Creating an 11-year longitudinal substance use harm cohort from linked health and census data to analyze social drivers of health

Abstract

Introduction: Research on substance use harm in Saskatchewan has been hampered by an absence of linked data to analyze and report on the social drivers of substance use harm.

Objectives: This study aims to address this gap by providing a fully annotated STATA do-file that links sociodemographic data to 11 years of hospitalization and death outcomes. This do-file will greatly facilitate the creation of provincial and national substance use cohorts using line-level data available through Statistics Canada's Research Data Centres (RDC) program.

Methods: We used Canadian Census Health and Environment Cohorts (CanCHEC) 2006 to create a cohort of Saskatchewanians followed from 2006 to 2016. We linked sociodemographic information of the 2006 Census (long-form) respondents to their hospitalization data captured in the Discharge Abstract Database (DAD) (2006 to 2016) and their mortality records in the Canadian Vital Statistics Death Database (CVSD) (2006 to 2016). We developed an algorithm to identify Saskatchewanians who experienced a substance use harm event. We validated the cohort by comparing our descriptive findings with those from other Canadian studies on substance use.

Results: We used CanCHEC, a national data resource, whereas most previous studies have used provincial data resources. Despite this difference in constructing the cohorts, our results showed trends consistent with previous studies, including an overrepresentation of individuals with lower socioeconomic status among the people who experienced substance use harm (PESUH). Similar to other Canadian studies, our results indicate an increasing rate of substance use harm from 2006 to 2016.

Conclusion: This study provides a STATA do-file that compiles a validated substance use cohort using CanCHEC, enabling comprehensive substance use research by linking sociodemographic data with health outcomes. The do-file is likely to save researchers hundreds of hours and accelerate research on the drivers of substance use harms in Canada.

Highlights

- This study provides a fully annotated Stata do-file, including a detailed walkthrough for using CanCHEC to create national or provincial substance use cohorts.
- CanCHEC links health system and census data allowing researchers to measure and examine inequalities in substance use harm across socioeconomic and ethnocultural dimensions over different periods and locations in Canada.
- There was a steady increase in people who experienced substance use harm in Saskatchewan, from 2006 to 2016.
- People who experienced substance use harm between 2006 and 2016 were overrepresented among individuals with an education level below high school, those in the lowest income quintile, residents of rural areas, and Indigenous population.

Introduction

Research on social drivers of substance use harm (SUH) in Canada has been hampered by an absence of infrastructure that supports routine access to and analysis of linked health administrative and sociodemographic data. This critical gap has prevented the production of a comprehensive population level and over-time portrait of substance use in Saskatchewan. Such data is essential for situating smaller-scale and local studies, untangling the "fundamental causes"[1,2] of health, informing decision-making, and improving service delivery and population health in Canada.[3,4]

Understanding social determinants of substance use harm is vital, given the profound influence they have on an individual's trajectory with using substances. Substance use is not simply a lifestyle "choice". Past circumstances, such as families, neighbourhoods, income, education, employment, and occupation can play significant parts in future substance use-related harm.[5] Moreover, these same social determinants also mediate access to care.[6,7] Furthermore, since socioeconomic status is often spatially concentrated, there is a geographical variation in substance use harm rates across Canada.[5] Recent Canadian research has found that although substance use harm is more concentrated and visible in urban areas, rates are actually higher in rural areas.[8] Additionally, mental health and substance use are known drivers of high-cost healthcare use in Canada.[9–11] To curb related healthcare spending, it is crucial to identify factors amenable to public policies that potentially can reduce these costs.[9] This also underscores the need for a comprehensive data infrastructure enabling researchers to examine related factors, such as socioeconomic status.

There have been efforts to create provincial cohorts to delve into substance use harm in regions like British Columbia[12] and Saskatchewan.[13] However, these studies had limited ability to provide socio-economic insights due to the absence of linked health data to social information. This gap can be bridged by using rich national databases, such as the Canadian Census Health and Environment Cohorts (CanCHEC), accessible at Research Data Centres (RDC)[14] throughout the country. The RDC facilitates research that uses sensitive microdata within a secure research environment managed by Statistics Canada.[14] Crucially, previous studies have also not published the code and documentation they used to compile their cohorts in ways that future researchers could replicate, verify, and build on their findings.

