

Surveillance des maladies chroniques au Québec par le jumelage des fichiers administratifs.

Louis Rochette, Valérie Émond¹

Résumé

En surveillance des maladies chroniques, la cueillette d'information est essentielle dans le but de mesurer l'ampleur des maladies, évaluer l'utilisation des services, identifier les groupes à risque et suivre l'évolution des maladies et des facteurs de risque dans le temps en vue de planifier et d'appliquer des programmes de prévention en santé publique. C'est dans ce contexte qu'a été créé le Système intégré de surveillance des maladies chroniques du Québec (SISMACQ). Le SISMACQ est une base de données issue du jumelage de fichiers administratifs, couvrant les années 1996 à 2013, représentant une alternative intéressante aux données d'enquête puisqu'elle porte sur la totalité de la population, n'est pas affectée par le biais de rappel et permet de suivre la population dans le temps et dans l'espace. Cette présentation a pour objet de présenter la pertinence de l'utilisation des données administratives comme alternative aux données d'enquête ainsi que de les méthodes choisies pour la construction de la cohorte populationnelle à partir du jumelage des différentes sources de données brutes ainsi que de décrire les traitements apportées dans le but de minimiser les biais. Nous discuterons également des avantages et des limites liés à l'analyse des fichiers administratifs.

Mots Clés : Données administratives; jumelage; surveillance.

1. Présentation du système intégré de surveillance des maladies chroniques du Québec

1.1 Introduction

Les maladies chroniques sont maintenant reconnues comme un problème majeur de santé publique. Plusieurs facteurs dont l'environnement, les habitudes de vie et le vieillissement de la population contribueront à l'augmentation du fardeau associé à ces maladies, au cours des prochaines années. Dans ce contexte, la surveillance des maladies chroniques ayant les prévalences les plus élevées telles que le diabète, les maladies cardiovasculaires, les maladies respiratoires, les maladies ostéoarticulaires, l'ostéoporose, les troubles mentaux et les démences ont été retenues afin de construire un système de surveillance des maladies chroniques pour la province de Québec.

Le mandat de développer la surveillance des maladies chroniques au Québec a été confié à l'Institut national de santé publique du Québec (INSPQ) par le ministère de la Santé et des Services sociaux du Québec (MSSS). Cette surveillance populationnelle s'appuie sur le jumelage de fichiers administratifs et vise autant une approche par maladie qu'une approche intégrée. Le financement du projet est assuré par l'Agence de la santé publique du Canada, le MSSS et l'INSPQ. Du financement ponctuel s'ajoute dans le cadre de projets de recherche liés à la validation des données et au développement méthodologique

1.2 Objectifs de surveillance

Les principaux objectifs du système de surveillance est de mesurer l'ampleur de chacune des maladies, d'analyser la mortalité, la survie, les incapacités, d'évaluer l'utilisation des services de santé, d'identifier les groupes à risque et d'analyser les trajectoires de soins. L'approche intégrée a comme objectifs de décrire et d'analyser les comorbidités

¹Louis Rochette, Institut national de santé publique du Québec, 945 avenue Wolfe, Québec, G1V 5B3
Louis.rochette@inspq.qc.ca; Valérie Émond, Institut national de santé publique du Québec, 945 avenue Wolfe, Québec, G1V 5B3 Valerie.Emond@inspq.qc.ca.

associées aux maladies, de caractériser les personnes atteintes de plus d'une maladie et de faire un portrait de la multimorbidité

1.3 Méthodologie.

Le système est basé sur le jumelage de données issues de fichiers administratifs et est appelé le système intégré de surveillance des maladies chroniques du Québec (SISMACQ) (Blais et al., 2014). Les données couvrent une période s'étendant du 1er janvier 1996 au 31 mars 2013 et sont mises à jour au mois de juin de chaque année.

Cinq fichiers administratifs jumelés composent la base de données du SISMACQ : le registre des personnes assurées à l'assurance-maladie du Québec, le fichier des hospitalisations, le fichier des décès, le fichier des services médicaux rémunérés à l'acte et le fichier des services pharmaceutiques (dans le cas de ce dernier fichier, pour les plus de 65 ans). Le numéro d'assurance maladie est utilisé comme clé de jumelage des fichiers.

