|  |  |
| --- | --- |
| **eTable 1. Description of health administrative databases.** | |
| **Database** | **Description** |
| ***Health Services*** | |
| Discharge Abstract Database (DAD) | The DAD is compiled by the Canadian Institute for Health Information (CIHI) and contains administrative, clinical (diagnoses and procedures/interventions), demographic, and administrative information for all admissions to acute care hospitals in Ontario.  Prior to April 1, 2002, diagnoses (up to 16 on a given DAD record) are captured using the International Statistical Classification of Diseases, Injuries, and Causes of Death, 9th Revision (ICD-9) coding system and procedures (up to 10 on a given DAD record) are captured using the Canadian Classification of Diagnostic, Therapeutic, and Surgical Procedures (CCP) coding system. Following April 1, 2002, diagnoses (up to 25 on a given DAD record) are captured using the International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Canada (ICD-10-CA) coding system and interventions (up to 20 on a given DAD record) are captured using the Canadian Classification of Health Interventions (CCI) coding system.  In a hospital medical record reabstraction study of 14,500 hospital discharges from 18 hospital sites between April 2002 and March 2004, DAD records were demonstrated to have excellent agreement (over 99%) for nonmedical information such as demographic and administrative data. Regarding diagnoses, median agreement between the original DAD records and the reabstracted records for the 50 most common most responsible diagnoses was 81% (Sensitivity 82%; Specificity 82%).1 The corresponding median agreement for the 50 most frequently performed surgical procedures was 92% (sensitivity 95%, positive predictive value 91%). |
| National Ambulatory Care Reporting System (NACRS) | The NACRS is compiled by the Canadian Institute for Health Information (CIHI) and contains administrative, clinical (diagnoses and procedures), demographic, and administrative information for all patient visits made to hospital- and community-based ambulatory care centres (emergency departments, day surgery units, hemodialysis units, and cancer care clinics) in Ontario.  Prior to April 1, 2002, diagnoses (up to 6 on a given NACRS record) are captured using the ICD-9 coding system and procedures (up to 10 on a given NACRS record) are captured using the CCP coding system. Following April 1, 2002, diagnoses (up to 10 on a given NACRS record) are captured using the ICD-10-CA coding system and interventions (up to 10 on a given NACRS record) are captured using the CCI coding system. NACRS emergency department diagnosis codes have been extensively validated. |
| Canadian Cancer Registry (CCR) | The Canadian Cancer Registry (CCR) is a population based registry that includes data collected and reported to Statistics Canada by each provincial/territorial cancer registry. The person based CCR collects information about each new primary cancer diagnosed among Canadian residents since 1992.2 |
| ***Population and demographics*** |  |
| 2006 Canadian Census Health and Environment Cohort (CanCHEC) | The Canadian Census Health & Environment Cohort (CanCHEC) are census long-form population linked datasets that that follow the non-institutional (i.e. household) population at time of census for different health outcomes such a mortality, cancer, and hospitalizations as well as annual place of residence. These data can be used to examine health inequalities by different population groups and by socioeconomic characteristics. With the inclusion of residential history (estimated via mailing address postal codes reported on tax returns), environmental data can be attached to these data in order to examine the association between environmental exposures and health outcomes. |
