

Chronic-disease surveillance in Quebec using administrative file linkage

Louis Rochette, Valérie Émond¹

Abstract

Information collection is critical for chronic-disease surveillance to measure the scope of diseases, assess the use of services, identify at-risk groups and track the course of diseases and risk factors over time with the goal of planning and implementing public-health programs for disease prevention. It is in this context that the Quebec Integrated Chronic Disease Surveillance System (QICDSS) was established. The QICDSS is a database created by linking administrative files covering the period from 1996 to 2013. It is an attractive alternative to survey data, since it covers the entire population, is not affected by recall bias and can track the population over time and space. In this presentation, we describe the relevance of using administrative data as an alternative to survey data, the methods selected to build the population cohort by linking various sources of raw data, and the processing applied to minimize bias. We will also discuss the advantages and limitations associated with the analysis of administrative files.

Keywords: Administrative data; Linkage; Surveillance.

1. Description of the Quebec Integrated Chronic Disease Surveillance System

1.1 Introduction

Chronic diseases are now recognized as a major public-health problem. A number of factors, including the environment, lifestyle and population aging, will contribute to the growing burden associated with these diseases in the coming years. In this context, the most prevalent chronic diseases, such as diabetes, cardiovascular diseases, respiratory diseases, bone and joint disorders, osteoporosis, mental disorders and dementia, were selected to construct a chronic-disease surveillance system for Quebec.

The ministère de la Santé et des Services sociaux du Québec (MSSS) mandated the Institut national de santé publique du Québec (INSPQ) to develop chronic-disease surveillance in Quebec. This population surveillance is based on the linkage of administrative files and involves both a disease-specific approach and an integrated approach. The Public Health Agency of Canada, the MSSS and the INSPQ provide funding for the project. Additional specific funding is allocated to research projects associated with data validation and methodological development.

1.2 Goals of surveillance

The main goals of the surveillance system are to measure the scope of each disease; analyze mortality, survival and disabilities; assess the use of health services; identify at-risk groups; and analyze care pathways. The purposes of the integrated approach are to describe and analyze the comorbidities associated with diseases, profile people with multiple diseases and paint a picture of multimorbidity.

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

1.3 Methodology

The system is based on the linkage of data from administrative files and is known as the Quebec Integrated Chronic Disease Surveillance System (QICDSS) (Blais et al. 2014). The data cover the period from January 1, 1996, to March 31, 2013, and are updated every year in June.

The QICDSS database is composed of five linked administrative files: Quebec's health insurance registry, the hospitalization file, the deaths file, the fee-for-service file, and the pharmaceutical services file (for people over 65). The health insurance number is used as the file linkage key.

More specifically, Quebec's health insurance registry contains demographic and geographic data as well as information about eligibility for public health insurance and drug insurance plans. The hospitalization file contains information about stays in Quebec hospitals. The data consist of information about hospital stays (location, length, and origin and destination of the individual), diagnoses, services, intensive care and procedures. The deaths file contains the records of all deaths of Quebecers, including deaths that occur outside the province. The information comes from the reporting physician or the coroner and includes the date and causes of death. The fee-for-service file gathers data related to fee-for-service billings by health professionals, i.e., payment claims that health professionals submit to the Régie d'assurance-maladie du Québec (RAMQ). This includes the procedure code and a diagnostic code. Also included are data on health professionals and the location where the service was provided. The pharmaceutical services file centralizes prescription drug claims submitted under the public drug insurance plan. The file does not include data from private insurance plans, which cover a large portion of the Quebec population. Since people aged 65 and over are automatically covered by the public plan, the file includes over 90% of this group (it excludes seniors covered by a private plan or living in an institution). The records contain information about prescription drugs (e.g., drug code, dose, number of renewals, and duration of treatment) and health professionals (class and specialty of the prescriber).

Since the above-mentioned files contain little socioeconomic information, the QICDSS includes the material and social deprivation index, an ecological substitute for socioeconomic status developed by the INSPQ (Pampalon et al. 2009). Based on indicators from the Canadian census at the dissemination-area level, the smallest geographic unit for which census data are produced, the index is assigned to individuals through a correspondence file that links dissemination areas and postal codes from the health insurance registry.

From the source data, the QICDSS identifies people with chronic diseases using specific case definitions. These definitions are constructed from combinations of diagnoses in the fee-for-service file or the hospitalization file, and from prescriptions of certain disease-specific drugs in the pharmaceutical services file. The required conditions vary with the disease being studied. For example, the case definition of hypertension is two diagnoses of hypertension in the fee-for-service file within a two-year period or one diagnosis of hypertension in the hospitalization file (drug prescriptions are not considered for this case definition).

