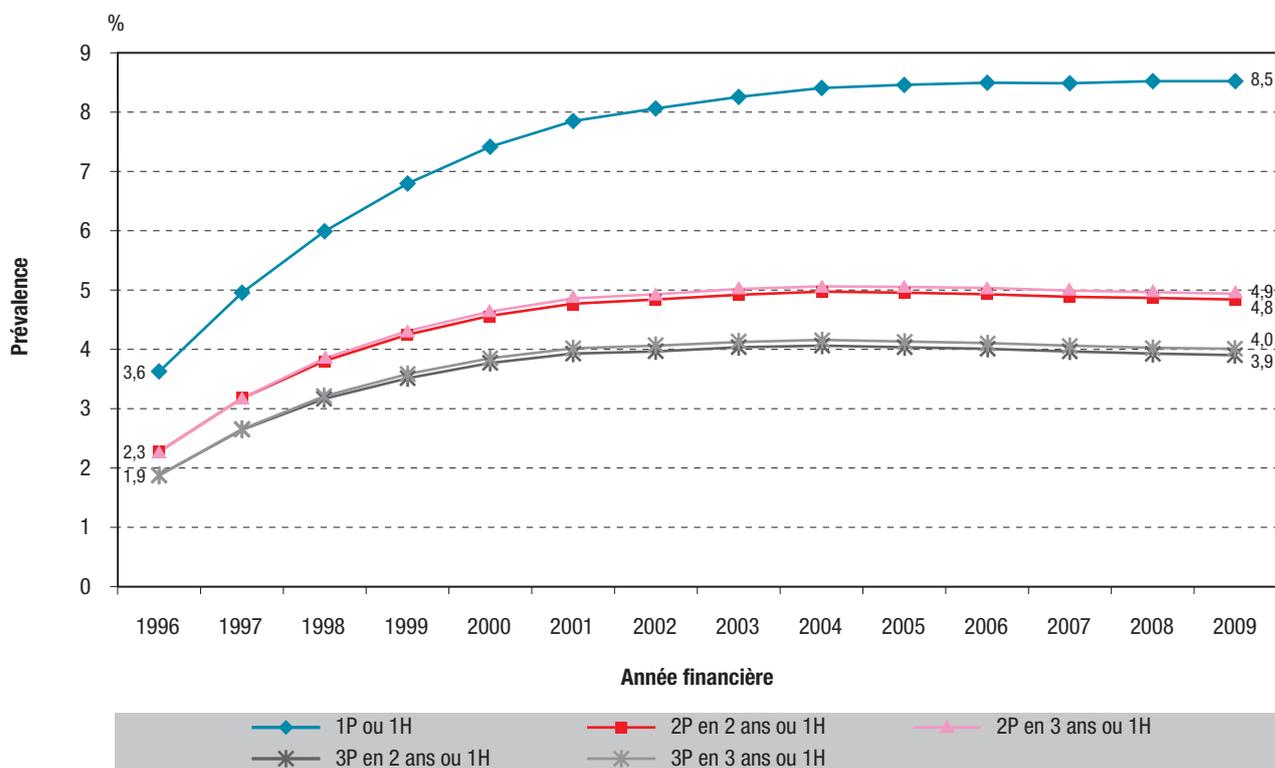
Pour la BPCO, la revue de la littérature a permis d'identifier cinq études de validation<sup>4-8</sup>, dont une a été retenue pour être testée<sup>8</sup>. L'étude de Gershon et coll. se distinguait par un échantillon représentatif de la population et un protocole mesurant la sensibilité, la spécificité et les valeurs prédictives positive et négative des définitions de cas évaluées. Dans le contexte de la validation d'une définition de cas à partir des DMA, la sensibilité correspond à la capacité de la définition d'identifier dans les DMA un individu avec une BPCO lorsque ce même individu a un diagnostic de BPCO dans son dossier médical ; la valeur prédictive positive fait référence à la probabilité que l'individu identifié comme cas de BPCO dans les DMA ait un diagnostic de BPCO dans son dossier médical. L'étude de Gershon et coll.<sup>8</sup> a évalué la validité de cinq définitions de cas, qui combinaient toutes un critère provenant du fichier des services médicaux rémunérés à l'acte et un critère provenant du fichier des

hospitalisations. Notre étude de faisabilité a analysé chacune de ces cinq définitions à partir des DMA.

Les résultats obtenus pour chacune des définitions de cas ont été comparés pour le Québec, ce qui a permis d'apprécier l'évolution de la prévalence de la BPCO chez les personnes de 35 ans et plus du 1<sup>er</sup> avril 1996 au 31 mars 2009 (figure 2). Suivant l'établissement d'un consensus entre des utilisateurs de DMA et des cliniciens québécois, nous avons retenu la définition de cas suivante parmi les cinq qui étaient évaluées : un enregistrement au fichier des services médicaux rémunérés à l'acte ou un enregistrement au fichier des hospitalisations avec un diagnostic de BPCO (CIM-9 : codes 491-492, 496 ou CIM-10 : codes J41-44, Classification internationale des maladies, 9<sup>e</sup> et 10<sup>e</sup> révisions). Dans l'étude de Gershon et coll.<sup>8</sup>, cette définition avait une sensibilité de 85,0% (intervalle de confiance (IC) à 95%:[77,0-91,0]), une spécificité de 78,4% (IC95%:[73,6-82,7]) et une valeur prédictive positive de 57,5% (IC95%:[49,6-65,1]). Qui plus est, le choix de cette définition pour la surveillance de la BPCO au Québec à partir des DMA s'est appuyé (figure 2)

Figure 2

**Prévalence relative ajustée de la BPCO chez les 35 ans et plus au Québec par définition de cas pour la période 1996-2009**



P : enregistrement au fichier des services médicaux rémunérés à l'acte (« Physician claim »), à la suite d'un diagnostic de BPCO ; H : enregistrement au fichier des hospitalisations à la suite d'un diagnostic primaire ou secondaire de BPCO.

Les cinq définitions de cas présentées (1P ou 1H ; 2P en 2 ans ou 1H ; 2P en 3 ans ou 1H ; 3P en 2 ans ou 1H ; 3P en 3 ans ou 1H) sont issues de l'étude de validation menée par Gershon et coll.[8], une étude à visée populationnelle utilisant des DMA. Chacune des définitions de cas repose sur des critères provenant du fichier des services médicaux rémunérés à l'acte (P), et du fichier des hospitalisations (H), deux bases de données administratives. Quatre de ces définitions ajoutent aussi un critère de temps, en spécifiant la période à l'intérieur de laquelle l'enregistrement au fichier des services médicaux rémunérés à l'acte doit être effectué.

Pour une année donnée, la prévalence se définit comme le nombre de cas identifiés par la définition retenue sur le nombre de personnes à risque formant la population générale.

sur l'estimation de la prévalence qui était la plus près de la réalité clinique observée au cours de la période 1996 à 2009 (figure 3). En effet, la prévalence relative brute de la BPCO au Québec en 2009 était de 9,2%, ce qui est comparable à la prévalence mesurée entre 2005 et 2009 par l'étude canadienne à visée populationnelle COLD (*Canadian Obstructive Lung Disease cohort*) chez les personnes avec BPCO de sévérité modérée à sévère (stade GOLD II+) <sup>9</sup>, qui sont de plus grands consommateurs de soins de santé et donc plus susceptibles de se retrouver dans les DMA (figure 3).

Comme pour la BPCO, une revue de la littérature des études canadiennes de validation des définitions utilisées pour identifier les cas d'asthme à partir des DMA a été effectuée. Une seule étude de validation a été retenue <sup>10</sup>. La méthode bayésienne <sup>11</sup> a ensuite été explorée afin de vérifier si elle pouvait fournir des estimations justes de la sensibilité et de la spécificité de plusieurs définitions de cas de l'asthme chez l'enfant, et ainsi contribuer à soutenir le choix d'une définition de cas en l'absence d'une étude de validation réalisée à partir de nos données. Cette approche consistait à combiner simultanément des distributions

de la prévalence, de la sensibilité et de la spécificité de l'étude retenue <sup>10,12</sup> avec les données du Sismacq, afin d'obtenir de nouvelles distributions pour les trois mêmes paramètres d'intérêt. Les analyses effectuées grâce à deux logiciels statistiques, WinBUGS® et R, ont rapidement permis de constater que ce type d'analyse ne semble pas s'appliquer à de très grandes populations, les simulations faisant émerger divers problèmes affectant la validité des résultats (exemples : auto-corrélation et non-convergence).

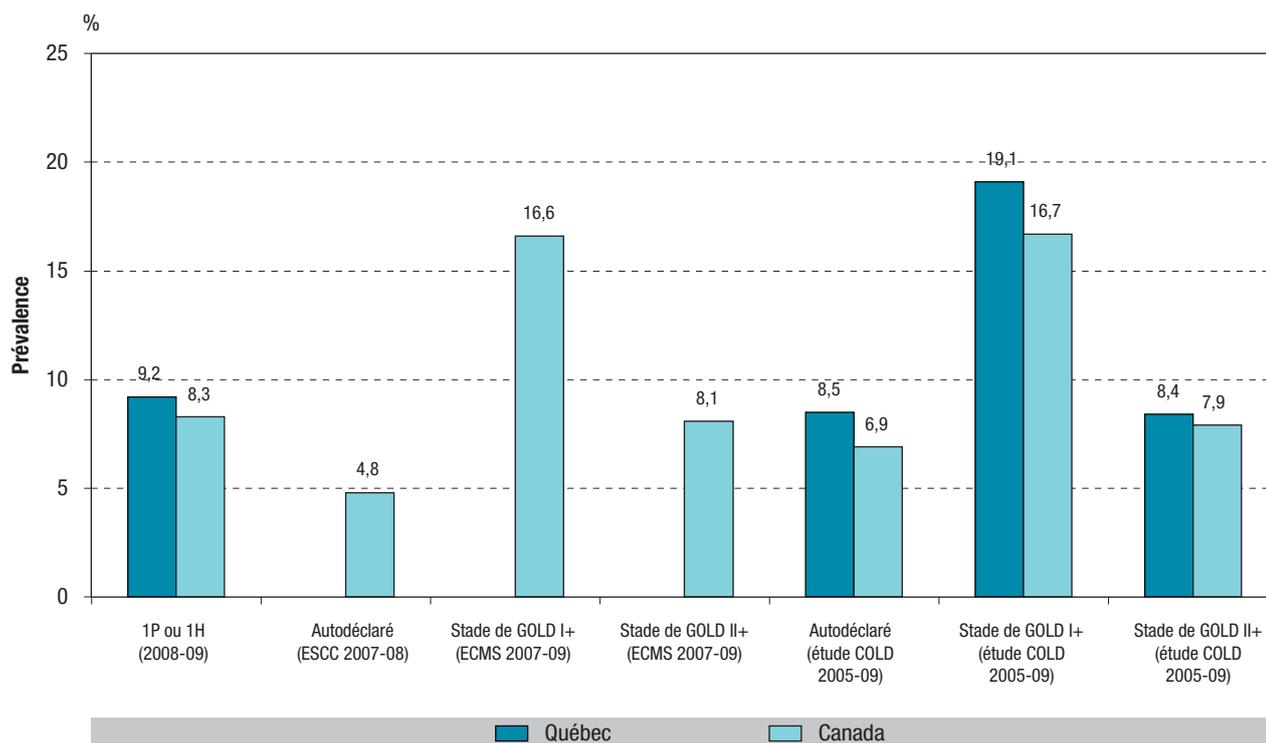
### Définir l'objectif du système de surveillance, les publics cibles et développer les messages de surveillance (étapes 2, 3 et 4)

La surveillance des maladies respiratoires doit permettre la production et la diffusion de messages de surveillance qui soient, d'une part, utiles à la prise de décision par les différents intervenants en santé publique et qui, d'autre part, facilitent les actions de prévention et de prise en charge de l'asthme et de la BPCO.

Les premiers publics cibles identifiés sont les décideurs gouvernementaux du MSSS et les acteurs de

Figure 3

### Prévalence relative de la BPCO au Québec et au Canada à partir de données médico-administratives et de données d'enquêtes, par définition, pour la période 2005-2009



ESCC : données provenant de l'Enquête sur la santé dans les collectivités canadiennes [13].

ECMS : données provenant de l'Enquête canadienne sur les mesures de la santé [13].

Étude COLD : données provenant de l'étude *Canadian Obstructive Lung Disease cohort* (cinq grandes villes de différentes provinces canadiennes participent à l'étude) [9].

GOLD : classification d'après l'initiative globale contre la BPCO (*Global Initiative for Chronic Obstructive Lung Disease*) ; stade I+ : sévérité légère à sévère ; stade II+ : sévérité modérée à sévère.

Les données comparées sont brutes. La prévalence brute de la BPCO chez les personnes de stade de GOLD II+ (sévérité modérée à sévère) qui a été estimée par l'étude COLD rejoint l'estimation de la prévalence brute estimée par la définition 1P ou 1H et mesurée à partir des DMA québécoises. C'est cette adéquation qui justifie notamment le choix de cette dernière définition de cas pour la surveillance de la BPCO au Québec à partir des DMA.

surveillance présents dans les nombreuses Agences régionales de santé du Québec. Leurs champs d'action concernent notamment l'allocation des ressources matérielles et humaines, pour les décideurs, et l'utilisation d'une méthodologie de surveillance rigoureuse par les acteurs de surveillance dans leur région respective. S'ajoutent ensuite les organismes sans but lucratif et les associations de chercheurs-cliniciens. Ces publics cibles secondaires utilisent les messages de surveillance afin de promouvoir leur organisation auprès des autorités gouvernementales pour les uns, et pour la formulation de nouvelles hypothèses de recherche, pour les autres.

