

*This is the peer reviewed version of the following article:* Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , **which has been published in final form at** <https://doi.org/10.1002/pds.4889>. **This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.**

## **Data variability across Canadian administrative health databases: differences in content, coding, and completeness**

**Authors:** Carla M. Doyle<sup>1</sup>, Lisa M. Lix<sup>2</sup>, Brenda R. Hemmelgarn<sup>3,4</sup>, J. Michael Paterson<sup>5,6,7</sup>, Christel Renoux<sup>1,8,9</sup>

### **Author affiliations:**

<sup>1</sup>Centre for Clinical Epidemiology, Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada

<sup>2</sup>Department of Community Health Sciences, College of Medicine, University of Manitoba, Winnipeg, Manitoba, Canada

<sup>3</sup>Department of Medicine, Division of Nephrology, University of Calgary, Alberta, Canada

<sup>4</sup>Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada

<sup>5</sup>University of Toronto, Toronto, Ontario, Canada

<sup>6</sup>Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada

<sup>7</sup>McMaster University, Hamilton, Ontario, Canada

<sup>8</sup>Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada

<sup>9</sup>Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada.

### **Address for Correspondence:**

Christel Renoux, Centre for Clinical Epidemiology, Lady Davis Research Institute, Jewish General Hospital, 3755 Cote Ste-Catherine, Montreal, Quebec H3T 1E2, Canada

Tel.: (514) 340-8222 ext 24561; Fax: (514) 340-7564

Email: [christel.renoux@mcgill.ca](mailto:christel.renoux@mcgill.ca)

**Short title:** Data variability in Canadian health databases

**Word Count:** 3858

**Tables:** 1; **Figures:** 0; **Online Appendices:** 1

**Keywords:** administrative health databases; provincial Canadian database; Research networks; pharmacoepidemiology

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

**Key points:**

- The Canadian Network for Observational Drug Effect Studies uses administrative databases across Canada to conduct drug safety and effectiveness studies.
- The populations and drugs covered by provincial programs, versions and precision of the International Classification of Diseases coding system in physician claims, and content and completeness of hospital discharge abstracts varies across provinces and over time.
- The heterogeneity amongst provincial databases can potentially introduce differences in study cohorts formed and affect study results.
- By illustrating this heterogeneity, we have emphasized the importance of considering the variability between databases in distributed networks and investigating the impact these differences might have on study results.

**Financial sponsors:** This research is funded by the Canadian Network for Observational Drug Effects Studies, a collaborative product of the Canadian Institutes of Health Research and Health Canada (Grant Number DSE-146021).

**Conflict of interest statement:** The authors declare no competing interests.

**Prior posting and presentation:** This work is the sole product of the authors and has never been submitted for publication; it has been presented at the International Conference for Pharmacoepidemiology<sup>1</sup>.

---

<sup>1</sup>Doyle CM, Lix L, Paterson JM, Hemmelgarn B, Renoux C. Data variability across the Canadian administrative health databases: differences in content, coding, and completeness. *Pharmacoepidemiology and Drug Safety*. 2017; 26: 3–636. doi: 10.1002/pds.4275.

*This is the peer reviewed version of the following article: Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , which has been published in final form at <https://doi.org/10.1002/pds.4889>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.*

## ABSTRACT

**Purpose:** The Canadian Network for Observational Drug Effect Studies (CNODES) is a network of Canadian research centres using administrative data to conduct distributed drug safety and effectiveness studies. In this study, we compare the provincial administrative databases and illustrate the potential impact of database differences on a CNODES study about domperidone and the risk of ventricular tachyarrhythmia and sudden cardiac death (VT/SCD).

**Methods:** We assessed the impact of varying versions and precision of the International Classification of Diseases coding system in physician claims data, and the content and completeness of hospital discharge abstracts across CNODES sites, as these variations can introduce differences in the study cohorts formed and affect study results.

**Results:** In our study of 214,962 patients, hospital diagnosis type (such as most responsible, admitting, or secondary diagnosis) was missing in some provinces, resulting in misclassification of the outcome and variation in rates and risk estimates. Incidence rates of VT/SCD ranged from 19.8 (95% confidence interval [CI] 17.7–22.2) per 10,000 person-years in British Columbia to 53.4 (95% CI 50.3–56.5) in Quebec. While most provinces reported an increased risk of VT/SCD, a null effect was observed in Quebec (rate ratio 1.06; 95% CI 0.79-1.41).

**Conclusions:** Distributed analyses allow for rapid responses to drug safety signals. However, variation in characteristics of the administrative data across research centres can influence study results. By

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

- 1 identifying the sources of database heterogeneity, one can evaluate the potential biases these
- 2 differences may introduce, highlighting the importance of considering such variation in distributed
- 3 networks.

4

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

## Introduction

Routinely collected administrative health data, while generally intended for payment purposes, are useful for the conduct of pharmacoepidemiologic research, as they are relatively inexpensive to access and process, are population based, and span multiple decades of healthcare delivery. A recent development in the field of pharmacoepidemiology is the creation of distributed networks of research centres with access to such data. By encompassing databases from multiple jurisdictions, these networks are able to increase the generalizability of study findings and include data on large numbers of individuals, resulting in an increased ability to rapidly respond to drug safety signals;<sup>1-4</sup> the Canadian Network for Observational Drug Effect Studies (CNODES), established in 2011 with funding from Canada's Drug Safety and Effectiveness Network, is one such distributed research network.<sup>1</sup> Others include the FDA Sentinel Initiative<sup>5</sup> and the Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium (PROTECT).<sup>6</sup> Comprised of researchers from across Canada, CNODES includes administrative data from seven Canadian provinces (Alberta, British Columbia, Manitoba, Nova Scotia, Ontario, Quebec, and Saskatchewan), the United Kingdom Clinical Practice Research Database, and the United States MarketScan, in order to conduct studies of drug safety and effectiveness.<sup>1</sup>