| 2006 Census of Population – long form | The 2006 Census of Population - conducted on May 16, 2006 – included both a short- and long-form questionnaire. Approximately, 80% of households received the short-form questionnaire which included information on date of birth, sex, and marital status of all household members. The remaining 20% of households received both the short- and long-form questionnaire. The long-form included questions related to education, ethnicity, mobility, income, and employment.  Only people who were counted by the census could be part of the cohort. Data quality reports found that the 2006 Census missed 4.3% of the Canadian population of all ages, an estimated 1,384,372. The missed individuals were more likely to be young, mobile, low income, of Aboriginal ancestry or homeless. There were also 22 incompletely enumerated Indian Reserves.  The institutional population at time of census was not eligible for the study cohort. To be considered an institutional resident, a person living in an institution on census day must have had no other residence in Canada, or have been living at the institution for at least six months. Thus, people experiencing short-term episodes of hospitalisation or incarceration were not considered institutional residents and were eligible to be part of the study cohort if their household received a long-form census questionnaire. |
| Ontario Marginalization Index (ON-Marg) | The ON-Marg was developed using census and geography data to quantify differences in population marginalization between regions within Ontario.3,4 The factor scores or quintile distributions obtained from application of the ON-Marg are used to identify inequalities in health and social well-being.  There are four dimensions of marginalization that are represented by the index, including residential instability, material deprivation, dependency, and ethnic concentration.3,4 Residential instability is measured using population dwelling characteristics, such as the proportion of the population living alone, the proportion of dwellings that are apartment buildings, and the proportion of the population who are single, widowed, or divorced. Material deprivation represents socioeconomic characteristics of the population, such as the proportion that is considered low income, the proportion receiving government transfer payments, and the proportion aged 15+ who are unemployed. The dependency dimension uses indicators of population dependency due to unemployment and younger or older age. Ethnic concentration is calculated using data indicating the proportion of the population that are recent immigrants (past 5 years) and the proportion who self-identify as a visible minority. |
| Canadian Vital Statistics – Death Database (CVSD) | This is an administrative survey that collects demographic and medical (cause of death) information annually and monthly from all provincial and territorial vital statistics registries on all deaths in Canada. The data are used to calculate basic indicators (such as counts and rates) on deaths of residents of Canada. Information from this database is also used in the calculation of statistics, such as cause-specific death rates and life expectancy. The cause of death variable in the death database is classified according to the World Health Organization "International Statistical Classification of Diseases and Related Health Problems" (ICD).  The central Vital Statistics Registry in each province and territory provides data from death registrations to Statistics Canada. The form for the registration of a death consists of personal information, supplied to the funeral director by an informant, and the medical certificate of cause of death, completed by the medical practitioner last in attendance, or by a coroner, if an inquest or enquiry was held.5 |
| Canadian Mortality Database (CMDB) | This dataset includes deaths that were reported to Statistics Canada after the reporting cut-off date for inclusion in the CVSD. |
| Historical Postal Code file | In Canada, income tax returns are submitted annually to the Canada Revenue Agency (CRA. The T1 Personal Master File (T1PMF), also known as the T1 General and Schedules, is a collection of the income tax returns shared by the CRA with Statistics Canada, and it provides income and demographic (e.g., date of death) information on tax filers in Canada. Every resident of Canada who earns taxable income is required to complete an income tax return, known as a T1 form, at the end of the year in which the income was received. Therefore, the T1PMF includes almost all individuals who filed an individual T1 tax return for the year of reference (i.e., some late filers may not be included) or those who received Canada Child Tax Benefits (CCTB) and their non-filing spouses.  The T1PMF is the principal data source for the Historical Postal Code file. Mailing address postal codes reported on these tax file were extracted to estimate a person’s place of residence for that reference year. Note that for some tax filers, the mailing addresses used for filing T1 tax records may not be associated with their place of residence (e.g. P.O. box, accountants’ or lawyers’ offices, parents’ addresses for young adults, children’s addresses for elderly parents). |