Les DMA ne constituent pas en elles-mêmes des messages de surveillance. Elles sont d'abord traduites en données statistiques, ce qui en fait des produits de surveillance. Ceux-ci sont interprétés par les analystes et transformés en messages de surveillance.

Les premiers messages de surveillance porteront sur la prévalence et l'incidence de l'asthme et de la BPCO au Québec entre 1996 et 2011. Ces estimations seront présentées sous forme de courbes de tendance présentant l'évolution temporelle de chacune de ces maladies au Québec. Des estimations annuelles pour toutes les régions du Québec seront aussi produites. Ces messages ne sont pas encore publiés, mais sur le point de l'être. Au cours des prochaines années, et grâce aux données du Sismacq, d'autres messages de surveillance seront développés et reliés à toutes les étapes de l'évolution de la maladie. Plus précisément, elles permettront la mise sur pied de cohortes longitudinales qui rendront possible le développement de messages portant sur les facteurs de risque associés au développement des maladies respiratoires. De plus, le suivi temporel de ces populations facilitera l'identification des problèmes émergents et des groupes ciblés par les maladies respiratoires chroniques. Le jumelage des données permettra aussi de produire des indicateurs sur l'utilisation des médicaments prescrits pour le traitement de l'asthme et de la BPCO, le recours aux services de santé et la mortalité des personnes atteintes de ces pathologies respiratoires chroniques.

### **Choisir les voies et moyens de communication et promouvoir les messages de surveillance (étapes 5 et 6)**

La diffusion des messages de surveillance prendra plusieurs formes afin d'être parfaitement adaptée aux publics cibles. Ainsi, seront destinés aux décideurs des documents de type « faits saillants » (format une page) rapportant l'essentiel du rapport de surveillance (format huit pages). Ces messages seront transmis par courriel aux intervenants en plus d'être accessibles à tous sur le site Web de l'INSPQ. Les acteurs de surveillance seront les principaux destinataires de fiches descriptives résumant la méthodologie utilisée et des résultats agrégés à l'échelle régionale et locale. Ces informations seront accessibles via un portail d'information Web (l'Infocentre de santé publique). Pour les organismes sans but lucratif, de brèves synthèses de deux pages seront rédigées et

transmises par courriel. Aux chercheurs, le résumé (format une page) d'articles scientifiques publiés et indexés dans MEDLINE® sera transmis par courriel.

Pour faciliter une diffusion efficace de l'information, un comité d'usagers sera mis en place. Ce comité sera formé des représentants de chacun des publics cibles. Il aura pour mandat de faire une rétroaction sur les messages de surveillance et leur diffusion, et de formuler des demandes concernant de futurs messages à développer. Ce faisant, il s'assurera que les messages de surveillance répondent aux besoins des publics cibles.

### **Mise en application (étape 7)**

Concrètement, le système de surveillance sera mis en application suivant deux grands volets (figure 4). Le premier, la création des messages, passe par le recueil, l'analyse et l'interprétation des DMA ; le second constitue le plan de diffusion des messages.

### **Évaluation du système de surveillance (étape 8)**

Une évaluation du système s'impose afin de valider l'atteinte de son objectif : la diffusion des messages de surveillance pour une action en santé publique. Pour ce faire, il s'agira, d'une part, d'évaluer la portée du plan de diffusion des messages, en comptabilisant par exemple le nombre de téléchargements de chacun des messages de surveillance qui aura été diffusé. D'autre part, l'utilisation concrète des messages par les publics cibles pour la prise de décisions et les actions entreprises pourra être évaluée à partir d'un questionnaire. Les données recueillies permettront de dresser un portrait de l'utilisation des messages, ce qui permettra d'orienter la production des messages subséquents.

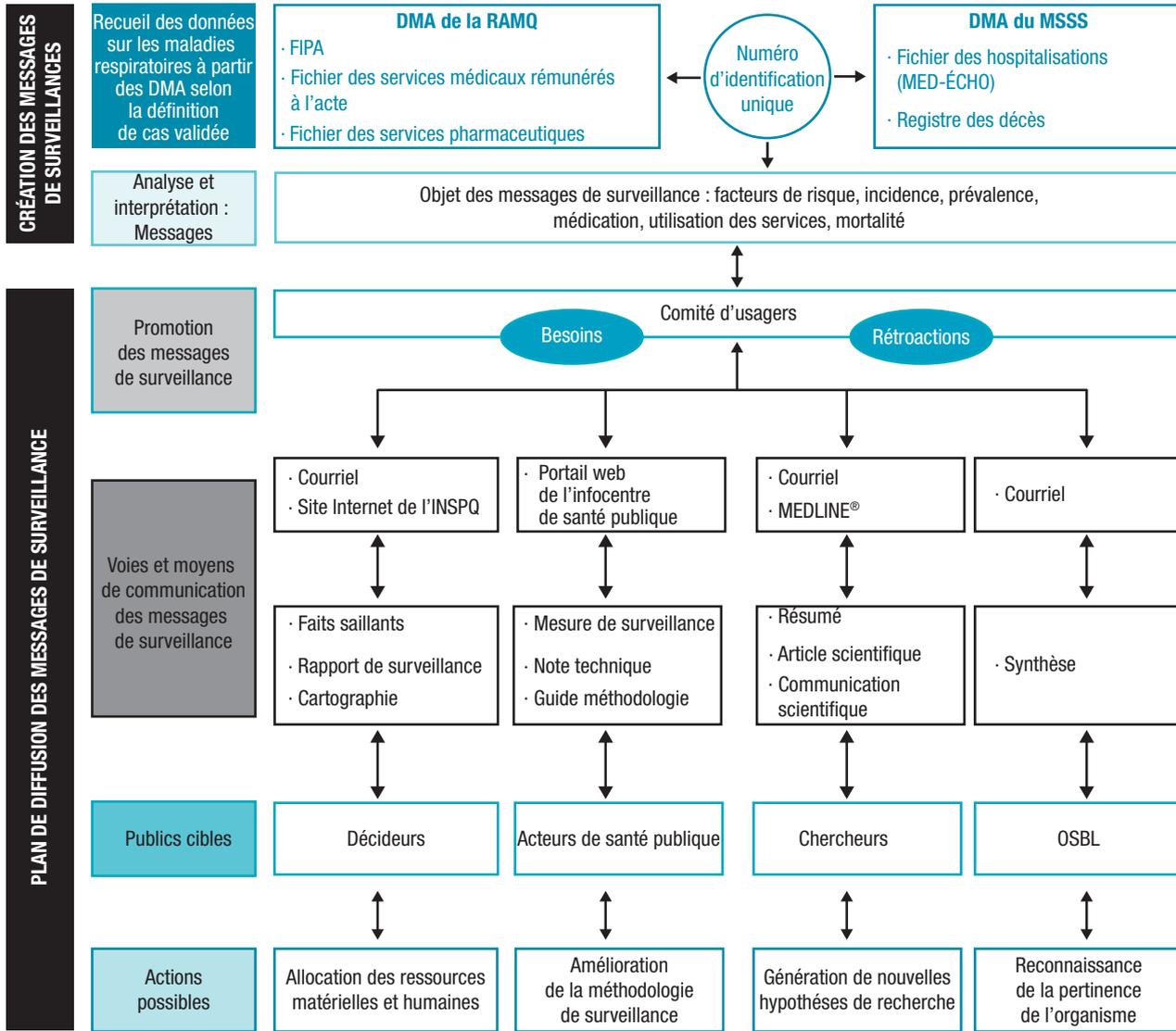
### **Défis et enjeux liés au développement**

Les deux premières étapes du développement du système de surveillance des maladies respiratoires ont été accomplies, alors que les étapes 3 à 6 sont en voie de l'être. Les deux dernières étapes (7 et 8) seront réalisées dans les prochaines années. Toutes ces étapes posent d'importants défis, qui sont surtout liés aux sources de données et à leur qualité. Par exemple, peu d'études de validation de définitions de cas sont disponibles pour l'asthme et la BPCO et les définitions de cas retenues présentent une sensibilité et une valeur prédictive positive faibles. L'étude de faisabilité est généralement le point de départ dans le choix de la meilleure définition de cas et une étude de validation à l'échelle québécoise semble, dans ce contexte, une étape importante.

Des défis s'imposent aussi en ce qui a trait à la diffusion des données de surveillance pour une action en santé publique. D'une part, il importe de trouver un équilibre entre les contraintes imposées par une méthodologie de surveillance rigoureuse, les délais inhérents à l'exploitation efficace des DMA et la nécessité de produire des données de surveillance en temps opportun pour les décideurs et intervenants. D'autre part, les produits de synthèse, tels

Figure 4

**Mise en application du volet « maladies respiratoires » du Système intégré de surveillance des maladies chroniques du Québec (Sismacq)**



DMA : données médico-administratives ; RAMQ : Régie de l'assurance-maladie du Québec ; FIPA : Fichier d'inscription des personnes assurées ; MSSS : Ministère de la Santé et des Services sociaux ; OSBL : Organismes sans but lucratif ; INSPQ : Institut national de santé publique du Québec.

La mise en application du volet « maladies respiratoires » du Système intégré de surveillance des maladies chroniques du Québec (Sismacq) se fera en deux parties principales : 1) création et 2) diffusion des messages de surveillance.

Afin de produire des messages de surveillance, les données nécessaires à la surveillance des maladies respiratoires chroniques sont extraites de cinq fichiers médico-administratifs à l'aide d'un identifiant unique. Les données brutes recueillies sont par la suite analysées et interprétées, ce qui conduit à la création de messages de surveillance. Les premiers messages développés portent sur la prévalence et l'incidence de l'asthme et de la BPCO. Les futurs messages porteront sur les facteurs de risque, l'utilisation de la médication et des services médicaux par les personnes atteintes de ces maladies, et la mortalité qui leur est attribuable.

Pour faciliter la diffusion des messages de surveillance, ceux-ci sont transmis à un comité d'usagers. Le comité se charge de faire la promotion des messages de surveillance, qui sont transmis selon différentes voies (courriel, site Internet, base de données - MEDLINE®) et différents moyens (synthèse, rapport de surveillance, cartographie, mesure de surveillance, note technique, guide méthodologique, article et communication scientifiques) aux différents publics cibles. Les publics cibles choisis sont les décideurs, les acteurs de surveillance, les chercheurs et les organisations sans but lucratif. L'objectif du volet « maladies respiratoires » du Sismacq est de générer des actions de santé publique : allocation des ressources, amélioration de la méthodologie de surveillance, génération de nouvelles hypothèses à inférer en recherche, reconnaissance de la pertinence d'organismes voués à l'amélioration de la santé respiratoire. L'action visée diffère selon le public cible. Le comité d'usagers a aussi pour mandat de recueillir les commentaires et rétroactions des publics cibles sur toutes les étapes du processus de surveillance. Les flèches bidirectionnelles symbolisent la modification possible des messages de surveillance et de leur plan de diffusion de manière à ce qu'ils s'arriment aux besoins des publics cibles.

que les faits saillants et les résumés de rapports, laissent peu de place aux éléments méthodologiques et aux informations sur les limites des données, qui sont souvent complexes. Cela pourrait entraîner une interprétation inadéquate du message. La mise en place d'un comité d'usagers posera également quelques défis, puisque ce groupe doit être composé de personnes reconnues et réputées, capables de faire la promotion du message. Or, celles-ci ont des disponibilités limitées.

La mise en application du système de surveillance sera complexe, car les données doivent être recueillies, analysées et transformées en messages adaptés aux différents publics cibles. En ce sens, harmoniser les besoins des publics et le processus d'extraction et d'analyse des données nécessitera communication et coordination de la part des analystes et des membres du comité d'usagers. Enfin, l'évaluation de la portée de la diffusion des messages et l'impact sur les actions ciblées posera d'autres défis, méthodologiques notamment.

## Conclusion

Le développement du système de surveillance des maladies respiratoires s'échelonne sur plusieurs années et assurera une meilleure connaissance de l'asthme et de la BPCO. Des messages de surveillance bien adaptés aux différents publics cibles faciliteront une meilleure prise de décisions et amélioreront la prise en charge des malades respiratoires au Québec. ■

## Remerciements

Le financement du projet est assuré par l'Agence de la santé publique du Canada, le ministère de la Santé et des Services sociaux du Québec et l'Institut national de santé publique du Québec. Les auteurs remercient sincèrement Patricia Lamontagne pour la révision de l'article.

## Références

- [1] Agence de la santé publique du Canada. La vie et le souffle : Les maladies respiratoires au Canada. Ottawa: Agence de la santé publique du Canada; 2007. 131 p.
- [2] Remington PL, Nelson DE. Communicating Public Health Surveillance Information for Action. In: Lee LM, Teutsch SM, Thacker SB, St Louis ME, editors. Principles and practice of public health surveillance. New York: Oxford University Press; 2010. pp. 146-65.