After the case definition is constructed, indicators of prevalence, incidence, mortality and service use are constructed. These data are compiled, interpreted and then disseminated in various forms. For example, aggregate data are released through a website of the Infocentre de santé publique du Québec [the Quebec public-health information centre]; summary aggregate data are provided to the Public Health Agency of Canada; and publications of various types are produced, such as INSPQ surveillance and methodology reports and scientific papers.

2. Administrative data as an alternative to survey data

2.1 Advantages and limitations of linked administrative data

Administrative data are an attractive alternative to survey data. They come from existing data sources, are produced on a timely basis, cover the entire population, and are not affected by recall bias. In addition, they can be collected over a long period of time and, consequently, can be used to track individuals over time.

However, administrative data were not designed for disease surveillance and therefore, are not optimal in some respects. They contain little detailed clinical or lifestyle information, and a number of variables in the databases are

not validated and may contain inaccuracies. Administrative data provide information only about people who have used health services and been diagnosed by a physician.

2.2 Comparison of indicators based on administrative data and indicators based on survey data

To measure the impact of administrative data on the indicators, estimates of hypertension and chronic obstructive pulmonary disease derived from administrative databases were compared with self-reported survey data and with survey data based on clinical measurements.

Table 2.2.1 shows the results of a Canadian comparison of chronic obstructive pulmonary disease between 2005 and 2009 (Doucet et al. 2013). The comparisons are based on data from the Canadian Community Health Survey (CCHS), the Canadian Health Measures Survey (CHMS) and the Canadian Cohort of Obstructive Lung Disease (CanCOLD) study. Estimates from administrative data were obtained through the Canadian Chronic Disease Surveillance System (CCDSS). Coordinated by the Public Health Agency of Canada, the CCDSS is a network of surveillance systems involving all provinces and territories. Quebec participates in the CCDSS through the QICDSS.

The CCHS is a cross-sectional survey conducted by Statistics Canada for which data were collected every two years between 2001 and 2005 and annually since 2007 (Statistics Canada 2014). The topic of chronic obstructive pulmonary disease was raised in the question “Have you been diagnosed with chronic bronchitis, emphysema or chronic obstructive pulmonary disease by a health professional?”

The CHMS is a cross-sectional survey administered by Statistics Canada that collects direct physical measures at a mobile examination centre (Statistics Canada 2014). One of these measures is spirometry, a lung function test used to diagnose certain respiratory pathologies.

The CanCOLD study is a cross-sectional survey that was conducted in five major cities in various Canadian provinces between 2005 and 2009; it included a spirometry test and a questionnaire in which respondents self-reported respiratory health problems (Tan et al. 2011).

The prevalence figures compared in Table 2.2.1 are unadjusted. The estimates based on clinical measurements are similar to the estimates based on administrative data, while the indicators based on self-reported data appear to underestimate the prevalence.

Table 2.2.2 shows the results of a comparison of hypertension estimates for Quebec in 2009 (Blais, Rochette and Hamel 2013). The survey data are from the CCHS and the CHMS, and the estimates based on administrative data are from the QICDSS. The CCHS estimate is based on self-reporting in response to the question “Have you been diagnosed with high blood pressure by a health professional?” The CHMS estimate is based on a series of blood pressure measurements and on self-reported use of medication for high blood pressure in the past month.

The prevalence figures presented in Table 2.2.2 are age-adjusted. Once again, the estimates based on administrative data are closer to the estimates based on clinical measurements, while the estimates based on self-reported data appear to underestimate the prevalence.

Table 2.2.1

Prevalence of chronic obstructive pulmonary disease in Canada, ages 35 and over, 2005 to 2009

Study	Prevalence (%)
Administrative data	8.3
CCHS self-reported data	4.8
CanCOLD study self-reported data	6.9
CanCOLD study clinical measurements	7.9
CHMS clinical measurements	8.1

Table 2.2.2**Adjusted prevalence of hypertension in Quebec, ages 20 to 79, 2009**

Study	Prevalence (%)
Administrative data	20.6
CCHS self-reported data	18.0
CHMS clinical measurements and use of medication	22.6

3. Data quality: Overview of the advantages and limitations of administrative files**3.1 Linkage quality**

The RAMQ was tasked with merging the five administrative files mentioned in Section 1.3. The RAMQ owns the health insurance registry, the fee-for-service file and the pharmaceutical services file. It is also responsible for issuing health insurance numbers, which are the unique identifiers used as the matching key. All the records in these three files are merged.