*This is the peer reviewed version of the following article:* Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , *which has been published in final form at* <https://doi.org/10.1002/pds.4889>. *This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.*

Health data privacy legislation, which is province-specific, governs access to administrative health data at each CNODES site. To ensure all legal requirements are met, and to take advantage of local expertise, the data used by CNODES remain ‘in situ’ at their respective provincial research centers<sup>1</sup>; as such, CNODES develops common study protocols that are implemented simultaneously at each site. A study’s overall result is then obtained by combining site-specific results using meta-analytic techniques. The common protocols and processes prepared by CNODES (<http://www.cnodes.ca/>) serve to standardize the study design and methodological approaches being used across its sites in order to reduce heterogeneity in the site-specific results. Differences in database structure and contents do, however, exist that can lead to some variation in, for instance, the study population or outcome definition used at each site. In this paper, we highlight some of the variations between the provincial administrative health databases used by CNODES and illustrate the potential impact of database differences on a CNODES study that examined the risk of ventricular tachyarrhythmia and sudden cardiac death (VT/SCD) associated with domperidone use among patients with Parkinson’s disease.<sup>7</sup> In this study, these differences influenced the case definition for VT/SCD and provincial effect size estimates.

## Methods

### Description of Canadian administrative health data sources and access

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

1 Canada has a universal, publicly-funded health insurance program that provides Canadian  
2 citizens with health insurance that is funded by the federal government, but managed by provincial  
3 and territorial governments.

4 As previously detailed by Quan et al., three particular features enrich the Canadian  
5 administrative health data: (1) each province or territory maintains a health insurance registry; (2)  
6 data are captured on all inpatient and outpatient medical services delivered (including information  
7 from physician consultations, emergency department visits, hospitalizations, and surgical procedures);  
8 and (3) each insured person is assigned a unique identifier that allows for the linkage of data within  
9 each province or territory.<sup>8</sup> As such, administrative health data on nearly the entirety of the Canadian  
10 population is captured. In addition, prescription drug dispensation data is collected within each  
11 province and territory and can be linked to the health databases via the unique identifier. Unlike  
12 health care however, a universal drug coverage program does not exist in Canada. As such, public  
13 drug coverage varies significantly according to province or territory and the drug data collected does  
14 not necessarily capture the entire population of each jurisdiction.<sup>9</sup>

15 In addition to a health insurance registry, each province and territory hosts a physician claims  
16 database, a hospital discharge abstract database, a drug claims database, and a vital statistics  
17 database. In some provinces, supplementary registries and databases, such as cancer registries or  
18 vaccine databases, may also be present. Many of these have been described in detail elsewhere.<sup>9-11</sup>

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

1           Although not the main focus of this article, it is important to highlight the differences in  
2           provincial drug claims data (see Table 1), as the study period and population are dependent on these  
3           data. First, the earliest date of provincial drug data availability ranges from 1983 in Quebec to 1997 in  
4           Ontario. Second, the populations covered by provincial public drug programs vary by province. In  
5           British Columbia, Manitoba, and Saskatchewan, drug claims data are available for the entire  
6           population whereas in Nova Scotia and Ontario, drug claims data are only available for those aged 66  
7           and older, as well as those receiving social assistance. In Alberta, drug claims data are available for the  
8           entire province as of 2008; for the years 1994 to 2007, data are only available for Albertans aged 65  
9           and older. In Quebec, drug claims data are available for those aged 66 and older and those receiving  
10          social assistance; also, as of 2001, data are available for those 65 and younger that purchase drug  
11          coverage from the province (i.e. those without a private insurance plan). Lastly, a drug's date of  
12          approval under the public program and its formulary restrictions (criteria for which all or part of the  
13          cost of a drug is covered by the provincial drug plan) typically differ by province, and over time within  
14          provinces.<sup>12</sup>

15          In the experience of CNODES, it has been important to consider four factors -- date of data  
16          availability, population covered, date of drug formulary coverage, and nature of formulary coverage --  
17          across the sites when determining the feasibility of a study, as well as when deciding which sites to  
18          include in a study. For instance, CNODES has included data from British Columbia, Manitoba and

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

1 Saskatchewan in studies of young people, as these provinces capture drug data for their entire  
2 populations. The challenges brought by various formulary restrictions across provinces have been  
3 illustrated in previous CNODES studies.<sup>13, 14</sup> For example, in the study by Filion and colleagues<sup>14</sup>,  
4 restrictions on access to proton pump inhibitors in Nova Scotia led to strong confounding by  
5 indication.

6 Another practical factor to consider is data accessibility. The process for data requests and  
7 wait times for its receipt are highly variable between provinces. Streamlined processes have been  
8 approved by the provincial ministries of health at some sites. For others, this streamlined process is  
9 still under negotiation.

10 The remainder of this paper will describe the heterogeneity amongst the data elements of the  
11 provincial health databases used by CNODES, with focus on the provincial physician claims, hospital  
12 discharge abstracts, and vital statistics databases. We will illustrate the potential impact of database  
13 differences on a CNODES study about domperidone and the risk of VT/SCD.<sup>7</sup>

14 Briefly, this study used the databases from seven Canadian provinces as well as the Clinical  
15 Practice Research Datalink (CPRD) in the UK to assess the risk of VT/SCD associated with the use of  
16 domperidone among patients with Parkinson's disease. Each site followed the same study protocol in  
17 order to conduct separate retrospective cohort studies with a nested case-control approach to

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

analysis; database-specific results were subsequently pooled via meta-analytic methods. The cohort included all patients aged 50 years or older newly diagnosed with Parkinson's disease between January 1, 1990 (or one year after the date of data availability, depending on the province) and June 30, 2012. We excluded patients with a history of VT, aborted cardiac arrest, implantation of a cardiac defibrillator, or cancer other than non-melanoma skin cancer, as well as patients diagnosed with secondary Parkinsonism; or living in a long-term care facility in the year before cohort entry.