[3] Agence de la santé publique du Canada, Canadian Chronic Disease Surveillance System, Chronic Respiratory Disease Working Group. Use of Provincial / Territorial Administrative Data for the Surveillance of Chronic Respiratory Disease. Feasibility Study. Summary of Results. Ottawa: Agence de la santé publique du Canada; 2010.

[4] Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. *J Clin Epidemiol.* 2004;57(2):131-41.

[5] Lacasse Y, Montori VM, Lanthier C, Maltis F. The validity of diagnosing chronic obstructive pulmonary disease from a large administrative database. *Can Respir J.* 2005;12(5):251-6.

[6] McKnight J, Scott A, Menzies D, Bourbeau J, Blais L, Lemiere C. A cohort study showed that health insurance databases were accurate to distinguish chronic obstructive pulmonary disease from asthma and classify disease severity. *J Clin Epidemiol.* 2005;58(2):206-8.

[7] Lacasse Y, Daigle JM, Martin S, Maltais F. Validity of chronic obstructive pulmonary disease diagnoses in a large administrative database. *Can Respir J.* 2012;19(2):e5-9.

[8] Gershon AS, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, To T. Identifying individuals with physician diagnosed COPD in health administrative databases. *COPD.* 2009;6(5):388-94.

[9] Tan WC, Bourbeau J, FitzGerald JM, Cowie R, Chapman K, Hernandez P, *et al.* Can age and sex explain the variation in COPD rates across large urban cities? A population study in Canada. *Int J Tuberc Lung Dis.* 2011;15(12):1691-8.

[10] Gershon AS, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, To T. Identifying patients with physician-diagnosed asthma in health administrative databases. *Can Respir J.* 2009;16(6):183-8.

[11] Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am J Epidemiol.* 1995;141(3):263-72.

[12] Gershon AS, Guan J, Wang C, To T. Trends in asthma prevalence and incidence in Ontario, Canada, 1996-2005: a population study. *Am J Epidemiol.* 2010;172(6):728-36.

[13] Statistique Canada. Enquête sur la santé dans les collectivités canadiennes. Composante annuelle (ESCC). 2012 [Internet]. [http://www23.statcan.gc.ca/imdb/p2SV\\_f.pl?Function=getSurvey&SurvId=3226&SurvVer=1&InstalD=15282&InstaVer=9&SDDS=3226&lang=fr&db=imdb&adm=8&dis=2](http://www23.statcan.gc.ca/imdb/p2SV_f.pl?Function=getSurvey&SurvId=3226&SurvVer=1&InstalD=15282&InstaVer=9&SDDS=3226&lang=fr&db=imdb&adm=8&dis=2)

## Citer cet article

Doucet M, Rochette L, Gagné ME, Gamache P, Pouliot S, Émond V. Développement de la surveillance des maladies respiratoires au Québec à partir des données médico-administratives. *Bull Epidemiol Hebd.* 2013;(Hors-série):36-42.

## ESTIMATION DE LA FRÉQUENCE DES HÉMORRAGIES OBSTÉTRICALES EN FRANCE À PARTIR DES DONNÉES TRANSFUSIONNELLES DU PMSI

// ADVANTAGES AND LIMITATIONS OF USING NATIONAL ADMINISTRATIVE DATA ON OBSTETRIC BLOOD TRANSFUSIONS TO ESTIMATE THE FREQUENCY OF OBSTETRIC HEMORRHAGES IN FRANCE

Catherine Quantin<sup>1,2</sup> (catherine.quantin@chu-dijon.fr), Éric Benzenine<sup>1</sup>, Cyril Ferdynus<sup>3</sup>, Mourad Sediki<sup>4</sup>, Bertrand Auverlot<sup>1</sup>, Michal Abrahamowicz<sup>5</sup>, Pascal Morel<sup>6</sup>, Jean-Bernard Gouyon<sup>3,7</sup>, Paul Sagot<sup>3,8</sup>

<sup>1</sup> Service de biostatistique et d'informatique médicale, CHU de Dijon, France

<sup>2</sup> Inserm U866, Université de Bourgogne, Dijon, France

<sup>3</sup> Centre d'épidémiologie des populations, EA 4184, Université de Bourgogne, Dijon, France

<sup>4</sup> Inserm CIC-P 803, CHU de Dijon, France

<sup>5</sup> Département d'épidémiologie, de biostatistique et de santé au travail, Université McGill, Montréal, Québec, Canada

<sup>6</sup> Établissement français du sang Bourgogne Franche-Comté, Besançon, France

<sup>7</sup> Faculté de médecine, Université de Bourgogne, Dijon, France

<sup>8</sup> Service de gynécologie-obstétrique, médecine fœtale et reproduction humaine, CHU de Dijon, France

Soumis le 05/06/2013 // Date of submission: 06/05/2013

### Résumé // Abstract

**Introduction** – Bien que les hémorragies obstétricales soient l'une des causes les plus fréquentes de mort maternelle dans le monde, leur fréquence est mal connue. L'objectif de cette étude de population était d'étudier l'intérêt des données du PMSI pour estimer la fréquence des hémorragies obstétricales en France.

**Méthode** – Nous avons chaîné le PMSI avec la base de données de l'Établissement français du sang (EFS, notre référentiel) sur la région Bourgogne, puis étudié les cas discordants. À partir de cette étude, nous avons élaboré un modèle de régression multivarié permettant de corriger le nombre total de femmes enceintes identifiées dans le PMSI national comme ayant bénéficié d'une transfusion. Cette étude a été réalisée à partir des données de la région Bourgogne et a concerné toutes les parturientes pour la période 2006-2007 (35 799 grossesses). La modélisation a porté sur tous les accouchements en France en 2006 et 2007 (1 629 537 grossesses).

**Résultats** – La sensibilité du PMSI lors de la détection de femmes enceintes transfusées a été estimée à 66,3% avec une valeur prédictive positive de 91,3%. Le nombre total de femmes enceintes identifiées dans le PMSI comme ayant bénéficié d'une transfusion au niveau national a été estimé, après correction, à 10 941 pour 2006-2007, soit 6,71‰.

**Conclusion** – La méthode proposée, permet d'obtenir une estimation corrigée de la fréquence des hémorragies obstétricales au niveau national.

**Introduction** – Currently, there are no tools to monitor obstetric hemorrhages, a frequent cause of maternal death worldwide. To assess the possibility of using hospital administrative databases, and especially their data on blood transfusion, the objective of this population-based study was to estimate the frequency of obstetric hemorrhages in France.

**Method** – This assessment was achieved by probabilistic linkage of PMSI data, collected by the perinatal network for all of the 35,799 deliveries in 2006 and 2007, with data from the EFS, the French national blood agency (gold standard). At the national level, the factors associated with discordance between the PMSI and EFS, estimated by logistic regression on the regional database, were applied to national PMSI data (1,629,537 deliveries) to obtain a corrected estimation of the number of delivery-related transfusions.

**Results** – In the Burgundy region, the sensitivity of the administrative data was estimated at 66.3% and the positive predictive value at 91.3%. At the national level, we estimated that after correction by modelization, 10,941 deliveries (6.71‰) with haemorrhage required a transfusion in France.

**Conclusion** – Thanks to anonymized linkage between PMSI and EFS data, it was possible to validate and correct the number of deliveries that required a transfusion.

---

**Mots-clés** : Hémorragies du *post-partum*, Transfusion, Mortalité maternelle, Chaînage, Programme de médicalisation du système d'information, France

// **Keywords**: Postpartum haemorrhage, Transfusion, Maternal morbidity, Linkage, Health administrative data, France

---

## Introduction

L'amélioration de la santé maternelle est le cinquième des huit objectifs du millénaire définis par l'Organisation mondiale de la santé (OMS) en septembre 2000<sup>1</sup>. Le recensement de tous les cas de morbidité maternelle sévère à l'échelle nationale est donc un enjeu de santé publique important. Les hémorragies graves obstétricales sont une des plus fréquentes causes de mortalité maternelle<sup>2</sup>. Bien qu'il n'y ait pas de consensus sur une définition standard des hémorragies « graves », il existe cependant une convergence d'opinions sur les critères permettant une telle qualification. Ainsi, le *Scottish Confidential Audit of Severe Maternal Morbidity* (SCASMM) recommande de considérer une hémorragie maternelle comme grave, si la perte de sang est supérieure à 2 500 ml ou si sa prise en charge nécessite la transfusion d'au moins cinq culots globulaires<sup>3</sup>. De son côté, la Haute Autorité de santé (HAS) a statué, qu'en France, une hémorragie grave pouvait tout autant se définir comme (i) une perte de sang supérieure ou égale à 2 500 ml, (ii) une chute de l'hémoglobine supérieure ou égale à 4 g/dl, ou (iii) la nécessité de transfuser au moins quatre culots globulaires. Dans cette étude, nous avons choisi de classer une hémorragie obstétricale comme grave lorsqu'au moins quatre unités globulaires avaient été dispensées.

Nous avons souhaité déterminer si le Programme de médicalisation des systèmes d'information (PMSI) pouvait être utilisé afin d'estimer l'incidence des hémorragies obstétricales au niveau national. Afin de pouvoir étudier la qualité des informations du PMSI relative à l'identification des cas de femmes enceintes pour lesquels une transfusion était enregistrée, nous les avons confrontées à la base de données gérée par l'Établissement français du sang (EFS), considérée comme notre référentiel. Un chaînage direct, sur le plan national, des bases PMSI et EFS n'a pas été possible du fait des méthodes d'anonymisation utilisées (procédure « Foin » - Fonction d'occultation d'identifiant nominatif - pour le PMSI). Nous avons donc réalisé cette confrontation entre les données du PMSI vs. celles de l'EFS au niveau de la région Bourgogne, et étudié les possibilités de l'extrapolation au niveau national de l'estimation ainsi obtenue de la fréquence des hémorragies maternelles.

L'objectif de notre étude était donc de déterminer si les données du PMSI peuvent être utilisées afin d'estimer l'incidence des hémorragies obstétricales en France. Pour ce faire, ce travail a consisté à :

- identifier les cas de femmes enceintes pour lesquelles une transfusion a été enregistrée, à partir de la base PMSI (région Bourgogne) ;
- comparer le PMSI avec la base de données gérée par l'EFS, considérée comme notre référence au niveau de la région Bourgogne ;
- étudier les possibilités d'extrapolation au niveau national : identifier, par modélisation, les déterminants des discordances afin d'appliquer ce modèle à la base nationale PMSI et corriger les estimations.

Le PMSI ne permet pas de repérer les hémorragies graves, puisque le nombre de culots administrés n'est pas renseigné. Grâce au croisement avec les données de l'EFS, nous avons également calculé le taux de transfusion pour hémorragie grave en région Bourgogne.

## Sources de données

### La base de données de l'EFS

L'EFS est chargé de la gestion des produits sanguins. Dans cette étude, nous avons extrait, pour chaque femme de Bourgogne (enceinte ou non) ayant bénéficié d'une transfusion, son identité, le type de produit sanguin (plaquettes, concentré globulaire, plasma), la date de l'acte et le nom de l'hôpital où il avait été effectué.

### La base de données du PMSI

La base de données PMSI couvre l'ensemble de la population française ayant bénéficié d'une hospitalisation. Depuis le passage à la tarification à l'activité, l'exhaustivité du recueil est quasi-parfaite et sa qualité s'est grandement améliorée. Toutefois, la qualité des données relatives aux transfusions sanguines est peu documentée, dans la mesure où le codage des diagnostics et des actes de transfusions sanguines ont peu d'impact sur la valorisation budgétaire des séjours correspondants.

### La base de données du Réseau périnatal de Bourgogne (RPB)

Depuis 2001, tous les établissements publics et privés (niveau 1, 2 et 3) de Bourgogne adhèrent à ce réseau. Précisons que les maternités sont classées en niveau 1 à 3 selon la présence ou non d'un service de néonatalogie ou de soins intensifs néonataux ou de réanimation néonatale. Les établissements de niveau 1 accueillent les femmes dont la grossesse et l'accouchement ne présentent *a priori* aucun risque, c'est-à-dire la majorité. Ceux de niveau 2 peuvent accueillir également les enfants dont la prématurité est supérieure à 33 semaines et les établissements de niveau 3 sont spécialisés dans le suivi des grossesses pathologiques.

Le RPB collecte tous les résumés produits par les hôpitaux dans le cadre du PMSI, pour toute femme enceinte au-delà de 22 semaines d'aménorrhée accouchant en Bourgogne (18 500 naissances annuelles)<sup>4</sup>. Pour les besoins de l'étude, nous avons sélectionné toutes les naissances ayant eu lieu en Bourgogne entre le 1<sup>er</sup> janvier 2006 et le 31 décembre 2007, enregistrées dans la base de données du RPB.

## Méthode

### Étude de validation à partir des données PMSI de la région Bourgogne

À partir des données PMSI de la base RPB, nous avons identifié les séjours d'hospitalisation pour

accouchement (codes CIM-10 : Z37, O80 à O84), que nous avons chaînés avec ceux de l'*ante* et *post-partum*.