| Postal Code Conversion File Plus (PCCF+) | The purpose of the Postal Code Conversion File Plus (PCCF+) is to provide a link between six-character postal codes produced by Canada Post, standard census geographic areas (such as dissemination areas, census subdivisions, and census tracts) produced by Statistics Canada, and supplementary administrative areas and neighbourhood income quintiles.  Postal codes do not respect census geographic boundaries and so may be linked to more than one standard geographic area, or assigned to more than one set of coordinates. Therefore, one postal code may be represented by more than one record. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **eTable 2. Administrative codes used to define variables.** | | | | |
| **Concept** | **Databases** | **Code Type** | **Codes** | **Algorithm** |
| ***Exclusion Criteria*** | | | | |
| Status epilepticus | DAD, NACRS | ICD-9 | 3452, 3453 | Detection of any code – algorithm not validated 6 |
| ICD-10 | G41 |
| ***Comorbidities*** |  |  |  |  |
| Alcohol or drug abuse or dependence | DAD, NACRS | ICD-10 | F10, F14, F16, F19, F55 | Detection of any code |
| Brain surgery | DAD, NACRS | CCP | 14.1, 14.2, 14.3, 14.4, 14.5, 14.82, 15.02, 15.14, 15.2, 15.9 | Detection of any code |
| CCI | 1AA, 1AB52MESJ, 1AB52MFSJ, 1AB52MQSJ, 1AC, 1AF27JA, 1AF27JADC, 1AF27JADE, 1AF27JADG, 1AF27JX, 1AF27JXDC, 1AF27JXDE, 1AF27JXDG, 1AF52SZ, 1AF52SZTS, 1AF53SEQS, 1AF54HAQS, 1AF54SEQS, 1AF55SEQS, 1AF59SZAD, 1AF59SZAW, 1AF59SZGX, 1AF59SZX7, 1AF87DAAG, 1AF87DAAZ, 1AF87DAGX, 1AF87LUAG, 1AF87LUAZ, 1AF87LUGX, 1AF87PRAG, 1AF87PRAZ, 1AF87PRGX, 1AF87SZAG, 1AF87SZAZ, 1AF87SZGX, 1AG, 1AJ, 1AK, 1AN, 1AP, 1BA, 1DL, 1DR, 1EA, 1EH, 1JW, 1JX, 2AA, 2AC, 2AE, 2AF, 2AG, 2AJ, 2AK, 2AN, 2AP, 2BA, 2EA, 2JW |
| Brain tumour | DAD, NACRS | ICD-9 | 191, 1920, 1921, 1928, 1929, 1943, 1944, 1983, 2250, 2252, 2258, 2259, 2370, 2371, 2375, 2377, 2379, 2396 | Detection of any code |
| ICD-10 | C700, C709, C710, C711, C712, C713, C714, C715, C718, C719, C72, D320, D330, D331, D332, D333, D337, D420, D430, D431, D432, D433, D437, D439, 938, 93900, 93903, 940-957 |
| Chronic kidney disease | DAD, NACRS | ICD-10 | E102, E112, E132, E142, I12, I13, N08, N18, N19 | Detection of any code |
| CNS Infection | DAD, NACRS | ICD-9 | 0065, 036, 013, 045, 046, 047, 048, 049, 0520, 0521, 0530, 0531, 0543, 0550, 062, 063, 064, 0662, 0721, 0722, 094, 1142, 1008, 320, 321, 322, 323, 3240, 3249, 325 | Detection of any code – algorithm not validated (ICD-9,7,8 ICD-108,9) |
| ICD-10 | A066, A17, A321, A390, A394, A399, A504, A521, A522, A80, A81, A82, A83, A84, A85, A86, A87, A88, A89, A922, A923, B003, B004, B010, B011, B020, B021, B050, B051, B060, B261, B262, B375, B384, B451, B574, B582, B602, B941, G00-G07, G36 |
| Dementia | DAD, NACRS | ICD-9 | 2900, 2901, 2902, 2903, 2904, 294, 3310, 3311, 3315, 33182 | Detection of any code 10 |
| ICD-10 | F00-F03, G30 |
| Depression or anxiety disorder | DAD, NACRS | ICD-9 | 2962, 2963, 3000, 3002, 3003, 3004, 311 | Detection of any code |
| ICD-10 | F32, F330, F331, F332, F333, F338, F339, F341, F40, F41 |
| Diabetes | DAD, NACRS | ICD-10 | E10, E11, E130, E1310, E1312, E140, E1410, E1420 | Detection of any code 11 |
| Epilepsy or seizures | DAD, NACRS | ICD-9 | 3332, 3450, 3451, 3454, 3455, 3457, 3458, 3459, 7803 | Detection of any code – algorithm not validated 6 |
| ICD-10 | G40 |
| Stroke | DAD, NACRS | ICD-9 | 431, 4240, 4241, 436 | Detection of any code |
| ICD-10 | G450, G451, G452, I61, I63, I64 |
| Traumatic brain injury (TBI) or mild TBI | DAD, NACRS | ICD-9 | 800, 801, 803, 804, 850-854, 9590 | Detection of any code – algorithm not validated (ICD-9,12,13 ICD-1014) |
| ICD-10 | F072, S020, S021, S023, S027, S028, S029, S06, S071, T902, T905 |
| ***Outcome*** | | | | |