The outcome was defined as a composite of VT or SCD. As data on prescriptions issued in hospital or long-term care facilities are generally not available, outcomes of VT/SCD occurring within these institutions were excluded. However, patients who were admitted to hospital for VT or resuscitated from SCD in an emergency department, or who died shortly (<24 h) after their arrival to hospital were included. The administrative record of all potential cases was manually reviewed in each centre to further exclude cases not meeting all inclusion and exclusion criteria. Each cohort was analyzed using a nested case-control approach, and each case of VT/SCD was matched to up to 30 controls on age, sex, date of cohort entry and duration of follow-up using risk set sampling.

Conditional logistic regression was used to estimate the rate ratio (RR) of VT/SCD associated with current use of domperidone. Any adaptation that was made to the study protocol was recorded at each site; a brief summary of these can be found in Table 2.

## **Physician Claims Data**

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

1           The physician claims databases in each province and territory contain information on all  
2   outpatient encounters. As summarized by Lix and colleagues, the earliest dates of data availability are  
3   from the 1970s in Manitoba and Saskatchewan, with data available in all other provinces as of the  
4   1990s<sup>11</sup> (see Table 1). Common to each database is a patient identifier, patient sex and date of birth,  
5   service date, physician identifier and specialty, a diagnosis code, and fee codes that identify the  
6   procedures and services rendered. Differences in these common data elements do, however, exist  
7   and must be considered when using service claims from multiple jurisdictions.

8           One of the more notable differences lies in the coding of diagnoses (Table 1). While all  
9   physician claim databases use diagnosis codes from the World Health Organization’s International  
10   Classification of Diseases (ICD) coding system, the version and precision varies across provinces. For  
11   instance, while both Alberta and Manitoba physician claims data are coded in ICD-9, the level of  
12   precision is five and three digits, respectively. In Ontario, diagnoses on physician claims are coded  
13   using modified ICD-8 codes. Finally, in Saskatchewan, ICD-9 codes are used along with a set of  
14   province-specific medical services diagnostic codes, known as “Z-codes”. Table 1 provides an overview  
15   of the ICD coding schemes used across Canadian physician claim databases, the precision of the  
16   diagnosis codes, and the maximum number of diagnosis codes listed on each claim.

17           These differences in the coding of physician claims data led to differences in the study  
18   populations at each participating site in the CNODES study of domperidone and the risk of VT/SCD.<sup>7</sup>

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

1 For instance, the study cohort included patients newly diagnosed with Parkinson's Disease (ICD-9  
2 code 332.0), but excluded patients with a diagnosis of secondary Parkinsonism (ICD-9 code 332.1). In  
3 Manitoba, Saskatchewan, and Ontario, where three digit diagnosis codes are captured in physician  
4 claims data, it was not possible to apply this exclusion criterion. Patients with a cardiac defibrillator  
5 were also to be excluded. For the same reason, these three provinces could not differentiate amongst  
6 patients who had previously received a cardiac pacemaker (ICD-9 code V45.01) or patients who had  
7 previously received a cardiac defibrillator (ICD-9 code V45.02). As such, they were unable to exclude  
8 patients with a cardiac defibrillator. Although the remainder of sites had ICD codes with up to five  
9 digits, some sites such as British Columbia reported that codes commonly had only three digits,  
10 indicating that these same exclusion criteria may not have fully applied at all sites. The consequences  
11 of the various applications of the inclusion and exclusion criteria are unlikely to have influenced the  
12 results in this particular study, mainly because of the small numbers of patients affected by these  
13 exclusion diagnoses (less than 0.1% of patients were excluded on these criteria in provinces where  
14 these diagnostic codes were available). However, this may not hold true in situations where failure to  
15 apply identical exclusion criteria across sites results in important differences in study populations.  
16 Similarly, inter-site differences in the precision of ICD codes could affect study outcomes.

## 17 **Hospital Discharge Abstract Data**

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

1           The Discharge Abstract Database (DAD) is a national database established by the Canadian  
2   Institute for Health Information (CIHI) to compile data from acute care institutions across the country.  
3   Submission of information to the DAD is mandatory in all provinces, except Quebec, with each  
4   participating province completing a DAD abstract for each hospital separation: discharge, transfer, or  
5   death. In some provinces, data on day surgeries also are recorded in the DAD, although provinces are  
6   increasingly capturing these data in the National Ambulatory Care Reporting System (NACRS).<sup>15</sup> In  
7   Quebec, hospital discharge data are captured in the MED-ÉCHO database. While these data are not  
8   captured in the DAD or NACRS, it is submitted to CIHI annually and stored together with DAD records  
9   to make up the Hospital Morbidity Database (HMDB).<sup>16</sup> The mandatory elements vary by province  
10   and, as reported by CIHI, the response rate for non-mandatory fields is typically low.<sup>15</sup> For instance, in  
11   Manitoba and British Columbia, recording the patient's residence code is optional,<sup>17</sup> making linkage to  
12   the provincial health insurance registry critical. Several data elements are captured in hospital  
13   discharge abstracts in all provinces, including Quebec: patient identifier, sex and date of birth, date of  
14   hospital admission and discharge, procedures, and diagnoses. However, the format and coding of  
15   certain data elements, such as diagnoses and procedures, vary across provinces. The diagnosis coding  
16   system, number of digits per code and maximum number of diagnosis codes on each hospital  
17   discharge abstract in each province are described in Table 1.

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

1       The earliest hospital discharge abstract data used in CNODES studies contain diagnoses coded  
2       using ICD-9 or ICD-9-CM;<sup>18</sup> the latter is a modified version of ICD-9 for use in clinical settings. Between  
3       2001 and 2002, most provinces switched to ICD-10-CA,<sup>19</sup> a Canadian version of ICD-10. The exceptions  
4       were Manitoba and Quebec, where this change occurred in 2004 and 2006, respectively. Notably, in  
5       Saskatchewan the change in ICD system did not happen uniformly; instead, part of the province  
6       switched from ICD-9 to ICD-10-CA in 2001, with the rest following in 2002.<sup>20</sup> In all provinces except  
7       Saskatchewan, the diagnosis codes in the hospital discharge abstract data contain 5 or 6 digits; in  
8       Saskatchewan these codes are reported with 4 (up to March 31, 2001/2002) and 7 digits (as of April 1,  
9       2001/2002).