### Anonymisation des données

En respect des lois européennes et françaises, toutes les données de l'étude ont dû être préalablement rendues anonymes. Concernant les données du RPB, celles-ci l'étaient déjà, et la méthode d'anonymisation, basée sur le *Standard Hash Algorithm*, utilisée en routine par ce réseau, provient de notre équipe (logiciel Anonymat<sup>5,6</sup>). Ce logiciel utilise une méthode de hachage appliquée à l'identité, assurant une transformation irréversible de chaque champ. Nous n'avons eu, dans le cadre de cette étude, qu'à l'appliquer aux données de l'EFS afin de rendre le chaînage des deux bases possible, après accord de la Commission nationale de l'informatique et des libertés (Cnil).

### Chaînage des données

Afin de chaîner les données du RPB avec celle de l'EFS, nous avons utilisé la méthode proposée par Jaro<sup>7</sup> en prenant en compte, comme variables d'identification, le nom de jeune fille, le prénom et la date de naissance. L'application de cette méthode de chaînage n'est pas gênée par l'utilisation des données hachées, au lieu des données originelles nominatives, puisqu'elle s'appuie sur l'étude de la concordance des champs et que l'impact du hachage est négligeable, comme en témoigne le taux de collision très faible (de l'ordre de  $10^{-48}$ ).

### Étude de validation sur les données régionales

La figure représente l'organigramme général du chaînage des données PMSI du RPB avec celles de l'EFS. Les faux négatifs (FN) ont été définis comme étant les cas où le résumé PMSI du RPB ne faisait pas mention de transfusion, mais pour lesquels l'EFS attestait d'un tel acte. À l'inverse, les cas où une transfusion était mentionnée dans les résumés du PMSI, mais pour lesquels la base de l'EFS ne contenait aucune information, étaient considérés comme des faux positifs (FP). Une fois leur identification réalisée, les FN et FP ont été analysés en détail afin de déterminer les causes d'erreurs. Concernant les FP, comme aucune information les concernant ne pouvait être retrouvée dans la base de l'EFS, nous sommes retournés aux dossiers médicaux, lorsque cela était possible, afin de comprendre les causes d'une telle erreur. Concernant les FN, c'est en étudiant la base de l'EFS que nous avons pu comprendre les raisons pour lesquelles l'acte n'était pas enregistré dans les résumés PMSI du RPB. La sensibilité, la spécificité, la valeur prédictive positive (VPP) et la valeur prédictive négative (VPN) des données PMSI ont été calculées d'une part pour le séjour de l'accouchement, et d'autre part pour l'ensemble de la grossesse. Concernant le séjour pour accouchement, nous avons séparé les hémorragies pouvant être qualifiées de « précoces », (le jour ou le lendemain

de l'accouchement) de celles survenant pendant le reste du séjour (« tardives »).

### Étude de modélisation : estimation du nombre total d'hémorragies dans la base nationale du PMSI

À partir des données du croisement PMSI-EFS sur la région Bourgogne, nous avons développé un modèle multivarié de régression logistique dont la méthodologie est décrite par ailleurs<sup>8</sup>, afin de rechercher les variables pouvant être associées à une augmentation de la proportion des FN. Nous avons choisi d'inclure comme variables indépendantes : les critères de sévérité clinique sélectionnés *a priori* (âge de la mère, durée de séjour, niveau de maternité), les facteurs géographiques et socioéconomiques, également sélectionnés *a priori* (zone géographique [rural vs. urbain] et statut de l'établissement [privé vs. public]) et les variables identifiées lors de l'étude de validation (césarienne, procédures d'hémostase, diagnostics d'hémorragie et d'anémie).

Les résultats de ce modèle ont ensuite été appliqués à tous les séjours d'accouchement du PMSI national ne comportant pas la mention d'une transfusion, afin d'estimer le nombre total de cas de transfusion manqués par le PMSI (faux négatifs).

Au final, le nombre total de transfusions au niveau national a été estimé en comptabilisant le nombre de cas de transfusions repérés dans la base PMSI, en additionnant le nombre de FN et en soustrayant le nombre de FP, calculés comme décrit dans l'article cité précédemment<sup>8</sup>. L'intervalle de confiance de 95% a été obtenu en combinant les variances estimées de ces trois composantes de l'estimation.

## Résultats

### Résultats de l'étude de validation en région Bourgogne

#### Fréquence des transfusions

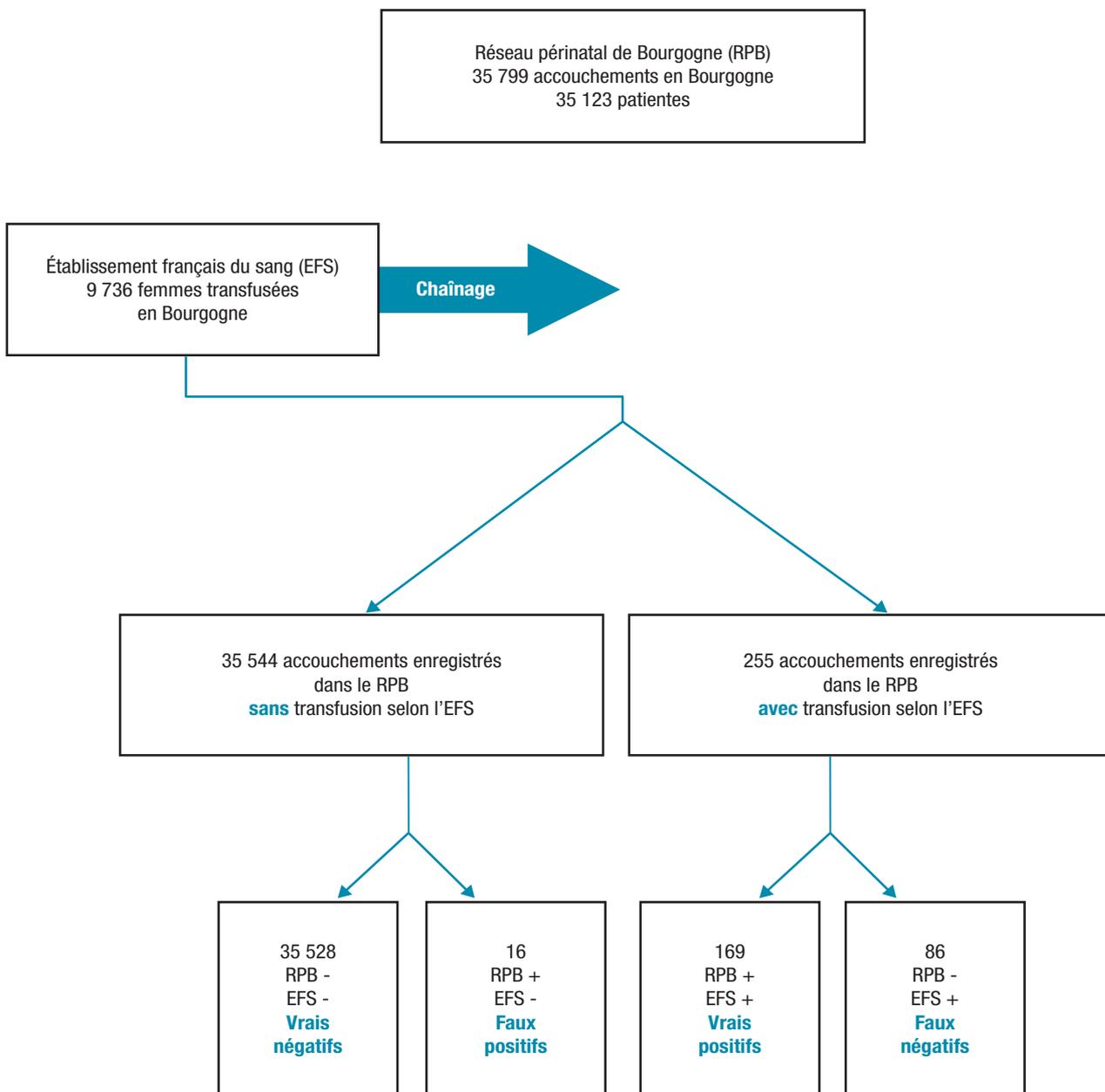
L'EFS a enregistré un total de 9 736 femmes transfusées (enceintes ou non) pour 2006 et 2007 en région Bourgogne (figure). Le RPB a identifié 35 799 accouchements sur la même période, pour 35 123 patientes. Dans les résumés de sortie du PMSI correspondants, nous avons retrouvé 195 transfusions, dont 185 réalisées durant le séjour pour accouchement.

#### Résultats du chaînage

Sur les 35 799 accouchements, le chaînage avec les données de l'EFS a permis d'en identifier 255 pour lesquels une transfusion était enregistrée dans la base de l'EFS (figure), soit un taux de transfusion de 7,12%. Aucune différence statistiquement significative n'a été retrouvée entre le nombre de transfusions de 2006 et de 2007.

La sensibilité des données PMSI a été évaluée à 66,3% et la VPP à 91,3%, pour les séjours relatifs

**Organigramme du chaînage entre les bases de l'Établissement français du sang et du Réseau périnatal de Bourgogne (séjour pour accouchement uniquement), données 2006-2007, France**



RPB + / - : transfusion enregistrée / non enregistrée dans la base RPB.

EFS + / - : transfusion enregistrée / non enregistrée dans la base EFS.

aux accouchements (tableau). Pour les hémorragies précoces, les résultats s'amélioraient : la sensibilité était à 72,8% et la VPP à 95,6%. Les résultats obtenus après chaînage des différents résumés d'une même grossesse étaient très similaires à ceux obtenus en ne prenant en compte que le séjour pour accouchement. En nous basant sur le nombre de culots dispensés (EFS), nous avons pu estimer le taux de transfusion pour hémorragies graves à 2,3% accouchements.

**Étude des cas discordants**

Un retour au dossier médical a été réalisé lorsque les données du PMSI étaient en discordance avec celles

de l'EFS : 77% des FP et 96% des FN ont pu être étudiés.

Les FP correspondaient essentiellement à des césariennes, et la majorité d'entre eux résultaient d'un abus de codage de la transfusion : soit un produit sanguin avait été codé à la place d'une autre substance (albumine, facteur de coagulation XI, inhibiteur de la C1 estérase), soit aucun produit n'avait en réalité été dispensé.

Concernant les FN, ces derniers survenaient préférentiellement dans des maternités de niveau 2 et essentiellement pour des hémorragies précoces du *post-partum*.

### Sensibilité et valeur prédictive positive des données du PMSI dans l'identification des cas de transfusion, dans le contexte obstétrical en région Bourgogne, France (2006-2007)

		Cas identifiés dans EFS	Cas identifiés dans RPB	Discordances RPB/EFS		Sensibilité [IC95%]	Spécificité	VPP	VPN	Taux de transfusion	
				Faux positifs	Faux négatifs					EFS	EFS + de 4 culots
Accouchements	Total	255	185	16	86	66,3% [60,5-72,1]	99,9%	91,3% [87,3-95,4]	99,8%	7,12‰	2,3‰
	Précoces	151	115	5	41	72,8% [65,7-79,9]	99,4%	95,6% [91,9-99,4]	95,6%	4,22‰	1,7‰
	Non précoces	104	70	11	45	56,7% [47,2-66,2]	99,9%	84,3% [75,8-92,8]	99,9%	2,90‰	0,6‰
Grossesses	Total	263	195	13	81	69,2% [63,6-74,8]	99,9%	93,3% [89,8-96,8]	99,8%	7,35‰	2,4‰
	Précoces	155	124	3	34	78,1% [71,6-84,6]	99,7%	97,6% [94,9-99,9]	96,3%	4,33‰	1,8‰
	Non précoces	108	71	10	47	56,5% [47,1-65,8]	99,9%	85,9% [77,8-94,0]	99,9%	3,02‰	0,6‰

PMSI : Programme de médicalisation des systèmes d'information ; RPB : Réseau périnatal de Bourgogne ; EFS : Établissement français du sang. VPP : valeur prédictive positive ; VPN : valeur prédictive négative.

#### Estimation de la fréquence des transfusions au niveau national

La base nationale du PMSI faisait état, entre 2006 et 2007, de 1 629 537 accouchements, avec un total de 6 932 (0,43%) transfusions.

Les variables significativement associées à une augmentation de la proportion de FN à partir du modèle de régression logistique multivarié (aire sous la courbe AUC=0,907) sont : césarienne, hémorragie, anémie, durée de séjour, procédure d'hémostase<sup>9</sup>. Le nombre total de FN a été estimé à 4 609 [3 990-5 996]. En additionnant ce chiffre aux 6 932 cas de transfusion retrouvés dans la base nationale PMSI, puis en leur soustrayant le nombre estimé de FP (600 cas), nous sommes arrivés à 10 941 [9 899-11 983] femmes enceintes transfusées au moins une fois pendant leur grossesse pour 2006-2007, soit 6,74‰ accouchements.

#### Discussion

Pour apprécier la pertinence des données du PMSI, dans la surveillance des hémorragies obstétricales, nous les avons confrontées aux données de l'EFS. Cette étude est originale car, du fait de l'anonymisation irréversible des informations de la base nationale du PMSI, il n'est pas possible de chaîner directement ces deux bases au niveau national. Néanmoins, ce chaînage a été possible au niveau de la région Bourgogne grâce aux données du RPB, qui incluent les résumés de sortie du PMSI, en appliquant la méthode d'anonymisation du RPB aux données de l'EFS.