Plus spécifiquement, le registre des personnes assurées à l'assurance-maladie du Québec contient des informations démographiques et géographiques ainsi que des renseignements sur l'admissibilité aux régimes publics d'assurance maladie et d'assurance médicaments. Le fichier des hospitalisations regroupe des informations relatives aux séjours en centres hospitaliers québécois. Les données comprennent de l'information sur les séjours hospitaliers (lieu, durée, provenance et destination de l'individu), les diagnostics, les services les soins intensifs et les interventions. Le fichier des décès inclut les enregistrements de tous les décès de Québécois, y compris ceux survenus à l'extérieur de la province. L'information provient du médecin qui constate le décès ou du coroner et comprend la date et les causes du décès. Le fichier des services médicaux rémunérés à l'acte rassemble les données liées à la facturation à l'acte des professionnels, c'est-à-dire les demandes de paiement transmises par les professionnels de la santé à la Régie d'assurance-maladie du Québec (RAMQ.) Cela comprend le code de l'acte médical rendu et un code de diagnostic. Sont également présentes les informations relatives aux professionnels de la santé ainsi que sur le lieu du service. Finalement, le fichier des services pharmaceutiques centralise les réclamations de médicaments dans le cadre du régime public d'assurance médicaments. Les informations issues des régimes privés d'assurances, auxquels souscrit une part importante de la population québécoise, ne s'y retrouvent donc pas. En raison de l'admissibilité automatique au régime public à partir de 65 ans, cette sous-population est couverte à plus de 90 % (exclusion des régimes privés et des personnes vivant en institution). Les données englobent les informations relatives aux médicaments prescrits (code du médicament, dose, nombre de renouvellements, durée du traitement, etc.), ainsi qu'aux professionnels de la santé (classe du prescripteur, spécialité).

Puisque les fichiers présentés contiennent peu d'informations à caractère socio-économique, l'indice de défavorisation matérielle et sociale, un substitut écologique du statut socio-économique développé à l'INSPQ est intégré au SISMACQ (Pampalon, 2009). Créé à partir d'indicateurs du recensement canadien à l'échelle de l'aire de diffusion, la plus petite unité géographique pour laquelle des données censitaires sont produites, l'indice est assigné aux individus grâce à un fichier de correspondance entre les aires de diffusion et les codes postaux du registre des personnes assurées.

À partir des données sources, les personnes atteintes d'une maladie sont identifiées à l'aide d'une définition de cas spécifique. Ces définitions sont construites à partir de combinaisons de diagnostics présents au fichier des services médicaux, ou au fichier des hospitalisations, ainsi que des prescriptions de certains médicaments spécifiques à la maladie enregistrés au fichier des services pharmaceutiques. Les conditions requises varient selon la maladie étudiée. Par exemple, la définition de cas de l'hypertension est deux diagnostics d'hypertension enregistrés au fichier des services médicaux à l'intérieur d'une période de deux ans ou un diagnostic d'hypertension recensé au fichier des hospitalisations (les prescriptions médicamenteuses ne sont pas considérées pour cette définition de cas).

Suite à la construction de la définition de cas, des indicateurs de prévalence, d'incidence, de mortalité et d'utilisation des services sont construits. Ces données sont compilées et interprétées pour être ensuite publiées selon plusieurs modes de diffusion tel que la transmission de données agrégées via un site web de l'Infocentre de santé publique du Québec, la transmission de données agrégées sommaires à l'Agence de la santé publique du Canada, et la production de différents types de publications tel que des rapports de surveillance et méthodologiques produits par l'INSPQ ainsi que des articles scientifiques.

2. Données administratives comme alternative aux données d'enquête

2.1 Avantages et limites des données administratives jumelées

Les données administratives représentent une alternative intéressante aux données d'enquête. Ces données utilisent les sources de données déjà existantes, produits au moment propice, couvre l'ensemble de la population, et ne sont pas influencées par le biais de rappel. De plus, ces données peuvent être recueillies sur une longue période de temps et permettre donc de suivre l'évolution des individus à travers le temps.