10       All diagnosis codes listed on the DAD abstract are classified according to type. Multiple types  
11       of diagnoses are recorded in the DAD abstract, including most responsible diagnosis, pre-admit  
12       comorbidity, post-admit comorbidity, secondary diagnoses, and admitting diagnosis (see Appendix  
13       1).<sup>21, 22</sup> In Quebec discharge abstracts, comorbidities are defined as admission, principal or  
14       secondary.<sup>23</sup> Between 1996 and 2005 the reporting of an admission diagnosis was mandatory, but  
15       became optional in 2006.<sup>24</sup> Accordingly, there is incomplete capture of admission diagnoses in  
16       Quebec. Currently, the only diagnosis type that is mandatory in Quebec is the principal diagnosis.

17       The omission of diagnosis type in Quebec's discharge abstract data was problematic in the  
18       CNODES domperidone study.<sup>7</sup> Cases were to have experienced the outcome of VT/ SCD either out-of-

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

1 hospital or upon admission to hospital. In Quebec, however, few of the VT and SCD cases identified  
2 over the study period had any additional diagnostic information recorded beyond the principal  
3 diagnosis. Therefore, for a substantial number of ‘potential’ cases such as VT or resuscitated cardiac  
4 arrest, it was unclear whether the event had occurred before or after admission to hospital. This  
5 phenomenon was also observed when reviewing potential cases in the Alberta database, where  
6 diagnosis type is a mandatory data element; however, in actuality, the only type identified was the  
7 most responsible diagnosis. In contrast, the type and date of diagnosis in the British Columbia or  
8 Ontario DAD abstracts, for instance, were more complete facilitating the exclusion of in-hospital  
9 events. Although the study investigators conducted chart reviews of all ‘potential’ cases in all  
10 provinces, in Quebec and Alberta for instance, the missing information on diagnosis type and date in  
11 the discharge abstracts precluded the exclusion of many VT/SCD events that may have occurred in  
12 hospital, which would have led to outcome misclassification and inflated outcome event rates in these  
13 provinces. The VT/SCD incidence rate was 198.0 (95% CI 177.0-222.0) per 100,000 person-years in  
14 British Columbia and 278.7 (95%CI 259.5-297.9) per 100,000 person-years in Ontario, while in Alberta  
15 and Quebec, the rates were 320.3 (95% CI 280.8-365.4) and 534.2 (95% CI 503.2-565.2) per 100,000  
16 person-years, respectively (Table 3). The largest two study cohorts, from Quebec and Ontario, had  
17 very similar age and sex structures, as well as similar percentage of patients exposed to domperidone  
18 during follow-up. However, the event rate in Quebec was nearly twice the rate in Ontario. Also, while

*This is the peer reviewed version of the following article: Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , which has been published in final form at <https://doi.org/10.1002/pds.4889>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.*

the outcome event rates in other provinces were more consistent and indicative of an increased risk of VT/SCD with current use of domperidone, we observed a null effect in Quebec (RR 1.06; 95% CI 0.79-1.41) (Table 3). Consequently, in the pooled analysis, the effect of current use of domperidone was diluted and the point estimate lowered (RR 1.22; 95% CI 0.99-1.50), because Quebec represented more than 50% of the weight in the meta-analysis. Excluding the province of Quebec from the meta-analysis would have yielded a 42% increased risk that was statistically significant (RR 1.42; 95% CI 1.05-1.90).

Another inconsistency between provinces was the frequency of codes used to initially identify potential cases of VT/SCD before individual chart review. Indeed, we found that the frequency of ‘unattended death’-- the least specific code to identify ‘potential’ cases -- was much higher in Quebec than in other provinces. This had a minor impact on the results in the domperidone study compared to the lack of information on admission diagnosis. However, these differences in coding and billing practices between provinces have not been studied and are not well understood, so their impact may vary across studies.

### **Vital Statistics Data**

By law, information on all deaths occurring in each province and territory is recorded and stored in the respective vital statistics databases of each jurisdiction. The information collected includes the demographic details of each decedent, as well as the presumed cause of death. While

*This is the peer reviewed version of the following article:* Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , **which has been published in final form at <https://doi.org/10.1002/pds.4889>**. **This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.**

some provinces and territories may collect more information than others, a minimum set of vital statistics data are shared by all with Statistics Canada to form a national vital statistics database. The data elements common to all provinces include age, sex, marital status, place of birth and residence, province and date of death occurrence, and primary cause of death. While only the primary cause of death is reported to Statistics Canada, multiple secondary causes of death may be recorded and stored in the provincial databases (see Table 1).<sup>25</sup>

In most provinces the primary and secondary causes of death are made available for research purposes. In Alberta, however, the vital statistics data released for research purposes normally includes only the primary cause of death. Accordingly, identification of non-arrhythmic cardiac or non-cardiac cause of death codes in the Alberta data was hampered in the CNODES domperidone study.<sup>7</sup> Indeed, VT/SCD events due to non-arrhythmic cardiac causes or acute life-threatening non-cardiac causes were not considered as cases in the study, and 'potential' cases with these cause of death codes listed on their death certificate were to be excluded during the individual chart review; in Alberta this exclusion was limited since only one cause of death was listed. Subsequently, CNODES has successfully requested information on secondary causes of death in Alberta.