Nos résultats ne sont pas très différents selon que l'on étudie le séjour pour accouchement (sensibilité

de 66,3%) ou l'ensemble des séjours au cours de la grossesse (sensibilité de 69,2%). La sensibilité globalement basse des données du PMSI confirme les résultats d'une autre étude de validation<sup>9</sup>, menée sur quatre Centres hospitaliers universitaires (CHU) dans le cadre de la morbidité maternelle sévère.

L'autre originalité de ce travail est d'avoir mené l'étude à l'échelle de la population d'une région, incluant tous les types d'établissements. De plus, nous avons intégré les résultats obtenus à l'échelle de la région Bourgogne, suite à la mise en œuvre d'une modélisation (recherche des facteurs associés aux FP et aux FN) aux données PMSI nationales, afin d'estimer le nombre total de femmes enceintes ayant bénéficié d'une transfusion. Les facteurs ayant un impact significatif sur la qualité des données se sont avérés être des caractéristiques cliniques ou de prise en charge (césarienne, hémorragie, anémie, procédure d'hémostase, durée du séjour). À l'opposé, les critères liés à l'établissement (zone géographique, statut public ou privé, niveau de la maternité) ne se sont pas révélés, ainsi que l'âge de la patiente, avoir un impact significatif sur la qualité des données. En appliquant notre méthode, nous avons estimé un taux de transfusion corrigé de 6,71‰ accouchements. Le résultat de notre étude est très proche de celui obtenu lors d'une étude randomisée sur 106 maternités françaises<sup>10</sup>. Notre estimation du taux de transfusion se situe également dans l'intervalle de 0,34-0,90% avancé par d'autres auteurs français<sup>11</sup> et se rapproche de celui observé au Canada (0,647% des naissances)<sup>12</sup>.

Il est donc possible de corriger l'estimation fournie par la base nationale du PMSI afin d'éviter une sous-évaluation du nombre de transfusions. Les résultats présentés dans cet article montrent que notre

méthode fait passer le nombre de transfusions de 6 932 à 10 941, soit une augmentation de près de 42%.

## Conclusion

Les informations du PMSI semblent donc permettre d'estimer la fréquence des hémorragies obstétricales, à condition de disposer d'une méthode de correction adaptée, fondée sur la comparaison avec une base de référence. Toutefois, le nombre de culots administrés n'étant pas renseigné au sein du PMSI, l'utilisation de ce dernier pour le repérage des hémorragies graves n'est pas réalisable.

Une alternative serait de mettre en place un chaînage national entre les bases PMSI et EFS. Un tel dispositif permettrait d'utiliser les données du PMSI pour le suivi de l'ensemble des pathologies nécessitant une transfusion et représenterait une grande avancée sur le plan de la santé publique. ■

## Références

- [1] Ronsmans C, Graham WJ, Lancet Maternal Survival Series steering group. Maternal mortality: who, when, where, and why. *Lancet*. 2006;368(9542):1189-200.
- [2] Khan KS, Wojdyla D, Say L, Gülmezoglu AM, Van Look PF. WHO analysis of causes of maternal death: a systematic review. *Lancet*. 2006;367(9516):1066-74.
- [3] Kilpatrick SJ, Crabtree KE, Kemp A, Geller S. Preventability of maternal deaths: comparison between Zambian and American referral hospitals. *Obstet Gynecol*. 2002;100(2):321-6.
- [4] Quantin C, Gouyon B, Avillach P, Ferdynus C, Sagot P, Gouyon JB. Using discharge abstracts to evaluate a regional perinatal network: assessment of the linkage procedure of anonymous data. *Int J Telemed Appl*. 2009;(2009):181842. doi:10.1155/2009/181842.
- [5] Quantin C, Bouzelat H, Allaert FA, Benhamiche AM, Faivre J, Dusserre L. Automatic record hash coding and linkage for epidemiological follow-up data confidentiality. *Methods Inf Med*. 1998;37(3):271-7.
- [6] Quantin C, Bouzelat H, Allaert FA, Benhamiche AM, Faivre J, Dusserre L. How to ensure data security of an epidemiological follow-up: quality assessment of an anonymous record linkage procedure. *Int J Med Inform*. 1998;49(1):117-22.
- [7] Jaro MA. Probabilistic linkage of large public health data files. *Stat Med*. 1995;14(5-7):491-8.
- [8] Quantin C, Benzenine E, Ferdynus C, Sediki M, Auverlot B, Abrahamowicz M, *et al*. Advantages and limitations of using national administrative data on obstetric blood transfusions to estimate the frequency of obstetric hemorrhages. *J Public Health*. 2013;35(1):147-56.
- [9] Chantry AA, Deneux-Tharoux C, Cans C, Ego A, Quantin C, Bouvier-Colle MH; GRACE study group. Hospital discharge data can be used for monitoring procedures and intensive care related to severe maternal morbidity. *J Clin Epidemiol*. 2011;64(9):1014-22.
- [10] Deneux-Tharoux C, Dupont C, Colin C, Rabilloud M, Touzet S, Lansac J, *et al*. Multifaceted intervention to decrease the rate of severe postpartum haemorrhage: the PITHAGORE6 cluster-randomised controlled trial. *BJOG*. 2010;117(10):1278-87.
- [11] François A, Courtois F, Collège national des gynécologues et obstétriciens français, Agence nationale d'accréditation et d'évaluation en santé. Management of blood products in the event of postpartum hemorrhage. *J Gynecol Obstet Biol Reprod (Paris)*. 2004;33(8 Suppl):4S120-4S9.
- [12] Liu S, Joseph KS, Bartholomew S, Fahey J, Lee L, Allen AC, *et al*. Maternal Health Study Group of the Canadian Perinatal Surveillance System. Temporal trends and regional variations in severe maternal morbidity in Canada, 2003 to 2007. *J Obstet Gynaecol Can*. 2010;32(9):847-55.

## Citer cet article

Quantin C, Benzenine E, Ferdynus C, Sediki M, Auverlot B, Abrahamowicz M, *et al*. Estimation de la fréquence des hémorragies obstétricales en France à partir des données transfusionnelles du PMSI. *Bull Epidémiol Hebd*. 2013;(Hors-série):43-8.

## CROISEMENT DE DEUX BASES DE DONNÉES MÉDICO-ADMINISTRATIVES : MÉTHODOLOGIE ET ÉTUDE DESCRIPTIVE POUR UNE APPLICATION À LA SURVEILLANCE ÉPIDÉMIOLOGIQUE DES CANCERS EN FRANCE

// CROSSING TWO HEALTH ADMINISTRATIVE DATABASES: METHODOLOGY AND DESCRIPTIVE STUDY FOR APPLICATION TO EPIDEMIOLOGICAL SURVEILLANCE OF CANCER IN FRANCE

Yao Kudjawu<sup>1</sup> (y.kudjawu@invs.sante.fr), Florence de Maria<sup>1</sup>, Elsa Decool<sup>1</sup>, Michelle Altana<sup>2</sup>, Jean-Marc Harlin<sup>3</sup>, Alain Weill<sup>4</sup>, Javier Nicolau<sup>1</sup>, Christian Leroy<sup>1</sup>, Isabelle Grémy<sup>1</sup>

<sup>1</sup> Institut de veille sanitaire, Saint-Maurice, France

<sup>2</sup> Régime social des indépendants, Saint-Denis, France

<sup>3</sup> Mutualité sociale agricole, Paris, France

<sup>4</sup> Caisse nationale d'assurance maladie des travailleurs salariés, Paris, France

Soumis le 26.07.2013 // Date of submission: 07.26.2013

### Résumé // Abstract

**Introduction** – L'utilisation des données croisées du programme de médicalisation des systèmes d'information (PMSI) et d'affections longue durée (ALD) constitue, en France, l'une des alternatives pour estimer localement l'incidence des cancers dans les zones dépourvues de registre. L'objectif de cette étude est de décrire la méthodologie de sélection des séjours PMSI et des bénéficiaires d'ALD et leur croisement, puis de comparer les effectifs obtenus aux données d'incidence de la zone registre de cancer.

**Matériel et méthode** – La période d'étude couvrait 2006-2008. Les données de 21 localisations cancéreuses étaient extraites du PMSI national, des bases nationales de l'assurance maladie et des registres de cancers. La sélection des informations dans le PMSI et les ALD, croisées entre elles grâce à un identifiant anonyme individuel et commun, reposait sur des algorithmes basés sur les codes CIM-10 de diagnostic de cancer.

**Résultats** – Le pourcentage de patients PMSI appariés aux bénéficiaires d'ALD variait de 19% à 60% selon la localisation cancéreuse. L'effectif de patients hospitalisés ou admis en ALD, considérés comme atteints pour la première fois d'un cancer invasif et dénombrés dans la base croisée PMSI-ALD, était plus élevé que le nombre de cas incidents dans la zone registre.

**Conclusion** – Cette description est une étape nécessaire à l'utilisation de la base croisée PMSI-ALD pour l'estimation infranationale de l'incidence des cancers. L'estimation reposera sur l'application, aux zones géographiques considérées, du rapport « incidence cancer/indicateur issu du PMSI-ALD croisé » calculé sur la zone registre.

**Introduction** – The use of crossed data from the French Hospital Discharge Data system (PMSI) and long-term illness (LTI) databases is one of alternatives to estimate subnational incidence of cancer in areas without registries in France. The objective of this study is to describe the methodology of hospital stays and LTI selection and their crossing, and to compare figures obtained from the crossed data to cancer incidence in registry area.

**Material and method** – The study period covered 2006-2008. Data for 21 cancers sites were extracted from national PMSI, health insurance databases and common database of cancer registries. Hospital stays and LTI beneficiaries' selection was based on algorithms related to ICD10 diagnosis codes and were crossed through an anonymous individual and common identifier.

**Results** – The percentage of PMSI patients matched with LTI varied from 19% to 60%, according to cancer site. The number of hospitalized or admitted to LTI patients, considered as suffering from invasive cancer for the first time in the crossed database, was higher than the number of incident cancers in registry area.

**Conclusions** – This description is a necessary step to use PMSI-LTI crossed data for subnational incidence estimation of cancer. The estimation will be based on "cancer incidence/number of patients from PMSI-LTI crossed database" ratio in registry area applied to considered geographical areas.

---

**Mots-clés** : Cancer, PMSI, ALD, Épidémiologie, France

// **Keywords** : Cancer, Hospital discharge data, Long-term illness data, Epidemiology, France

---

## Introduction

En France, la surveillance épidémiologique des cancers repose principalement sur les données des registres des cancers. Ces structures, créées le plus souvent sur une base départementale, constituent les seules sources de données exhaustives et continues d'enregistrement des cas de cancer (selon la localisation, le type histologique, le stade au diagnostic...) sur leur zone géographique. La recherche active des cas et leur validation par les registres nécessitent des moyens humains et financiers importants représentant un obstacle majeur à la mise en place de registres dans la totalité des 101 départements français. Les registres français couvrent environ 20% de la population générale<sup>1,2</sup>. À partir de leurs données, des estimations nationales de l'incidence des cancers sont produites depuis plusieurs années<sup>3</sup>. Elles reposent sur l'extrapolation à la France du rapport « Incidence/Mortalité » observé dans la zone registre (*i.e.* ensemble des départements couverts par un registre). Si ce rapport peut être raisonnablement utilisé pour les estimations nationales, il ne peut l'être au niveau infranational, en raison de sa variabilité interdépartementale<sup>4</sup> empêchant toute estimation d'incidence dans les zones dépourvues de registres.

Les bases nationales de données médico-administratives que sont le Programme de médicalisation des systèmes d'information (PMSI) et les bases contenant les Affections longue durée (ALD) n'ont pas été créées pour mesurer l'incidence des cancers<sup>5,6</sup>. Cependant, depuis quelques années, les ALD sont utilisées pour suivre les tendances récentes de l'incidence nationale des cancers<sup>7</sup>. Des développements méthodologiques ont également été réalisés avec ces bases pour estimer l'incidence des cancers au niveau infranational. Ils reposent sur l'utilisation du rapport « Incidence/PMSI (ou ALD) de la zone registre » appliqué au PMSI (ou ALD) des départements (ou régions) sans registre<sup>1,8</sup>. Des estimations départementales (ou régionales) ont ainsi pu être fournies pour certains cancers mais pas pour tous, car l'hypothèse d'égalité du rapport « Incidence/PMSI (ou ALD) » entre les départements (ou régions) n'a pu être validée pour plus de la moitié des localisations cancéreuses<sup>8,9</sup>. Pour ces localisations, l'une des orientations repose sur la construction d'un nouvel indicateur, proxy de l'incidence, obtenu en croisant les bases PMSI et ALD, qui serait soumis à la même méthodologie pour produire des estimations infranationales. Cette démarche doit être précédée d'une étude exploratoire de l'indicateur construit à partir du croisement PMSI-ALD, afin de comprendre les phénomènes induits par le croisement d'informations médico-administratives de différente nature.

L'objectif de l'étude exploratoire présentée dans cet article est de :

1) décrire la méthodologie de :

- sélection des patients atteints pour la première fois d'un cancer invasif dans chacune des bases ;
- croisement des bases PMSI et ALD ;
- construction de l'indicateur proxy de l'incidence issu du croisement PMSI-ALD ;

2) comparer cet indicateur avec les cas incidents de cancers invasifs de la zone registre qui est la référence.