| Status Epilepticus | DAD | ICD-10 | G41 | Detection of any code– algorithm not validated6 |
| \*CCP= Canadian Classification of Diagnostic, Therapeutic and Surgical Procedures  \*CCI= Canadian Classification of Health Interventions  Note: Code descriptions are available upon request. | | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **eTable 3. Data type and category groupings used for covariates, by analysis.** | | | |
| **Variable** | **Analysis** | **Data Type** | **Category Groupings** |
| Age | Bivariate | Continuous | N/A |
| Bivariate and Multivariable | Categorical | 1=18-24  2=25-34  3=35-44  4=45-54  5=55-64  6=65-74  7=75-84  8=85+ |
| Sex | Bivariate and Multivariable | Categorical | 1=Female  2=Male |
| Educational attainment | Bivariate and Multivariable | Categorical | 1= Less than high school  2= High school  3= Post-secondary certificate or diploma  4= University degree |
| Marital status | Bivariate and Multivariable | Categorical | 1= Divorced or separated  2= Married  3= Never married  4= Widowed |
| Employment status | Bivariate and Multivariable | Categorical | 1= Employed  2= Unemployed or not in labour force |
| Race | Bivariate and Multivariable | Categorical | 1= White  2= Non-white |
| Immigration Status | Bivariate and Multivariable | Categorical | 1=Non-immigrants  2= Immigrants  3= Non-permanent residents |
| Second generation Canadian | Bivariate and Multivariable | Categorical | 1= Second generation  2= Not second generation |
| Canada’s official language | Bivariate and Multivariable | Categorical | 1= English only or French only  2= Both  3= Neither |
| Mother tongue | Bivariate and Multivariable | Categorical | 1= English  2= French  3= Other |
| Weekly number of hours spent on unpaid housework | Bivariate and Multivariable | Categorical | 1= Less than 5 hours  2= 5 to 14 hours  3= 15 to 29 hours  4= 30 or more hours |
| Weekly number of hours spent on unpaid childcare | Bivariate and Multivariable | Categorical | 1= Less than 5 hours  2= 5 to 14 hours  3= 15 or more hours |
| Weekly number of hours spent on unpaid senior care | Bivariate and Multivariable | Categorical | 1= Less than 5 hours  2= 5 or more hours |
| Household conditions: in need of repairs | Bivariate and Multivariable | Categorical | 1= Not in need of repairs  2= In need of minor repairs  3= In need of major repairs  4= Not applicable |
| Individual-level household income | Bivariate | Categorical | 1= Quintile 1 (lowest)  2= Quintile 2  3= Quintile 3  4= Quintile 4 or 5 |
| Multivariable | Categorical | 1= Quintile 1 (lowest)  2= Quintile 2  3= Quintile 3  4= Quintile 4  5= Quintile 5 |
| Neighbourhood-level household income | Bivariate and Multivariable | Categorical | 1= Quintile 1 (lowest)  2= Quintile 2  3= Quintile 3  4= Quintile 4  5= Quintile 5  6=Missing |
| Neighbourhood marginalization: instability | Bivariate and Multivariable | Categorical | 1= Quartile 1 (lowest)  2= Quartile 2  3= Quartile 3  4= Quartile 4 |
| Neighbourhood marginalization: deprivation | Bivariate and Multivariable | Categorical | 1= Quartile 1 (lowest)  2= Quartile 2  3= Quartile 3  4= Quartile 4 |
| Neighbourhood marginalization: dependency | Bivariate and Multivariable | Categorical | 1= Quartile 1 (lowest)  2= Quartile 2  3= Quartile 3  4= Quartile 4 |
| Neighbourhood marginalization: ethnic concentration | Bivariate and Multivariable | Categorical | 1= Quartile 1 (lowest)  2= Quartile 2  3= Quartile 3  4= Quartile 4 |
| Rurality | Bivariate and Multivariable | Categorical | 1= Rural  2= Urban |
| Alcohol or drug abuse | Bivariate and Multivariable | Categorical | 0=No  1=Yes |
| Brain surgery | Bivariate and Multivariable | Categorical | 0=No  1=Yes |
| Brain tumour or cancer | Bivariate and Multivariable | Categorical | 0=No  1=Yes |
| CKD | Bivariate and Multivariable | Categorical | 0=No  1=Yes |
| CNS infection or TBI | Bivariate and Multivariable | Categorical | 0=No  1=Yes |
| Dementia | Bivariate and Multivariable | Categorical | 0=No  1=Yes |
| Depression/anxiety | Bivariate and Multivariable | Categorical | 0=No  1=Yes |
| Diabetes | Bivariate and Multivariable | Categorical | 0=No  1=Yes |
| Epilepsy/seizures | Bivariate and Multivariable | Categorical | 0=No  1=Yes |
| Stroke | Bivariate and Multivariable | Categorical | 0=No  1=Yes |
|  |  |  |  |