The hospitalization file and the deaths file are from other sources and are therefore subject to linkage errors caused by missing or incorrectly coded health insurance numbers. The proportion of linked records is 99% for the hospitalization file and 97% for the deaths file. The majority of the missing hospitalization records relate to newborns for whom there was a transcription error between the temporary identifier assigned at birth and the permanent identifier assigned a few months later. The missing records from the deaths file produce little differential bias, with the exception of violent deaths for which a coroner conducted an inquest. The coroner is not required to record the health insurance number on the death certificate. Consequently, there may be a bias in the evaluation of causes of death, particularly for people diagnosed with mental disorders, as they are more likely to die as a result of accident or injury.

3.2 Strengths and limitations of administrative files

The health insurance registry is the best available source for defining the population cohort. Health insurance is publicly funded and covers almost the entire population. In addition, the estimates of surveillance indicators based on the registry are more precise than the demographic estimates based on the census, since the cases included in the numerator are all included in the denominator.

We nevertheless compared the population estimates based on the registry with those based on the census to make sure that the denominator considered in calculating the indicators was consistent with the population distribution in Quebec. The age–sex population distributions from the two sources are very similar to each other except for young adults, whose estimated population based on the health insurance registry is 5% to 10% lower than the census demographic estimates for the years covered by the QICDSS. A detailed review showed that a high proportion of young adults were late in renewing their health insurance cards or failed to do so. The RAMQ issues health insurance cards valid for four years, and users are responsible for renewing them. The proposed solution was to automatically assign a health insurance card to men aged 18 to 30 and to women aged 18 to 24. There was a risk that this procedure could result in an overestimation caused by individuals who were no longer eligible for health insurance. However, comparisons made following the adjustment showed very little difference between the estimate based on the registry and census data.

The fee-for-service file is an excellent source of information, with more than 1 billion records for the 17-year history of the surveillance system. This file is the main source of data for identifying cases. It contains useful data such as diagnostic codes, procedure codes, type of institution and physician specialty. The main limitation of this file is that the information covers only fee-for-service claims. Physicians in Quebec may also be paid on a fixed-amount, sessional or salary basis. This might constitute a potential bias, since these alternative forms of payment have increased since the inception of the surveillance system. According to the Canadian Institute for Health Information

(ICIS 2014), the proportion of physician remuneration by alternative means rose from 16% in 1999/2000 to 24% in 2011/2012, with a peak of 26% in 2008/2009. Verification of this information over the next few years will be very important in documenting potential bias in the evolution of the number of cases.

The hospitalization file contains validated information on all admissions to hospitals for short-term care. Medical records administrators review admission files and code the summary data provided by physicians. The information in this file is therefore valid, and a single diagnosis is generally sufficient for the identification of a case from this file. The main limitation of this file is that it does not include records for people who are treated in the emergency department and subsequently not admitted. These individuals are identified only if the physician submitted a fee-for-service claim for the medical services provided in the emergency department.

The pharmaceutical services file is also an excellent source of information, since it contains more than 1 billion records for the study period. However, the population in question is limited to seniors aged 65 and over who participate in the public drug insurance plan. The file does not include medications prescribed for seniors covered by private drug insurance plans or for people who are hospitalized or in long-term care. An estimated 10% of seniors are excluded as a result.

The main value of the deaths file is that it documents causes of death. This file is usually not available until two or three years after the other files. Consequently, the health insurance registry is used for data on the date of death. A potential source of bias in the deaths file is the change in the coding method made in January 2000. Before then, only one cause of death coded under the International Classification of Diseases, Ninth Revision, was recorded. Since then, the primary cause and as many as 10 secondary causes coded under the International Classification of Diseases, Tenth Revision, have been recorded. It is therefore important to study the impact of these changes over the study period.

4. Conclusion

The Quebec Integrated Chronic Disease Surveillance System contains an impressive quantity of information; its strengths far outweigh its limitations. The system is inexpensive and produces reliable data for the entire population on a timely basis. It is nevertheless important to understand and document the limitations of the system, because they are generally different from those of sample surveys.

References

- 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. and Émond, V. (2014), "Quebec Integrated Chronic Disease Surveillance System (QICDSS), an innovative approach", *Chronic Diseases and Injuries in Canada*, Vol. 34, no. 4, p. 226-235. Forthcoming.
- Blais, C., Rochette, L. and 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.
- Doucet, M., Rochette, L., Gagné, M.E., 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). Available at: <https://secure.cihi.ca/estore/productFamily.htm?locale=en&pf=PFC2368>.

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

Statistics Canada (2014). Available at: www.statcan.gc.ca.

Tan, W.C., Bourbeau, J., Fitzgerald, J.M., Cowie, R., Chapman, K., Hernandez, P., Buits, S., Sin, D.D. (2011). "Can age and sex explain the variation in COPD rates across large urban cities? A population study in Canada", *International Journal of Tuberculosis and Lung Disease* 15 (12): 1691–1698.