## Matériel et méthode

La période d'étude couvrait 2006-2008 et concernait 21 localisations cancéreuses (tableau) identifiées en utilisant la Classification internationale des maladies, 10<sup>e</sup> révision (CIM-10).

### Trois informations différentes et une information de référence

#### Les séjours hospitaliers issus du PMSI (période 2004-2008)

Le PMSI enregistre exhaustivement les hospitalisations des établissements (privés et publics) de court séjour. Chacune produit un résumé de sortie anonyme comportant des informations administratives (âge, sexe, code postal de résidence) et médicales : code CIM-10 de diagnostics principal (DP), relié (DR) et associé significatif (DAS) enregistré sur quatre caractères, actes réalisés. Pour mémoire, le DP représente (jusqu'en 2009) le « motif de prise en charge qui a mobilisé l'essentiel de l'effort médical et soignant au cours de l'hospitalisation ». Le DR « a pour rôle, en association avec le DP et lorsque celui-ci n'y suffit pas, de rendre compte de la prise en charge du malade en termes médico-économiques » : il s'agit de tout diagnostic permettant d'éclairer le contexte pathologique, essentiellement lorsque le DP n'est pas lui-même une affection. Pour le cancer, il est fréquemment complété avec un code de cancer invasif lors de la réalisation de séances de chimiothérapie ou de radiothérapie, elles-mêmes étant codées en DP. Les DAS sont notamment « les diagnostics, symptômes et autres motifs de recours significatifs d'une majoration de l'effort de soins et de l'utilisation des moyens, par rapport aux DP et DR » : il peut s'agir d'une complication du DP ou du couple DP+DR, d'une complication du traitement ou d'une affection distincte supplémentaire (source : <http://www.atih.sante.fr>).

Chaque année l'Institut de veille sanitaire (InVS), dans le cadre de ses missions de surveillance, reçoit de l'Agence technique de l'information sur l'hospitalisation (Atih) une extraction du PMSI national. Pour cette étude, cinq bases annuelles de séjours PMSI pour cancers de 2004 à 2008 ont été utilisées. La base PMSI de chaque année N regroupe tous les séjours hospitaliers de l'année N, en lien avec un cancer.

#### Les consommations de soins pour cancer reconnu comme ALD, issues du Sniiram (période 2006-2008)

Le Sniiram (Système national d'information inter-régimes de l'assurance maladie) est un ensemble de bases de données informationnelles des régimes d'assurance maladie. Il est constitué des données d'activité et de dépenses détaillées, individualisées et chaînées, couvrant les soins de ville, les soins hospitaliers sanitaires et sociaux, publics et privés.

Dans ces bases figure l'information d'ALD, un dispositif financier de l'assurance maladie de prise en charge totale des soins et traitements en lien avec une maladie « dont la gravité et/ou le caractère chronique nécessitent un traitement prolongé et une thérapeutique particulièrement coûteuse »<sup>10</sup> comme les cancers. Ces données d'ALD contiennent des informations médicales (code CIM-10 de la pathologie ayant entraîné la mise en ALD, enregistré sur trois caractères, dates de début et de fin d'ALD). Ces informations peuvent être chaînées avec des informations administratives (année et mois de naissance, sexe, code postal de résidence, régime d'affiliation) présentes dans le Sniiram. Trois extractions du Sniiram 2006-2008 relatives aux cancers, transmises par l'assurance maladie, ont été utilisées. La base Sniiram de l'année N regroupe toutes les consommations de soins de l'année N associées à une ALD cancer, que celle-ci ait été accordée l'année N ou les années précédentes (exemple : la base Sniiram 2006 contient des informations sur les patients mis en ALD en 2006 ou avant 2006 qui ont consommé des soins en lien avec le cancer en 2006).

#### Les assurés sociaux bénéficiant pour la première fois du dispositif ALD de la Mutualité sociale agricole (MSA) en raison d'un cancer (période 2006-2008)

En raison de leur absence du Sniiram, les données des assurés sociaux de la MSA ont spécifiquement été extraites des bases des bénéficiaires d'ALD pour cancer de la MSA (informations administratives et médicales). Tout comme le Sniiram, trois extractions de bases MSA, correspondant chacune à une année, de 2006 à 2008, ont été utilisées. La base MSA de l'année N regroupe tous les patients atteints d'un cancer reconnu comme ALD par la MSA pour la première fois l'année N.

#### La référence : les cas de cancer invasif de la zone registre (période 2006-2008)

L'indicateur proxy de l'incidence issu du croisement PMSI-ALD a été comparé à l'incidence de la zone registre. Les données des registres ont été extraites de la base commune des registres des cancers.

#### Étape préliminaire : un identifiant anonyme commun

Les données des bases utilisées ont bénéficié d'une anonymisation qui reposait sur le numéro d'inscription au répertoire, le sexe et la date de naissance complète, grâce à la fonction d'occultation d'informations nominatives de l'assurance maladie. Il en est résulté un identifiant anonyme unique par personne, commun aux bases médico-administratives.

#### Sélection des patients hospitalisés ou admis en ALD atteints pour la première fois d'un cancer invasif

La méthode de sélection des patients diffère entre les bases PMSI, Sniiram et MSA en raison de leur mode de constitution non identique (figure 1).

#### Sélection dans les bases PMSI

Dans la base PMSI de l'année N, les séjours avec un code CIM-10 de cancer invasif en DP ont été sélectionnés et chaînés entre eux (même code CIM-10 en DP) puis avec ceux des deux années précédentes (mêmes codes CIM-10 en DP ou DR ou DAS) afin d'exclure les cancers invasifs prévalents. Les patients avec un code CIM-10 de cancer invasif l'année N sans séjour hospitalier pour ce même cancer en N-1 ou N-2 ont été retenus, permettant de constituer trois bases d'étude : PMSI 2006, PMSI 2007 et PMSI 2008.

#### Sélection dans les bases Sniiram et MSA

Dans la base Sniiram de l'année N, les consommations de soins avec un code CIM-10 de cancer invasif au niveau de la pathologie associée à l'ALD, et avec une date de mise en ALD l'année N, ont été chaînées entre elles pour obtenir les consommateurs de soins avec un code CIM-10 de cancer invasif dont la date de mise en ALD correspondait à l'année du fichier Sniiram. Au final, trois bases d'étude ALD-Sniiram ont été constituées, de 2006 à 2008.

Dans la base MSA de l'année N, la sélection reposait sur les assurés avec un code CIM-10 de cancer invasif au niveau de la pathologie associée à l'ALD. Au final, trois bases d'étude ALD-MSA ont été constituées, de 2006 à 2008.

Pour chacune des trois années, les consommateurs ALD-Sniiram ont été fusionnés avec les assurés ALD-MSA. Les doublons ont été éliminés afin de tenir compte des éventuels transferts d'assurés entre régimes.

#### Croisement des bases PMSI et ALD-Sniiram/MSA et traitement des non appariés

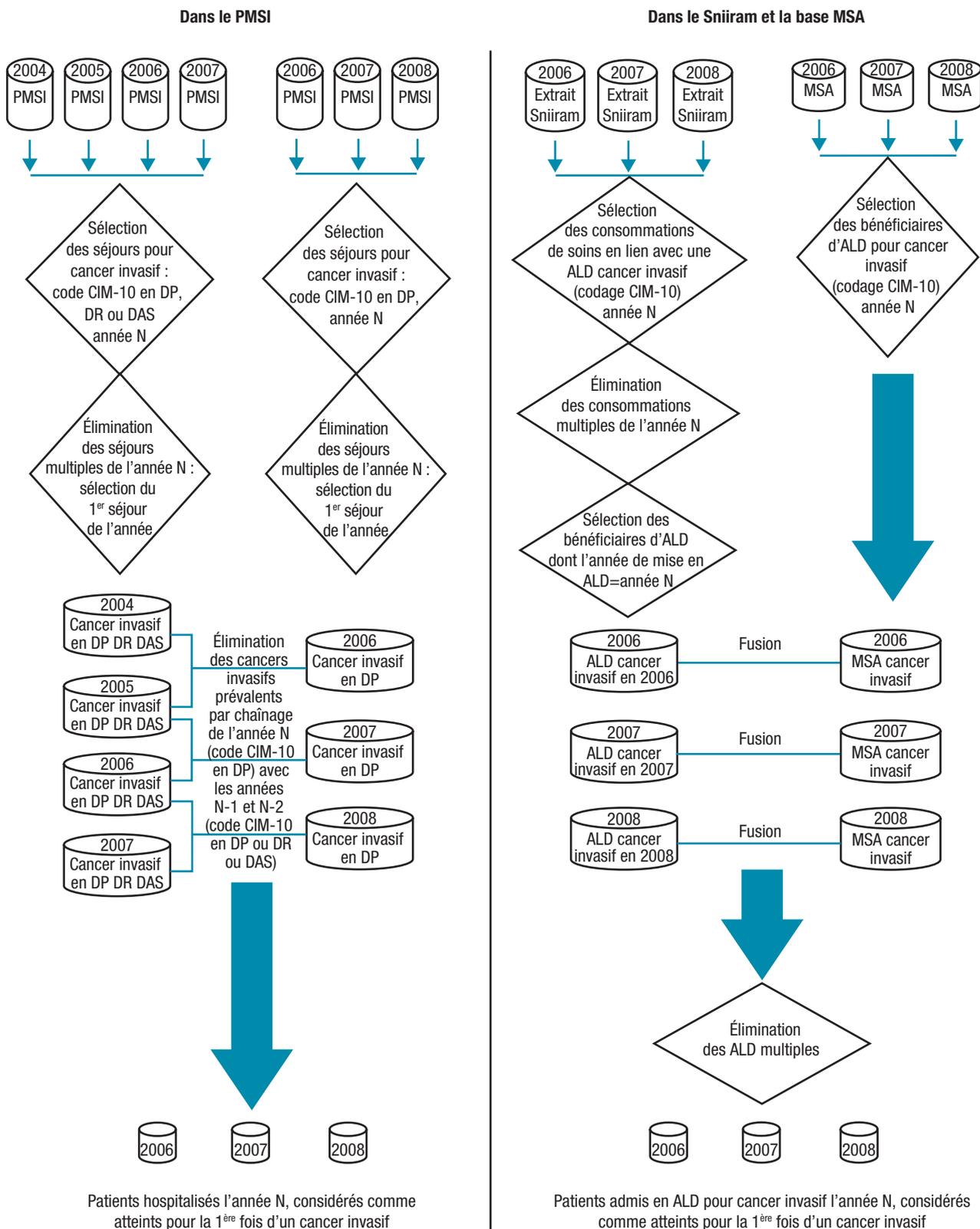
Les patients issus des bases d'étude PMSI et ALD-Sniiram/MSA (appelées dans la suite de l'article « bases ALD ») ont été croisés grâce à l'identifiant individuel anonyme commun et aux codes CIM-10 de cancer invasif (figure 2). À l'issue du croisement, trois groupes de patients ont été identifiés : patients PMSI-ALD appariés, patients PMSI non appariés aux bénéficiaires ALD et bénéficiaires ALD non appariés aux patients PMSI.

Pour chaque patient non apparié au sein d'une source, une vérification du caractère initial ou invasif du cancer a été réalisée à partir des informations contenues dans l'autre source. Elle consistait :

- pour le caractère initial du cancer d'un patient non apparié mis en ALD pour cancer, à rechercher parmi les bénéficiaires d'ALD non appariés de l'année N ceux qui avaient effectué un séjour hospitalier dans les deux années précédentes pour le même cancer, à l'aide des bases PMSI N-1 et PMSI N-2 ;
- pour le caractère initial du cancer d'un patient non apparié enregistré dans le PMSI, à rechercher parmi les patients PMSI non appariés de l'année N, ceux dont le cancer avait été reconnu comme ALD antérieurement à l'année N,

Figure 1

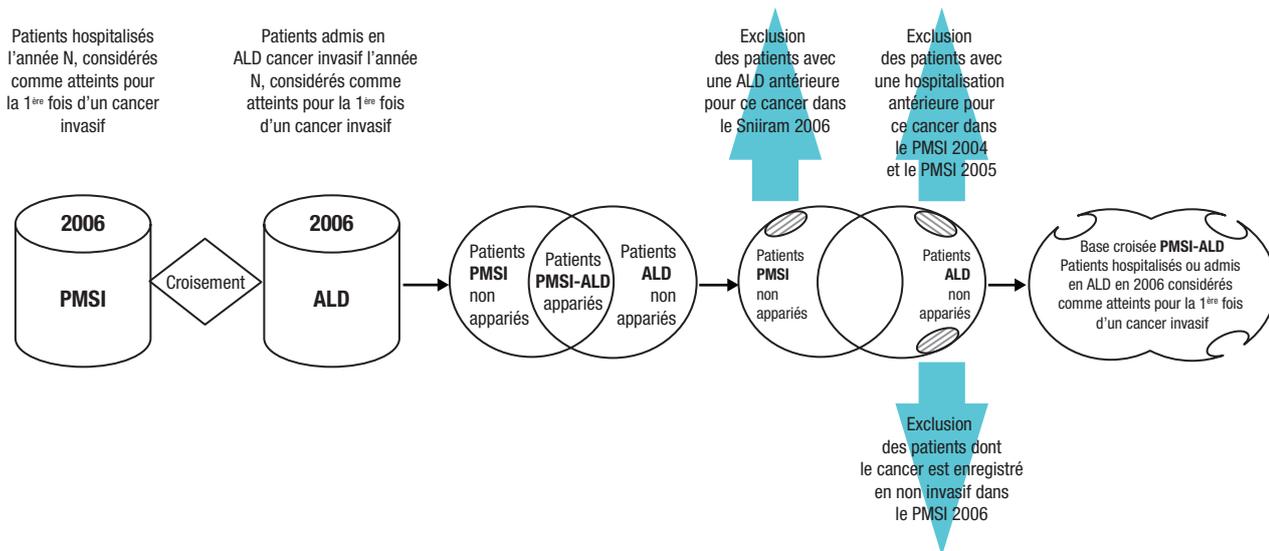
**Sélection des patients hospitalisés ou admis en ALD, atteints pour la première fois d'un type de cancer invasif**



ALD : Affection longue durée ; PMSI : Programme de médicalisation des systèmes d'information ; Sniiram : Système national d'information inter-régimes de l'assurance maladie ; MSA : Mutualité sociale agricole ; DP : diagnostic principal ; DR : diagnostic relié ; DAS : diagnostic associé, significatif.