Cependant, les données administratives ont été conçus à d'autres fins que la surveillance des maladies et ne sont donc pas optimales sur certains aspects. Ces données renferment peu d'informations cliniques détaillées et sur les habitudes de vie, plusieurs variables contenues dans ces banques ne sont pas validées et peuvent comporter certaines inexacuitudes. Les données administratives ne comprennent de l'information que sur les gens ayant eu recours aux services de santé et dont l'état a été diagnostiqué par un médecin.

2.2 Comparaison des indicateurs obtenus à partir des données administratives et des enquêtes.

Dans le but de mesurer l'influence de l'utilisation des données administratives sur les indicateurs, des estimations d'hypertension et de maladies pulmonaires obstructives chroniques obtenus à l'aides des bases de données administratives ont été comparées avec des données d'enquête portant sur de l'auto-déclaration ainsi qu'avec des données d'enquêtes où des mesures cliniques étaient enregistrées.

Le tableau 2.2.1 présente les résultats d'une comparaison canadienne des maladies pulmonaires obstructives chroniques entre 2005 et 2009 (Doucet et al., 2013). Les comparaisons s'appuient sur des données provenant de l'Enquête sur la santé dans les collectivités canadiennes (ESCC), l'Enquête canadienne sur les mesures de la santé (ECMS) et sur l'étude *Canadian Obstructive Lung Disease cohort* (COLD). Les estimations sur les données administratives ont été obtenus via le système canadien de surveillance des maladies chroniques coordonné par l'Agence de santé publique du Canada qui est un réseau de système de surveillance impliquant toutes les provinces et territoires où la participation du Québec se fait à travers du SISMACQ.

L'ESCC est une enquête transversale menée par Statistique Canada dont les données sont collectées annuellement depuis 2007 ainsi qu'à chaque deux ans entre 2001 et 2005 (Statistique Canada, 2014). Le thème des maladies pulmonaires obstructives chroniques était abordé à travers la question « *Avez-vous été diagnostiqué par un professionnel de la santé de bronchite chronique, d'emphysème ou d'une maladie pulmonaire obstructive chronique?* »

L'ECMS est une enquête transversale administrée par Statistique Canada qui permet de recueillir des mesures physiques directes à l'aide d'un centre d'examen mobile (Statistique Canada, 2014). Une de ces mesures est la spirométrie qui est un test de contrôle de la fonction pulmonaire permettant de diagnostiquer certaines pathologies respiratoires.

COLD est une enquête transversale qui a été menée dans cinq grandes villes de différentes provinces canadiennes entre 2005 et 2009 et qui incluait un questionnaire où les problèmes de santé respiratoires étaient auto-déclarés ainsi qu'un test de spirométrie (Tan et al., 2011).

Les prévalences comparées au tableau 2.2.1 sont brutes. Les estimations obtenues par des mesures cliniques sont près de l'estimation obtenue à l'aide des données administratives alors que les indicateurs obtenus par auto-déclaration semblent sous-estimer la prévalence.

Le tableau 2.2.2 présente les résultats d'une comparaison des estimations de l'hypertension pour le Québec en 2009 (Blais et al. 2013). Les données comparées proviennent des enquêtes ESCC et ECMS alors que le SISMACQ a été utilisé pour les estimations provenant des données administratives. L'estimation de l'ESCC a été obtenue par auto-déclaration à travers la question « *Avez-vous été diagnostiqué par un professionnel de santé d'hypertension?* ». L'estimation de l'ECMS s'appuie sur séries de mesures de tension artérielle ainsi que d'une auto-déclaration de prise de médicaments antihypertenseurs dans le dernier mois.

Les prévalences présentées au tableau 2.2.2 sont ajustées selon l'âge. Encore une fois, les estimations obtenues à l'aide des données administratives sont plus près des estimations basées sur des mesures cliniques alors que les estimations auto-déclarées semblent sous-estimer la prévalence.

Tableau 2.2.1

Prévalence des maladies pulmonaires obstructives chroniques au Canada, 35 ans et plus, de 2005 à 2009.

Étude	Prévalence
Données administratives	8,3%
Enquête ESCC auto-déclarée	4,8%
Enquête COLD auto-déclarée	6,9%
Enquête COLD mesures cliniques	7,9%
Enquête ECMS mesures cliniques	8,1%

Tableau 2.2.2

Prévalence ajustée de l'hypertension au Québec, 20-79 ans, en 2009.