## Discussion

Distributed networks of administrative health data centres incorporate data from multiple jurisdictions to increase the sample size, power, and generalizability of pharmacoepidemiologic

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

1 studies. CNODES is one such distributed network accessing administrative databases across Canada to  
2 conduct drug safety and effectiveness studies. The common protocols, statistical analysis plans, and  
3 analytic code established by CNODES aim to reduce heterogeneity in the results obtained from each  
4 site. Common exposure, covariate and outcome definitions are used for all sites and we make  
5 maximum use of shared common, tested, statistical analysis code.<sup>1</sup> All study outcomes and analyses  
6 are pre-specified, and site-specific protocol deviations are permitted for technical reasons only, as  
7 was the case in this study. All steps to conduct of the project are systematically enumerated to ensure  
8 that the protocol can be consistently implemented in all participating sites. However, database  
9 differences between Canadian provinces can contribute to variation in study results across sites. Here  
10 we compared the provincial administrative databases and illustrated the potential impact of  
11 differences among provinces on a CNODES study about domperidone and the risk of VT/SCD.<sup>7</sup> We  
12 showed, in particular, how differences in available information from the provinces lead to large  
13 variability in case definitions, event rates, and effect size estimates across some provinces.

14 While databases across Canadian provinces share many similarities, we described several  
15 variations in the information available across sites. First, the populations and drugs covered by  
16 provincial programs differ across provinces and over time, which can potentially lead to inter-  
17 provincial differences in study populations and accrual periods. Further, the version and precision of  
18 the ICD coding system used in physician claims data varies, and this can introduce differences in the

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

1 study cohorts. Variation in the content and completeness of hospital discharge abstracts across  
2 provinces can similarly affect the study results obtained. We showed that this source of variation led  
3 to different outcome rates and ultimately variation in point estimates of the association between  
4 exposure to domperidone and VT/SCD across provinces. Finally, while vital statistics data collected  
5 across provinces are similar, accessibility of these data for research purposes can vary between  
6 provinces.

7 Few studies have provided specific examples of the challenges posed by differences in  
8 databases on the conduct and outcomes of distributed analyses. Some have investigated the  
9 feasibility of combining data from European electronic health record databases for drug safety  
10 monitoring.<sup>26</sup> These authors highlighted the differences in the databases used and concluded that the  
11 heterogeneity observed arose from variation in language, coding systems, healthcare systems, as well  
12 as possible differences in risk factors and diagnostic testing across countries. Another study examined  
13 the heterogeneity in the meta-analyzed risk estimates of 45 drug-outcome associations across 8 US  
14 claims and electronic medical records databases.<sup>27</sup> The authors quantified the heterogeneity using  
15 the I-squared statistic and graphically compared a weighted measure of each data source's influence  
16 on results to a weighted measure of their contribution to overall heterogeneity. In doing so, a  
17 relatively high level of heterogeneity was found across data sources. Due to its complexity, no single

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

method for handling heterogeneity was recommended; however, the authors illustrated how such investigation can help to identify and understand data source heterogeneity.

The data variability that we have described in the context of the Canadian provincial databases may be relevant to other settings, as similar differences may exist elsewhere. As shown in our example, administrative health data collected separately in different regions within the same country can differ in content and quality. Such data heterogeneity and how it might influence study results should be considered when assessing study feasibility. The detail and precision of the information available to carry out such a database study should be carefully reviewed within each individual site to ensure it can acceptably satisfy the necessary case definitions and procedures (i.e. definitions of the study population, exposures, outcomes, covariates, etc.). Upon such careful assessment, researchers can make an informed decision as to whether the study is feasible at each potential participating site, or whether some should abstain from participating in the study before it is carried out. While not an issue in the study presented here<sup>7</sup>, drug indications and formulary restrictions should also be considered for each site because differences could affect the ability to effectively address confounding by indication across sites.<sup>14</sup> Finally, identifying sources of data variability and evaluating their impact on study results represents an important avenue for future research.

## **Conclusions**

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

- 1           A distributed network can allow for rapid and precise responses to drug safety signals.
- 2           However, variability in the characteristics of site-specific administrative databases can influence study
- 3           results. Identifying sources of database heterogeneity and testing their impact on study findings
- 4           through empirical and simulation studies can strengthen the design and analysis of network studies.
- 5           By including and utilizing the expertise of a large team of researchers, CNODES incorporates such
- 6           methods investigations into its research activities to ensure its assessments of drug safety and
- 7           effectiveness are conducted using high-quality methods that are most appropriate for its databases.

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

**Table 1 – Summary of Canadian administrative health data by province**

Province	Prescription Drug Data		Physician Claims Data			Hospital Discharge Abstract Data		Emergency Department Data	Vital Statistics Data	
	Time Period Covered	Population Covered	Time Period Covered	Diagnosis Code Type; # of Digits	Max. # of Diagnosis Codes per Claim	Diagnosis Code Type; # of Digits	Max. # of Diagnosis Codes per Claim		COD Code Type; # of Digits	# of Codes
Alberta	From 1994	65+, from 1994; All, from 2008	From 1994	ICD-9; 5	3	ICD-9-CM (1994 – March 31, 2002), ICD-10-CA (April 1, 2002 – and onward); 5 (1994 – March 31, 2002), 6 (April 1, 2002 and onward)	16 (1994 – March 31, 2002), 25 (April 1, 2002 and onward)	Yes	ICD-9 (1994 – December 31, 1999), ICD-10 (January 1, 2000 and onward); 5	1*
British Columbia	From 1996	All	From 1986	ICD-9; 5	3	ICD-9 (up to March 31, 2001), ICD-10-CA (April 1, 2001 and onward); 5	16 (up to March 31, 2001), 25 (April 1, 2001 and onward)	No	ICD-9 (up to December 31, 1999), ICD-10 (January 1, 2000 and onward); 5	1 underlying COD; 14 record axis codes†
Manitoba	From 1995	All	From 1973	ICD-9-CM; 3	1	ICD-9 (1979 – March 31, 2004), ICD-10-CA (April 1, 2004 and onward); 5 (1979 – March 31, 2004), 6 (April 1, 2004 and onward)	16 (1979 – March 31, 2004), 25 (April 1, 2004 and onward)	Partial (emergency departments in the Winnipeg health region)	ICD-9 (up to December 31, 1999), ICD-10 (January 1, 2000 and onward); 4	Primary, 20 underlying CODs
Nova Scotia	From 1989	66+ and persons receiving social	From 1989	ICD-9-CM; 4 (1989 – 1996), 5 (1997 and onward)	1 (1989 – 1996), 3 (1997 – forward)	ICD-9-CM (1989 – March 31, 2001), ICD-10-CA (April 1, 2001 and onward); 5 (1989 –	7 (1989 - 1994), 16 (1995 - 2000), 25 (2001 and onward)	Yes	ICD-9-CM (up to December 31, 1999), ICD-10-CM	1 underlying COD, 13 CODs