**References**

1. Juurlink D, Preyra C, Croxford R, et al. Canadian Institute for Health Information Discharge Abstract Database: a validation study. Toronto; 2006.

2. Statistics Canada. Canadian Cancer Registry (CCR) [online]. 2023. Accessed at: https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3207. Accessed October 25, 2023.

3. Matheson FI, Dunn JR, Smith KLW, Moineddin R, Glazier RH. Ontario Marginalization Index: user guide version 1.0. Toronto, Ontario; 2012.

4. Matheson F, van Ingen T. 2011 Ontario Marginalization Index: technical document. Toronto, Ontario: Ontario Agency for Health Protection and Promotion (Public Health Ontario); 2017.

5. Statistics Canada. Statistics Canada, Canadian Vital Statistics - Death database (CVSD) [online]. 2023. Accessed at: https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3233. Accessed October 25, 2023.

6. Jetté N, Reid AY, Quan H, Hill MD, Wiebe S. How accurate is ICD coding for epilepsy? Epilepsia. 2010;51:62–69.

7. Hasbun R, Rosenthal N, Balada-Llasat JM, et al. Epidemiology of meningitis and encephalitis in the United States, 2011-2014. Clinical Infectious Diseases. Oxford University Press; 2017;65:359–363.

8. Gedeborg R, Furebring M, Michaëlsson K. Diagnosis-dependent misclassification of infections using administrative data variably affected incidence and mortality estimates in ICU patients. J Clin Epidemiol. Elsevier USA; 2007;60:155.e1-155.e11.

9. Ahlers FS, Benros ME, Dreier JW, Christensen J. Infections and risk of epilepsy in children and young adults: A nationwide study. Epilepsia. Blackwell Publishing Inc.; 2019;60:275–283.

10. Jaakkimainen RL, Bronskill SE, Tierney MC, et al. Identification of Physician-Diagnosed Alzheimer’s Disease and Related Dementias in Population-Based Administrative Data: A Validation Study Using Family Physicians’ Electronic Medical Records. J Alzheimers Dis. 2016;54:337–349.

11. Jiang J, Southern D, Beck CA, James M, Lu M, Quan H. Validity of Canadian discharge abstract data for hypertension and diabetes from 2002 to 2013. CMAJ Open. Joule Inc.; 2016;4:E646–E653.

12. Mcgarry LJ, Thompson D, Millham FH, et al. Outcomes and costs of acute treatment of traumatic brain injury. Epub 2002.

13. Bazarian JJ, Veazie P, Mookerjee S, Lerner EB. Accuracy of mild traumatic brain injury case ascertainment using ICD-9 codes. Academic Emergency Medicine. 2006;13:31–38.

14. Chen AY, Colantonio A. Defining neurotrauma in administrative data using the International Classification of Diseases Tenth Revision. Emerg Themes Epidemiol 2011.