Étude	Prévalence
Données administratives	20,6%
Enquête ESCC auto-déclarée	18,0%
ECMS mesures cliniques / prise médicaments	22,6%

3. Qualité des données : survol des avantages et limites des fichiers administratifs

3.1 Qualité du jumelage

La Régie de l'assurance-maladie du Québec a été mandatée pour fusionner les cinq fichiers administratifs mentionnés à la section 1.3. La Régie est propriétaire du registre des personnes assurées, du fichier des services médicaux et du fichier des services pharmaceutiques. La Régie est de plus responsable de la création du numéro d'assurance-maladie qui est l'identifiant unique désigné comme la clé d'appariement. Ainsi, tous les enregistrements de ces trois fichiers sont fusionnés.

Le fichier des hospitalisations et le fichier des décès proviennent d'autres sources et sont donc sujet à des erreurs de jumelage en raison de numéros d'assurance-maladie absents ou mal codés. Pour le fichier des hospitalisations, la proportion d'enregistrements jumelés s'élève à 99% alors qu'il est de 97% pour le fichier des décès. La majorité des enregistrements manquants pour les hospitalisations concerne certains nouveau-nés pour lesquels il y a eu une erreur de transcription entre leur identifiant temporaire assigné à leur naissance et l'identifiant permanent assigné quelques mois plus tard. Les enregistrements manquants du fichier des décès induisent peu de biais différentiel à l'exception des morts violentes ayant fait l'objet d'enquête du coroner. En effet, le coroner n'est pas obligé d'inscrire le numéro d'assurance-maladie dans le bordereau de décès. Il est donc possible qu'un biais dans l'évaluation des causes de décès soit présent en particulier pour les gens diagnostiqués avec troubles mentaux qui sont plus enclins à mourir par accidents ou traumatismes.

3.2 Forces et limites des fichiers administratifs.

Le registre des personnes assurées est la meilleure source disponible pour définir la cohorte populationnelle. L'assurance-maladie est financée par le secteur public et couvre presque toute la population. De plus, les estimations des indicateurs de surveillance sont plus précises en utilisant le registre en comparaison avec les estimations démographiques issues du recensement puisque les cas inclus au numérateur sont tous inclus au dénominateur.

Nous avons tout de même comparé les estimations de population provenant du registre à ceux du recensement pour s'assurer que le dénominateur considéré pour le calcul des indicateurs était conforme à la répartition démographique du Québec. La distribution âge sexe de la population issue des deux sources est fort semblable à l'exception des jeunes adultes où la population estimée par le registre des personnes assurées est de 5 à 10% inférieure à celle des estimations démographiques du recensement selon les années de couverture du SISMACQ. Une vérification détaillée a permis de constater une proportion élevée de jeunes adultes qui tardait ou qui ne renouvelait pas leur carte d'assurance-maladie. La Régie d'assurance-maladie émet des cartes pour une durée de 4 ans et l'utilisateur a la responsabilité de procéder au renouvellement. La solution proposée a été d'assigner automatiquement une carte d'assurance-maladie aux hommes de 18-30 ans et aux femmes de 18-24 ans. Cette procédure risquait de provoquer une surestimation en raison des individus n'étant effectivement plus éligible à l'assurance-maladie. Cependant, les comparaisons de l'estimation du registre suite à l'ajustement ont montré très peu de différences avec les données du recensement.

Le fichier des services médicaux rémunérés à l'acte est une puissante source d'informations avec plus d'un milliard d'observations pour les 17 ans d'existence du système de surveillance. Ce fichier est la principale source de données pour identifier les cas. Parmi les informations utiles, on recense les codes de diagnostics, les codes d'actes médicaux, le type d'établissement et la spécialité du médecin. La principale limite de ce fichier est que l'information est restreinte à la rémunération à l'acte. Or, les médecins du Québec peuvent également se faire payer par forfaits, par vacations, ou à salaire. Un biais potentiel pourrait être présent puisque ces formes de paiements alternatifs ont augmenté depuis le début de l'existence du système de surveillance. Selon l'Institut Canadien de l'information de la santé (ICIS, 2014), la proportion de modes alternatifs de rémunération clinique des médecins est passée de 16% en 1999-2000 à 24% en 2011-2012 atteignant un sommet de 26% en 2008-2009. La vérification de cette information au cours des prochaines années sera très importante afin de pouvoir documenter les biais potentiels dans l'évolution du nombre de cas.