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

		assistance/ Family Pharmacare				March 31, 2001), 6 (April 1, 2001 and onward)			(January 1, 2000 and onward); 5 (up to December 31, 1999), 4 (January 1, 2000 and onward)	
Saskatchewan	From 1990	All <sup>†</sup>	From 1996	ICD-9 and Z-codes <sup>§</sup> ; 3	1	ICD-9 (up to March 31 2001/2002), ICD-10-CA (April 1, 2001/2002 and onward (half of the province switched to ICD-10-CA by April 1, 2001; the other half by April 1, 2002); 4 (up to March 31, 2001/2002), 7 (April 1, 2001/2002 and onward)	3 (1996/1997-1998/1999), 16 (1999/2000-2000/2001), 25 (2001/2002 and onward)	Partial (Tertiary Care Sites)	ICD-9/ICD-10 (1996 and onward); 4 and 7	Primary, 20 underlying CODs
Ontario	From 1997	66+ and persons receiving social assistance	From 1991	Modified ICD-8; 3	1	ICD-9 (up to March 31, 2002), ICD-10-CA (April 1, 2002 and onward); 5	16 (up to March 2002), 25 (April 1, 2002 and onward)	Yes	ICD-9 (up to December 31, 1999), ICD-10 (January 1, 2000 and onward); 4 (up to 1999), 5 (2000 and onward);	Immediate COD, 11 antecedent CODs

**This is the peer reviewed version of the following article:** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , **which has been published in final form at** <https://doi.org/10.1002/pds.4889>. **This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.**

Quebec	From 1983	66+ and persons receiving social assistance, from 1983; 66+, persons receiving social assistance, persons without a private insurance plan, from 2001	From 1983	ICD-9; 4	1	ICD-9 (up to March 2006), ICD-10-CA (April 1, 2006 and onward); 5	No maximum	Yes	ICD-9/ICD-10 (1997 and onward); 4	Primary COD, 25 secondary CODs
--------	-----------	---	-----------	----------	---	---	------------	-----	-----------------------------------	--------------------------------

Abbreviations: International Classification of Diseases (ICD); Cause of death (COD)

\* In addition to the underlying COD code, there are up to 15 additional ICD codes available if special permission is obtained to use this data (though for the CNODES domperidone study, only one code was available).

† Record axis codes represent any immediate, antecedent, contributing or life factor causes associated with death. The record axis codes are received together as one code of 100 digits and translated using a length of 5 digits to differentiate each ICD code. The first code is usually the same as the underlying COD.

‡ As per the *Prescription Drugs Act*, pharmacies transmit all prescription drug dispensation data to the Saskatchewan (SK) Drug Plan for residents registered with SK Health (i.e. who have a SK Health Services number), including data on prescription drugs that are not benefits under the SK formulary. Adjudicated drug data (drug claims reviewed and reimbursed by the SK Ministry of Health) are available as of April 1, 1990; Non-adjudicated drug data (private drug insurance claims, out of pocket payments, or drug claims paid for Registered Indians, veterans, and inmates of federal prisons by federal government programs) is available as of April 1, 2008.

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

<sup>§</sup>Z-codes: SK medical services diagnostic codes

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

**Table 2 – Summary of site-specific protocol adaptations as applied in the CNODES study of domperidone and the risk of VT/SCD<sup>7</sup>**

Province	Study period	Study cohort	Exposure identification	Covariate identification	Outcome identification
Alberta	N/A	<ul style="list-style-type: none"> <li>Patients were aged 66 years and older</li> <li>Could not identify patients in LTC for exclusion. Instead, cases were excluded and selected controls were ineligible if the location of all physician claims in the three months prior to the event/index date was specified as a LTC functional centre code.</li> <li>Could not identify the date of first loss of continuous health or drug plan enrollment. Instead, cases were excluded and controls were considered ineligible if there were no drug dispensations in the three months prior to the index/event date.</li> </ul>	N/A	N/A	<ul style="list-style-type: none"> <li>Could not exclude all outcomes that might have occurred in-hospital due to omission of diagnosis type in Alberta's hospital discharge abstract data.</li> <li>Could not identify outcomes from vital statistics data after December, 2010.</li> <li>Could not exclude potential outcomes due to non-arrhythmic cardiac causes or acute life threatening non-cardiac causes as only one COD code was available in vital statistics data.</li> </ul>
British Columbia	N/A	N/A	N/A	N/A	N/A
Manitoba	January 1, 1996 - June 30, 2012	All hospital diagnoses for inclusion/exclusion criteria were based on exact codes; all physician claims diagnoses used for inclusion/exclusion criteria were based on 3-digit ICD-9 codes.	N/A	All hospital diagnoses for covariates were based on exact codes; all physician claims diagnoses used for covariates were based on 3-digit ICD-9 codes.	<ul style="list-style-type: none"> <li>Could not identify outcomes from emergency department data.</li> <li>All hospital diagnoses and vital statistics for outcome identification were based on exact codes; all physician claims diagnoses used for outcome identification were based on 3-digit ICD-9 codes.</li> </ul>
Nova Scotia	N/A	<ul style="list-style-type: none"> <li>Patients were aged 66 years and older</li> <li>Could not identify patients in LTC for exclusion.</li> </ul>	N/A	N/A	N/A
Quebec	January 1, 1997 -	Patients were aged 66 years and older,	N/A	N/A	Could not exclude all outcomes that