Figure 2

**Croisement des bases PMSI et ALD (Sniiram/MSA) puis traitement des non appariés : exemple pour l'année 2006**



ALD : Affection longue durée ; PMSI : Programme de médicalisation des systèmes d'information ; Sniiram : Système national d'information inter-régimes de l'assurance maladie ; MSA : Mutualité sociale agricole.

enregistrés dans la base Sniiram de l'année N. Cette recherche ne pouvait pas être réalisée au sein des bases ALD-MSA en raison de leur indisponibilité pour les années antérieures à 2006 ;

- pour le caractère invasif du cancer d'un patient non apparié mis en ALD pour cancer, à rechercher parmi les bénéficiaires d'ALD non appariés de l'année N, ceux dont le diagnostic de cancer était identifié comme *in situ* (ou tumeur bénigne pour le système nerveux central) dans la base PMSI de l'année N (code CIM-10 en DP) ;
- pour le caractère invasif du cancer d'un patient non apparié enregistré dans le PMSI, la recherche n'a pas été réalisée en raison de la qualité du diagnostic final enregistré dans le PMSI (basé sur des informations histologiques ou hématologiques issues d'examen complémentaires), considérée *a priori* comme supérieure à celle accompagnant les mises en ALD.

Les patients PMSI ou ALD non appariés pour lesquels ces recherches étaient positives ont été exclus.

**Obtention de la population d'étude : les patients hospitalisés ou admis en ALD considérés comme atteints pour la première fois d'un cancer invasif**

Pour chaque cancer, le nouvel indicateur de surveillance était constitué de l'union :

- des patients PMSI appariés aux bénéficiaires ALD ;
- des patients PMSI non appariés aux bénéficiaires ALD, desquels ont été exclus ceux dont le cancer a été identifié comme prévalent grâce aux informations ALD ;
- des bénéficiaires ALD non appariés aux patients PMSI, desquels ont été exclus ceux dont le

cancer a été identifié comme prévalent ou non invasif grâce aux informations PMSI.

**Étude du nouvel indicateur : description et comparaison avec la référence**

Une recherche de cohérence entre les patients appariés, concernant le sexe, l'âge et le lieu de résidence, a été réalisée. Le taux d'appariement (nombre de patients appariés divisé par le nombre de patients issus de l'union PMSI-ALD) a été également calculé. Enfin, l'indicateur a été comparé aux cas incidents de cancers de la zone registre à travers le calcul d'un ratio (figure 3).

**Résultats**

Sur la période d'étude, une stabilité relative dans le temps des effectifs issus du croisement était observée pour chaque cancer. Le nombre de patients hospitalisés ou admis en ALD considérés comme atteints pour la première fois d'un cancer invasif variait selon le cancer (tableau) de 6 800 (lymphome de Hodgkin) à 187 500 (cancer de la prostate).

**Cohérence des informations entre les patients appariés**

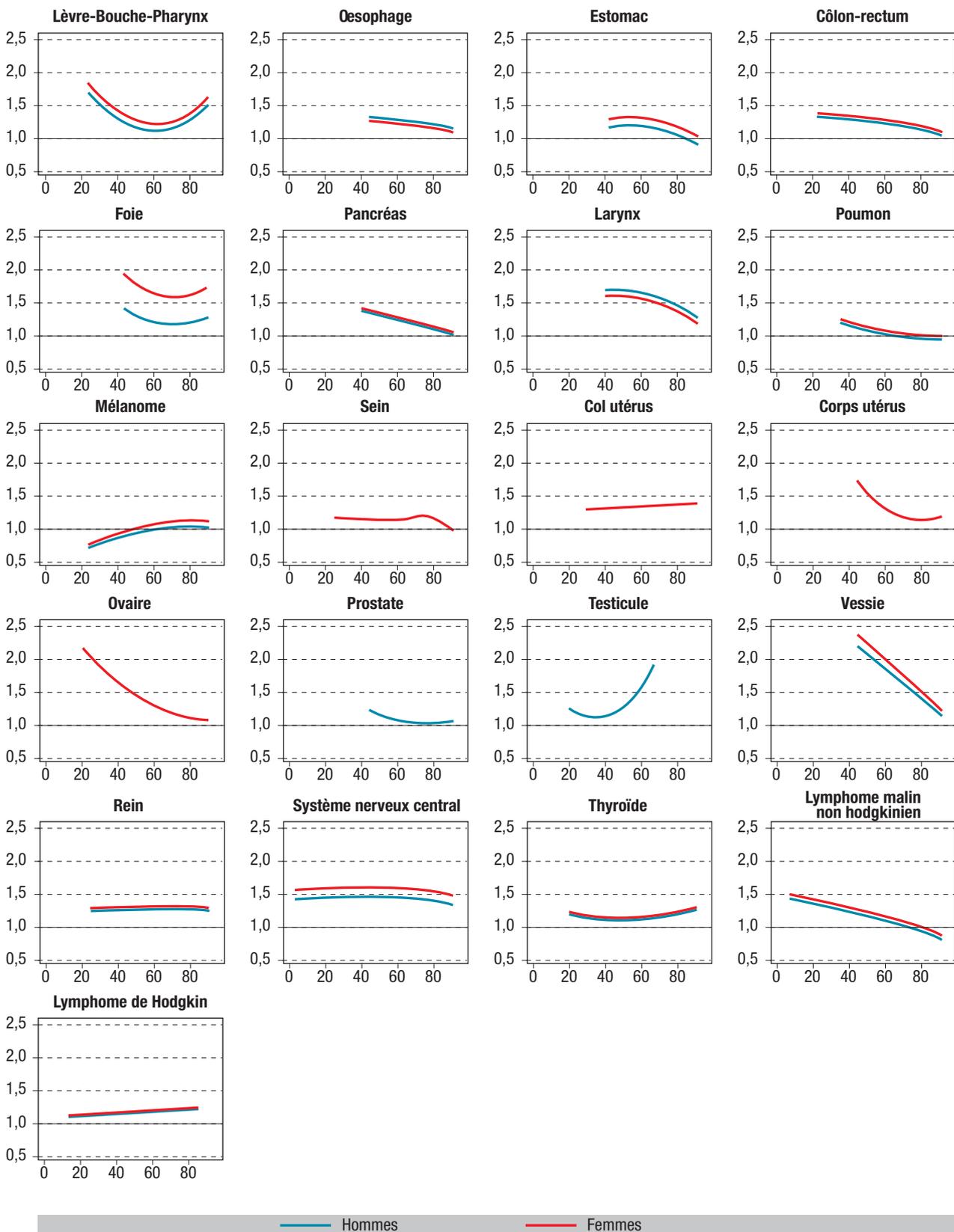
La comparaison individuelle des patients PMSI appariés aux bénéficiaires ALD montrait une cohérence parfaite sur le sexe et sur l'âge. La cohérence de l'information relative au département de résidence variait de 94% à 98% selon le cancer.

**Exclusion des cancers prévalents ou non invasifs, et taux d'appariement**

La majorité des patients exclus grâce au croisement étaient des patients hospitalisés mis en ALD avant leur hospitalisation. Le pourcentage

Figure 3

**Rapport entre le nombre de patients hospitalisés ou admis en ALD considérés comme atteints pour la première fois d'un cancer invasif dénombrés dans la base croisée PMSI-ALD, et le nombre de cas incidents, selon l'âge, par sexe, zone registre, 2006-2008**



L'axe des abscisses correspond à l'âge et l'axe des ordonnées au rapport entre le nombre de patients hospitalisés ou admis en ALD, considérés comme atteints pour la première fois d'un cancer invasif dénombrés dans la base croisée PMSI-ALD, et le nombre de cas incidents de cancer de la zone registre.

**Répartition des patients hospitalisés ou admis en ALD sélectionnés, considérés comme atteints pour la première fois d'un cancer invasif dans la base PMSI-ALD croisée, par localisation cancéreuse, effectifs cumulés 2006-2008**

Localisation cancéreuse	Code CIM-10	Apport du croisement : exclusion supplémentaire de faux-positifs				Patients hospitalisés ou admis en ALD considérés comme atteints pour la première fois d'un cancer invasif		
		Effectif et % de patients PMSI non appariés avec une ALD antérieure à l'hospitalisation (2)	Effectif et % de patients ALD non appariés avec une hospitalisation antérieure à l'ALD (3)	Effectif et % de patients appariés retrouvés parmi les patients PMSI atteints de tumeurs in situ ou bénignes (4)	Effectif « net », après exclusion des cancers non initiaux ou non invasifs (5) <sup>a</sup>	% de patients PMSI appariés aux patients ALD (taux d'appariement)		
Lèvre-Bouche-Pharynx	C00 à C14	5 241	9,7%	247	0,5%	346	48 105	41%
Œsophage	C15	1 071	5,4%	102	0,5%	74	18 452	41%
Estomac	C16	1 286	4,8%	131	0,5%	34	25 419	35%
Côlon-Rectum-Anus	C18 à C21	8 186	5,4%	876	0,6%	516	140 965	46%
Foie	C22	1 306	4,4%	262	0,9%	4	27 877	19%
Pancréas	C25	1 346	4,0%	134	0,4%	7	32 003	33%
Larynx	C32	1 200	6,6%	113	0,6%	182	16 687	31%
Poumon	C33-C34	7 144	6,3%	488	0,4%	127	105 170	39%
Mélanome de la peau	C43	1 585	5,0%	121	0,4%	398	29 646	24%
Sein	C50	208 841	13,6%	323	0,2%	7 402	172 735	59%
Col de l'utérus	C53	14 794	6,1%	30	0,2%	527	13 334	32%
Corps de l'utérus et utérus sans autre indication	C54-C55	25 888	5,2%	136	0,5%	119	24 294	42%
Ovaire	C56	19 318	5,5%	53	0,3%		18 195	38%
Prostate	C61	209 105	9,5%	1 624	0,8%	43	187 547	44%
Testicule	C62	7 803	7,2%	8	0,1%	9	7 222	60%
Vessie	C67	80 581	8,4%	2 107	2,6%	689	71 016	31%
Rein	C64 à C66, C68	41 978	6,0%	194	0,5%	23	39 229	36%
Système nerveux central	C70 à C72	23 127	5,7%	53	0,2%	1 151 <sup>b</sup>	20 599	30%
Thyroïde	C73	28 104	9,6%	228	0,8%	59	25 120	44%
Lymphome malin non hodgkinien	C82 à C85	42 439	7,9%	322	0,8%		38 771	33%
Lymphome de Hodgkin	C81	7 422	7,6%	16	0,2%		6 842	44%

<sup>a</sup> (5) = (1) - [(2) + (3) + (4)]

<sup>b</sup> Correspond à des tumeurs bénignes ou des kystes cérébraux du système nerveux central.

Les zones grisées correspondent à des localisations cancéreuses pour lesquelles il n'existe pas de code CIM-10 de cancer *in situ*.

ALD : Affection longue durée ; PMSI : Programme de médicalisation des systèmes d'information ; CIM-10 : Classification internationale des maladies, 10<sup>e</sup> révision.

de patients PMSI appariés aux bénéficiaires d'ALD variait selon la localisation cancéreuse. Le cancer du foie et le mélanome cutané présentaient un faible taux d'appariement (19% et 24%) et les cancers du sein et du testicule un taux élevé (59% et 60%). Ce taux variait de 30% à 46% pour les autres cancers (tableau).

### Comparaison sur la zone registre, des effectifs issus du croisement PMSI-ALD avec l'incidence

Pour tous les cancers, le nombre de patients hospitalisés ou admis en ALD considérés comme atteints pour la première fois d'un cancer invasif dénombrés dans la base croisée PMSI-ALD était plus élevé que le nombre de cas incidents de cancer invasif de la zone registre (figure 3). Le rapport PMSI-ALD/Incidence variait avec l'âge.