Le fichier des hospitalisations contient des informations validées sur l'ensemble des admissions hospitalières de soins de courte durée. Des archivistes médicaux révisent les dossiers d'admission et codent les données sommaires fournies par les médecins. L'information enregistrée dans ce fichier est donc valide et un seul diagnostic suffit généralement pour identifier un cas à partir de ce fichier. La principale limite de ce fichier est que les individus soignés à l'urgence et qui n'ont pas fait l'objet d'une admission hospitalière subséquente n'ont pas d'enregistrements dans le fichier des hospitalisations. Ces individus sont identifiés seulement si le médecin a facturé à l'acte les services médicaux rendus à l'urgence.

Le fichier des services pharmaceutiques est également une puissante source d'information puisqu'elle contient plus de un milliard d'observations pour l'ensemble de la période à l'étude. La population à l'étude est cependant limitée aux aînés de 65 ans et plus participant au régime public d'assurance-médicaments. La banque n'enregistre pas les médicaments prescrits aux aînés cotisant à des régimes privés d'assurance-médicaments ainsi qu'aux personnes en hébergement ou hospitalisées. Il est estimé qu'environ 10% des aînés sont ainsi exclus.

Le fichier des décès a comme principale utilité de documenter les causes de décès. La disponibilité de ce fichier est généralement retardée de deux ou trois ans en comparaison avec les autres fichiers. Le registre des personnes assurées est donc utilisé pour le recensement de la date de décès. Une source potentielle de biais de ce fichier est le changement du mode de codification qui a été apporté à partir de janvier 2000. Avant cette date, une seule cause de décès codée selon la 9^e révision de la classification internationale des maladies était enregistrée. Depuis ce temps, la cause principale ainsi qu'une possibilité de 10 causes secondaires codées selon la 10^e révision de la classification internationale des maladies sont saisies. Il est donc important d'étudier l'impact de ces changements au cours de la période d'étude.

4. Conclusion

Le système intégré de surveillance des maladies chroniques du Québec renferme une quantité impressionnante d'informations. Les forces associées à ce système dépassent largement les limites. Ce système s'avère peu coûteux et produit au moment propice des données fiables pour l'ensemble de la population. Il demeure néanmoins important de bien connaître les limites et de les documenter car celles-ci sont généralement différentes des enquêtes par échantillonnage.

Bibliographie

Blais C, Rochette L, Hamel D (2013), Prévalence de l'hypertension au Québec, comparaison entre les données médico-administratives et les données d'enquête, BEH, Institut de veille sanitaire, Paris, France.

Blais C, Jean S, Sirois C, Rochette L, Plante C, Larocque I, Doucet M, Ruel G, Simard M, Gamache P, Hamel D, St-Laurent D, Émond V (2014), « Le système intégré de surveillance des maladies chroniques du Québec (SISMACQ) , une approche novatrice (Quebec Integrated Chronic disease surveillance system (QICDSS), an Innovative approach)», *Maladies chroniques et blessures au Canada*, Vol. 34 no 4, p. 247-256.Sous presse

Doucet M, Rochette L, Gagné ME, Gamache P, Pouliot S, Émond V.(2013) Développement de la surveillance des maladies respiratoires au Québec à partir des données médico-administratives, BEH, Institut de veille sanitaire, Paris, France.

ICIS (2014), Site Web,<https://secure.cihi.ca/estore/productFamily.htm?locale=fr&pf=PFC2368>

Pampalon R, Hamel D, Gamache P, Raymond G. (2009) A deprivation index for health planning in Canada. *Chronic Dis Can*; 29(4):178-191.

Statistique Canada (2014), Site Web, www.statcan.gc.ca.

Tan WC, Bourbeau J, Fitzgerald JM, Cowie R, Chapman K Hernandez P, Buits S, Sin DD. (2011), « 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):1691-1698.