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

	December 31, 2011	unless receiving social assistance, or without a private insurance plan (after 2001)			might have occurred in-hospital due to omission of diagnosis type in Quebec's hospital discharge abstract data.
Saskatchewan	N/A	All hospital diagnoses for inclusion/exclusion criteria were based on exact codes; all physician claims diagnoses for inclusion/exclusion criteria were based on 3-digit ICD-9 codes.	N/A	All hospital diagnoses for covariates were based on exact codes; all physician claims diagnoses used for covariates were based on 3-digit ICD-9 codes	All hospital diagnoses for outcome identification were based on exact codes; all physician claims diagnoses used for outcome identification were based on 3-digit ICD-9 codes.
Ontario	N/A	<ul style="list-style-type: none"> <li>• Patients were aged 66 years and older</li> <li>• All hospital diagnoses for inclusion/exclusion criteria were based on exact codes; all physician claims diagnoses used for inclusion/exclusion criteria were based on 3-digit modified ICD-8 codes.</li> </ul>	N/A	All hospital diagnoses used for covariates were based on exact codes; all physician claims diagnoses used for covariates were based on 3-digit modified ICD-8 codes.	All hospital diagnoses for outcome identification were based on exact codes; all physician claims diagnoses used for outcome identification were based on 3-digit modified ICD-8 codes.

Abbreviations: Not applicable (N/A; this indicates the site had no protocol adaptations to report in that section); Long-term care (LTC); International Classification of Diseases (ICD), Cause of Death (COD)

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

**Table 3 – Site-specific incidence rates and incidence rate ratios obtained in the CNODES study of domperidone and the risk of VT/SCD<sup>7</sup>**

Site	Incidence rate of VT/SCD (95% CI) per 100,000 person-years	Incidence rate ratio of VT/SCD (95% CI)*
Alberta	320.3 (280.8 – 365.4)	1.11 (0.38 – 3.27)
British Columbia	198.0 (177.0 – 222.0)	1.39 (0.59 – 3.28)
Manitoba	215.7 (180.9 – 250.6)	1.28 (0.44 – 3.69)
Nova Scotia	901.0 (727.7 – 1074.2)	1.14 (0.39 – 3.34)
Ontario	278.7 (259.5 – 297.9)	1.28 (0.84 – 1.95)
Quebec	534.2 (503.2 – 565.2)	1.06 (0.79 – 1.41)
Saskatchewan	191.7 (149.2 - 246.3)	2.96 (0.89 – 9.92)

Abbreviations: Ventricular tachyarrhythmia and sudden cardiac death (VT/SCD); Confidence interval (CI); United Kingdom Clinical Practice Research Datalink (CPRD)

\*Comparing current use of domperidone to non-use, matching on sex, age, duration of follow-up, and calendar time. Models are adjusted for comorbid conditions (diabetes, hyperlipidemia, hypertension, cardiovascular disease, peripheral vascular disease, cerebrovascular disease, chronic obstructive pulmonary disease, renal failure, liver disease, epilepsy, schizophrenia, cancer), health utilization, and use of antihypertensive medications, anticoagulants, antiplatelets, antiparkinsonian drugs, drugs that could prolong the QT interval, and strong inhibitors of CYP3A4).

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

## **Acknowledgements**

This study was made possible by The Canadian Network for Observational Drug Effect Studies (CNODES), a collaborating centre of the Drug Safety and Effectiveness Network (DSEN) that is funded by the Canadian Institutes of Health Research (Grant Number DSE-146021). CNODES research is made possible through data sharing agreements between CNODES member research centres and the respective provincial governments of Alberta, British Columbia, Manitoba, Nova Scotia, Ontario, Quebec, and Saskatchewan. The opinions, results, and conclusions reported in this paper are those of the authors. No endorsement by the provincial governments is intended or should be inferred.

**This is the peer reviewed version of the following article:** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , **which has been published in final form at <https://doi.org/10.1002/pds.4889>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.**

## References

1. Suissa S, Henry D, Caetano P, et al. CNODES: the Canadian Network for Observational Drug Effect Studies. *Open Med.* 2012;6(4):e134-40.
2. Popovic JR. Distributed data networks: a blueprint for Big Data sharing and healthcare analytics. *Annals of the New York Academy of Sciences.* 2016.
3. Popovic JR. Distributed data networks: a paradigm shift in data sharing and healthcare analytics. Proceedings of the 2015 Pharmaceutical Industry SAS Users Group Conference; Orlando, FL. 2015.
4. Brown JS, Holmes JH, Shah K, et al. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. *Medical care.* 2010;48(6 Suppl):S45-51.
5. Behrman RE, Benner JS, Brown JS, et al. Developing the Sentinel System--a national resource for evidence development. *The New England journal of medicine.* 2011;364(6):498-9.
6. Reynolds RF, Kurz X, de Groot MC, et al. The IMI PROTECT project: purpose, organizational structure, and procedures. *Pharmacoepidemiology and drug safety.* 2016;25 Suppl 1:5-10.
7. Renoux C, Dell'Aniello S, Khairy P, et al. Ventricular tachyarrhythmia and sudden cardiac death with domperidone use in Parkinson's disease. *Br J Clin Pharmacol.* 2016;82(2):461-72.
8. Quan H, Smith M, Bartlett-Esquilant G, et al. Mining administrative health databases to advance medical science: geographical considerations and untapped potential in Canada. *Can J Cardiol.* 2012;28(2):152-4.
9. Moride Y, Metge CJ. Canadian provincial databases. *Pharmacoepidemiology*: Wiley-Blackwell, Oxford, UK; 2012. p. 259-69.
10. Moride Y, Metge CJ. Data sources to support research on real world drug safety & effectiveness in Canada: an environmental scan & evaluation of existing data elements. In: *Drug Safety and Effectiveness Network*, editor.: Health Canada; 2010. p. 1-179.
11. Lix LM, Walker R, Quan H, et al. Features of physician services databases in Canada. *Chronic Dis Inj Can.* 2012;32(4):186-93.
12. Wong T, Jepson G. Canada's Provincial Drug Formulary System. *Regulatory Focus.* 2011 November.
13. Renoux C, Lix LM, Patenaude V, et al. Serotonin-Norepinephrine Reuptake Inhibitors and the Risk of AKI: A Cohort Study of Eight Administrative Databases and Meta-Analysis. *Clinical journal of the American Society of Nephrology : CJASN.* 2015;10(10):1716-22.
14. Filion KB, Chateau D, Targownik LE, et al. Proton pump inhibitors and the risk of hospitalisation for community-acquired pneumonia: replicated cohort studies with meta-analysis. *Gut.* 2014;63(4):552-8.
15. Canadian Institute for Health Information. Data Quality Documentation, Discharge Abstract Database - Multi-Year Information 2012 [1-14]. Available from: [https://www.cihi.ca/en/dad\\_multi-year\\_en.pdf](https://www.cihi.ca/en/dad_multi-year_en.pdf).