## Discussion

Les informations des bases médico-administratives ne sont pas recueillies à des fins épidémiologiques. Cependant, malgré leurs limites<sup>11</sup>, elles présentent des avantages qui leur permettent d'alimenter certains systèmes de surveillance comme celui des cancers en France. L'interprétation des résultats doit toutefois rester prudente.

La description de l'indicateur issu du croisement montre un taux d'appariement variable de 19% à 60% selon le cancer et un indicateur supérieur à l'incidence de la zone registre, suggérant la persistance de cancers prévalents ou non invasifs parmi les patients non appariés, malgré les exclusions déjà réalisées. En effet, si un appariement faible était attendu pour quelques cancers comme ceux nécessitant un geste chirurgical sans traitement long et coûteux ou ceux essentiellement traités en ambulatoire (mélanome de la peau) ou encore ceux à pronostic sombre à court terme ou siège de tumeur secondaire (foie et système nerveux central), il aurait dû être élevé pour les autres, ce qui n'est globalement pas observé.

Compte tenu de cette limite, la discussion sur l'appariement obtenu ne peut porter que sur ses valeurs extrêmes - élevées pour le testicule et le sein, faibles pour le foie et le mélanome cutané - et vise à confirmer des hypothèses sur le parcours et la prise en charge des patients.

Le cancer du testicule est un cancer rare, qui touche principalement les adultes jeunes entre 20 et 35 ans<sup>12</sup>. Les jeunes atteints sont non seulement des primo-accédants aux ALD, mais également des patients nécessairement pris en charge en secteur hospitalier, compte tenu du mode radical de traitement de ce cancer.

Pour le cancer du sein, la nécessité d'une demande de mise en ALD liée à une prise en charge coûteuse et de longue durée, après un diagnostic de plus en plus précoce grâce au programme de dépistage organisé, expliquerait le taux d'appariement élevé.

Le pronostic du cancer du foie est très mauvais. La survie brute à un an est de 36% et de 16% à 3 ans (période 1989-2007)<sup>13</sup>. Ce mauvais pronostic pourrait être à l'origine de peu de demandes spécifiques d'ALD pour ce cancer au profit de l'utilisation de l'ALD « maladies chroniques actives du foie et cirrhose » le plus souvent à l'origine de ce cancer.

Quant au mélanome cutané, son traitement chirurgical peut être réalisé en ambulatoire ou dans un cabinet privé (où il n'existe pas de PMSI). Par ailleurs, le traitement post-chirurgical peut ne pas s'avérer coûteux ou ne pas faire l'objet d'un long suivi.

Peu d'études françaises se sont intéressées à la qualité du PMSI ou des ALD comme indicateur de suivi épidémiologique des cancers à l'aide de données nominatives<sup>6,14</sup>. Outre la persistance de cancers prévalents difficilement repérables, ces études évoquent aussi, pour expliquer la persistance de faux positifs, des erreurs de codage de certaines lésions tumorales en raison :

- de leur localisation à la frontière d'organes (œsophage-estomac, cancers ORL...);
- d'un cancer métastatique considéré comme primitif (foie, poumon, système nerveux central); par exemple, dans notre étude, le rapport élevé PMSI-ALD/Incidence pour le cancer du foie chez la femme reflète probablement le codage de métastases hépatiques de cancers gynécologiques primitifs;
- de tumeurs bénignes du système nerveux central ou de tumeurs à stade inclassable de la vessie, codées en tumeurs malignes en raison d'une prise en charge hospitalière identique à celle des cancers invasifs, ce qui, dans notre étude, pourrait expliquer en partie le faible appariement observé (environ 30%) pour ces localisations.

Par ailleurs, le passage de la situation d'ayant-droit à celle d'assuré social serait également à l'origine de faux positifs. Le statut d'ayant-droit permet à une personne non reconnue comme assuré social par l'assurance maladie (enfant, conjoint qui n'a jamais travaillé...) de bénéficier d'une protection sociale. Dans les bases de données médico-administratives, les ayants-droits n'ont pas de numéro spécifique et sont enregistrés sous le numéro de l'assuré social dont ils dépendent. D'après une étude de l'assurance maladie, cette situation est très fréquente avant 25 ans et marginale ensuite (de l'ordre de 0,5% à 2%)<sup>15</sup>. Son impact serait limité sur les tumeurs solides, qui surviennent principalement chez des personnes âgées.

Une des possibilités pour exclure davantage de faux positifs consisterait à rendre plus spécifique les algorithmes de sélection. Dans la littérature, l'utilisation épidémiologique du PMSI repose sur la sélection des séjours à partir des codes diagnostiques de la pathologie étudiée, associés ou non aux actes thérapeutiques. Ainsi, pour les tumeurs

du système nerveux central, la sélection en « DP ou DR » permettrait d'observer une diminution des faux positifs avec une hausse de la valeur prédictive positive, mais au prix d'une diminution de la sensibilité<sup>14</sup>. Dans notre étude, la sélection PMSI reposait sur le DP afin d'exclure d'emblée les cancers prévalents (patient en cours de suivi médical, patient dont le cancer est connu et qui est hospitalisé pour une autre pathologie, récurrence) et de diminuer, pour certains cancers, le risque de variabilité géographique du rapport Incidence/PMSI par rapport à une sélection incluant les actes chirurgicaux<sup>16</sup>. La sélection PMSI associant « DP ou DR » a été testée : la proportion de faux positifs exclus avec une sélection en « DP » variait, selon le cancer, de 4,4% à 17,3% contre 5,4% à 20,4% avec une sélection en « DP ou DR » sans gain sur le taux d'appariement, sauf pour le mélanome cutané (+1 point) et le cancer du poumon (+3 points).

Autre option envisageable au niveau de la sélection PMSI : l'allongement de la durée de chaînage, fixée à deux ans dans l'étude, afin d'exclure également d'emblée les cancers prévalents. Les travaux relatifs aux estimations d'incidence à partir du PMSI retrouvés dans la littérature<sup>8,9</sup> ne comportant pas de chaînage entre les années, il était difficile de se prononcer sur une durée de chaînage pertinente. Toutefois, l'extension du chaînage à quatre années antérieures, réalisée pour l'année 2008, permettait de montrer qu'au moins 90% des cancers prévalents étaient repérés dans les deux premières années de chaînage, sauf pour les cancers du sein, de la prostate et de la vessie, pour lesquels une année supplémentaire de chaînage était nécessaire pour atteindre ce résultat.

Dans le Sniiram, la sélection des bénéficiaires d'ALD reposait sur les patients mis en ALD une année donnée et ayant réalisé leur première consommation de soins en lien avec l'ALD la même année. Ce processus ne permettait pas de prendre en compte les 4% à 7% de bénéficiaires ayant débuté leur consommation de soins postérieurement à l'année de leur mise en ALD. Toutefois, la sélection élargie ne permettait ni d'exclure davantage de faux positifs ni d'augmenter le taux d'appariement.

## Conclusion et perspectives

La méconnaissance de l'ampleur et du devenir des faux positifs issus des deux sources croisées semble poser davantage de questions qu'en résoudre, en termes d'impact sur les faux positifs : communs aux deux sources ou propres à une seule source ? En l'absence de littérature relative au croisement de bases médico-administratives, l'évaluation de la qualité de l'indicateur PMSI-ALD croisé pour le suivi épidémiologique des cancers devrait être envisagée à l'aide de données nominatives issues des registres. Ces travaux complèteraient la perspective prochaine de production d'estimations infranationales de l'incidence à partir de l'indicateur PMSI-ALD, sous réserve de sa validation. ■

## Remerciements

Les auteurs remercient vivement Zoé Uhry, qui a réalisé la comparaison de l'indicateur avec la zone registre, Laurent Duchet, David Dias, Sophie Gosselin, Mathilde Risse-Fleury et Pascale Grosclaude pour leur aide précieuse, ainsi que les relecteurs pour leurs remarques pertinentes.

## Références

- [1] Bossard N, Gomez F, Remontet L, Uhry Z, Olive F, Mitton N, *et al.* Utilisation des données du PMSI pour estimer l'incidence des cancers en France à l'échelon infranational: la démarche du groupe Oncepi. *Bull Epidemiol Hebd.* 2012;(5-6):77-80. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=10441](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=10441)
- [2] Grémy I. Portrait des registres des cancers en France. *Environ Risques Santé.* 2012;11(2):162-5. [http://www.jle.com/fr/revues/sante\\_pub/ers/e-docs/00/04/76/3E/article.phtml](http://www.jle.com/fr/revues/sante_pub/ers/e-docs/00/04/76/3E/article.phtml)
- [3] Binder-Foucard F, Belot A, Delafosse P, Remontet L, Woronoff A, Bossard N. Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2012. Étude à partir des registres des cancers du réseau Francim - Partie 1 : tumeurs solides. Saint-Maurice: Institut de veille sanitaire; 2013. 122 p. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=11619](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=11619)
- [4] Colonna M, Mitton N, Remontet L, Belot A, Delafosse P, *et al.* Éléments d'interprétation des estimations régionales de l'incidence du cancer en France sur la période 1980-2005. *Rev Epidemiol Santé Publique.* 2008;56(6):434-40.
- [5] Couris C, Polazzi S, Olive F, Remontet L, Bossard N, Gomez F, *et al.* Breast cancer incidence using administrative data: correction with sensitivity and specificity. *J Clin Epidemiol.* 2009;62(6):660-6.
- [6] Grosclaude P, Dentan C, Trétarre B, Velten M, Fournier E, Molinié F. Utilité des bases de données médico-administratives pour le suivi épidémiologique des cancers. Comparaison avec les données des registres au niveau individuel. *Bull Epidemiol Hebd.* 2012;(5-6):63-7. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=10438](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=10438)
- [7] Uhry Z, Remontet L, Grosclaude P, Belot A, Colonna M, Bossard N, *et al.* Tendances récentes des données d'affections de longue durée (ALD) : intérêt pour la surveillance nationale de l'incidence des cancers, période 1997-2009, France. *Bull Epidemiol Hebd.* 2012;(5-6):58-63. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=10437](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=10437)
- [8] Uhry Z, Remontet L, Colonna M, Belot A, Grosclaude P, Mitton N, *et al.* Cancer incidence estimation at a district level without a national registry: a validation study for 24 cancers sites using French health insurance and registry data. *Cancer Epidemiol.* 2013;37(2):99-114.
- [9] Estimations régionales de l'incidence par cancer, 2008-2010. In: Surveillance épidémiologique des cancers. Dossier thématique. Saint-Maurice: Institut de veille sanitaire (site Internet). <http://www.invs.sante.fr/Dossiers-thematiques/Maladies-chroniques-et-traumatismes/Cancers/Surveillance-epidemiologique-des-cancers/Estimations-de-l-incidence-et-de-la-mortalite/Estimations-regionales-de-l-incidence-par-cancer-2008-2010>.
- [10] Définition d'une affection de longue durée exonérante et d'une affection de longue durée non exonérante. Assurance maladie. <http://www.ameli.fr/assures/soins-et-remboursements/combien-serez-vous-rembourse/en-cas-d-affection-de-longue-duree/qu-est-ce-qu-une-affection-de-longue-duree-ald.php>.
- [11] Kudjawa Y, Rudant J, Decool E, Danzon A, Grémy I, Musset A, *et al.* Système multi sources de surveillance des cancers. Croisement des données ALD et PMSI. Description des données 2006-2008 et construction d'indicateurs. Saint-Maurice: Institut de veille sanitaire; 2012. 70 p. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=10637](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=10637).

[12] Kudjawu Y, Danzon A, Bloch J. Évolution nationale et variations régionales du taux de patients opérés pour cancer du testicule en France, 1998-2008. *Bull Epidémiol Hebd.* 2012;(7-8-9):106-10. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=10472](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=10472).

[13] Grosclaude P, Remontet L, Belot A, Danzon A, Rasamimanana-Cerf N, Bossard N. Survie des personnes atteintes de cancer en France 1989-2007. Étude à partir des registres des cancers du réseau Francim. Saint-Maurice: Institut de veille sanitaire; 2013. 412 p. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=11316](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=11316).

[14] Coureau G, Baldi I, Savès M, Jaffré A, Barat C, Gruber A, *et al.* Évaluation des performances du PMSI pour l'identification des tumeurs incidentes du système nerveux central par rapport à un registre spécialisé en Gironde, France, en 2004. *Rev Epidémiol Santé Publique.* 2012;60(4):295-304.

[15] Chaignot C, Weill A, Blotière PO, de Roquefeuil L, Ricordeau P, Alla F, *et al.* Intérêt de l'identifiant bénéficiaire anonyme unique vie entière dans le Sniiram et de son utilisation après chaînage dans le PMSI. *Rev Epidémiol Santé Publique.* 2013;61S:S16-S17.

[16] Mitton N, Colonna M, Trombert B, Olive F, Gomez F, Iwaz J, *et al.* A suitable approach to estimate cancer incidence in area without cancer registry. *J Cancer Epidemiol.* 2011;2011:418968. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3065037/>.

#### Citer cet article

Kudjawu Y, de Maria F, Decool E, Altana M, Harlin JM, Weill A, *et al.* Croisement de deux bases de données médico-administratives : méthodologie et étude descriptive pour une application à la surveillance épidémiologique des cancers en France. *Bull Epidémiol Hebd.* 2013; (Hors-série):49-58.