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at*** <https://doi.org/10.1002/pds.4889>. ***This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

16. Canadian Institute for Health Information. Data Quality Documentation, Discharge Abstract Database - Current-Year Information, 2015-2016 2016 [1-18]. Available from: [https://www.cihi.ca/sites/default/files/document/dad-data-quality\\_15-16\\_en.pdf](https://www.cihi.ca/sites/default/files/document/dad-data-quality_15-16_en.pdf).
17. Canadian Institute for Health Information. DAD Data Elements 2016-2017 2016 [1-25]. Available from: [https://www.cihi.ca/en/dad\\_data\\_elements\\_2016\\_2017\\_en.pdf](https://www.cihi.ca/en/dad_data_elements_2016_2017_en.pdf).
18. Centers for Disease Control and Prevention. International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) 2015 [updated June 18, 2013; cited 2017 July 3]. Available from: <https://www.cdc.gov/nchs/icd/icd9cm.htm>.
19. Canadian Institute of Health Information. ICD-10-CA [www.cihi.ca](http://www.cihi.ca): CIHI 1996 - 2016; [cited 2016 December 2]. Available from: <https://www.cihi.ca/en/data-and-standards/standards/classification-and-coding/icd-10-ca>.
20. Canadian Institute for Health Information. ICD-10-CA / CCI Implementation Schedule: CIHI 1996 - 2016; [cited 2016 December 2]. Available from: <https://www.cihi.ca/en/data-and-standards/standards/classification-and-coding/icd-10-ca-cci-implementation-schedule>.
21. Canadian Institute for Health Information. Canadian Coding Standards for ICD-10-CA and CCI for 2006. 2006. p. 1-318.
22. Canadian Institute for Health Information. Canadian Coding Standards for Version 2015 ICD-10-CA and CCI 2015 [1-511]. Available from: [https://secure.cihi.ca/free\\_products/Coding%20standard\\_EN\\_web.pdf](https://secure.cihi.ca/free_products/Coding%20standard_EN_web.pdf).
23. Régie de l'assurance maladie du Québec. Banque de données ministérielles MED-ÉCHO: Gouvernement du Québec; [cited 2016 December 2]. Available from: <http://www.ramq.gouv.qc.ca/fr/donnees-statistiques/sur-demande/donnees-msss/Pages/med-echo.aspx#haut>.
24. Québec: Ministère de la santé et des services sociaux. Cadre normatif du système Med-Écho 2000 [updated April 2006; cited 2017 July 3]. Available from: <http://collections.banq.qc.ca/ark:/52327/bs31498>.
25. Statistics Canada. Vital Statistics - Death Database: Stastics Canada; 2015 [updated December 9; cited 2016 December 12]. Available from: <http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3233>.
26. Coloma PM, Schuemie MJ, Trifirò G, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiology and drug safety.* 2011;20(1):1-11.
27. Hansen RA, Zeng P, Ryan P, et al. Exploration of heterogeneity in distributed research network drug safety analyses. *Research synthesis methods.* 2014;5(4):352-70.

***This is the peer reviewed version of the following article:*** Doyle, CM, Lix, LM, Hemmelgarn, BR, Paterson, JM, Renoux, C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf.* 2019; 29( S1): 68– 77. , ***which has been published in final form at <https://doi.org/10.1002/pds.4889>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.***

## **Appendix 1 – CIHI Discharge Abstract Database: Diagnosis Type Definitions**

The following diagnosis type definitions are presented as provided by CIHI in the Canadian Coding Standards for Version 2015 ICD-10-CA and CCI.<sup>22</sup>

**Most Responsible Diagnosis:** the one diagnosis or condition that can be described as being most responsible for the patient’s stay in hospital. If there is more than one such condition, the one held most responsible for the greatest portion of the length of stay or greatest use of resources (for example, operating room time, investigative technology) is selected.

- If no interventions were performed, select the first-listed diagnosis as the most responsible diagnosis.
- If no definite diagnosis was made, select the main symptom, abnormal finding or problem as the most responsible diagnosis.

**Pre-Admit Comorbidity:** a condition that existed prior to admission, has been assigned an ICD-10-CA code and has been determined to meet at least one of the three *criteria for significance*\*.

**Post-Admit Comorbidity:** a condition that arose post-admission, has been assigned an ICD-10-CA code and has been determined to meet at least one of the three *criteria for significance*\*.

**Secondary Diagnosis:** a secondary diagnosis or condition for which a patient may or may not have received treatment, that has been assigned an ICD-10-CA code and that does not meet any of the three *criteria for significance*\*.

- also used for ICD-10-CA codes that are assigned to provide detail but that in themselves do not represent a condition.

**Admitting Diagnosis:** the admitting diagnosis when it differs from the most responsible diagnosis code.

- Its use is determined at the jurisdictional or facility level.

*\*The condition:*

1. *Requires treatment beyond maintenance of the pre-existing condition;*
2. *Increases the length of stay (LOS) by at least 24 hours; and/or*
3. *Significantly affects the treatment received.*