In response to the increased demand for including social data in epidemiology research, Statistics Canada has launched the CanCHEC project to relate social information gathered through its bi-decennial censuses (long form) and National Household Survey (NHS) to health administrative and registry databases at the individual level.[15] This includes the Discharge Abstract Database (DAD), Canadian Vital Statistics Death Database (CVSD), National Ambulatory Care Reporting System (NACRS), and Canadian Cancer Registry (CCR). CanCHEC also links these data to Canada Revenue Agency (CRA) tax files to trace where people have historically resided, making the data ideal for environmental exposure research.[15]

CanCHEC provides researchers with an excellent opportunity to measure and examine health inequalities across socioeconomic and ethnocultural dimensions for different periods and locations in Canada.[15] Yet, to date, only one study has used CanCHEC to examine the social drivers of substance use harm in Canada.[17] Carrière et al. used it to describe the socioeconomic characteristics of those experiencing hospitalizations due to opioid poisoning between 2011 and 2016.[17]

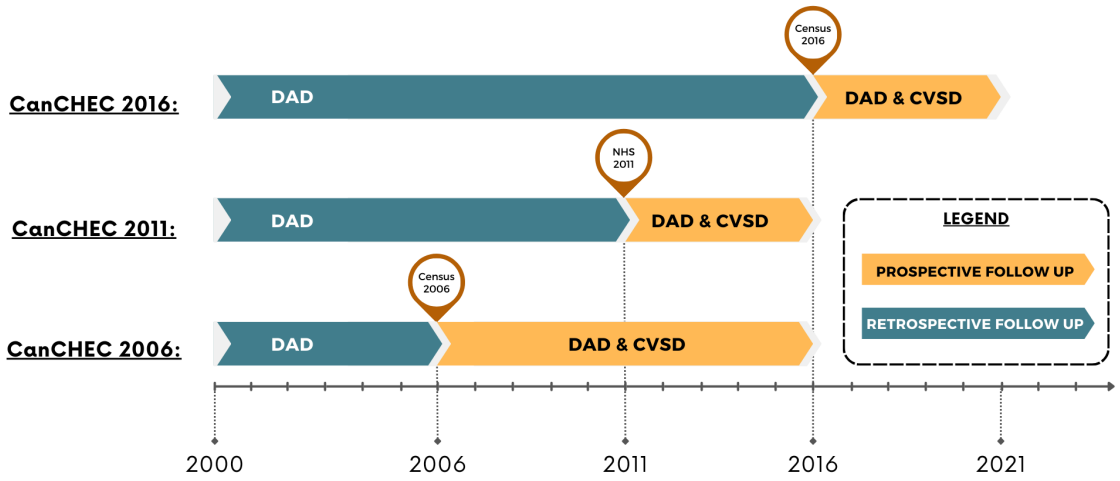
The main objective of this study is to develop a STATA do-file that links sociodemographic data to health outcomes to create provincial and national substance use cohorts, facilitating further research on substance use harm at national, provincial, and local levels. To validate this cohort, we compared our descriptive findings against those from previous studies on substance use in Canada, such as the British Columbia[12] and Saskatchewan[13] substance use cohorts. An ancillary objective is to highlight the potential CanCHEC holds for a micro-level linkage of socioeconomic and ethnocultural data to health administrative data.

Methods

Data sources

"CanCHEC is a series of population-based, probabilistically linked datasets that combine data from respondents to the long-form census or 2011 National Household Survey (NHS) with administrative health data (e.g., mortality, cancer incidence, hospitalizations, emergency ambulatory care) and annual postal code history".[15] As of June 2024, CanCHEC includes six cycles spanning from 1991 to 2016.[16] Figure 1 illustrates the period where each Census 2006, NHS, and Census 2016 respondents' socioeconomic and ethnocultural information is linked to DAD and CVSD. While Census 2006 and NHS are linked to DAD from 2000 to 2016, Census 2016 is linked to this database from 2000 to 2021. Amongst these three cycles, CanCHEC 2006 offers the longest prospective study period (10 years) and the shortest retrospective period (5 years), whereas Census 2016 provides the shortest prospective study period (5 years) and the longest retrospective follow-up (15 years).[16] Given that retrospective follow-up leads to survivorship bias,[18,19] we selected the 2006 cycle for this study to maximize the prospective follow-up duration (10 years).

Figure 1. CanCHEC cycles 2006, 2011, and 2016.



Notes: DAD: Discharge Abstract Dataset. CVSD: Canadian Vital Statistics Death Database. NHS: National Household Survey.

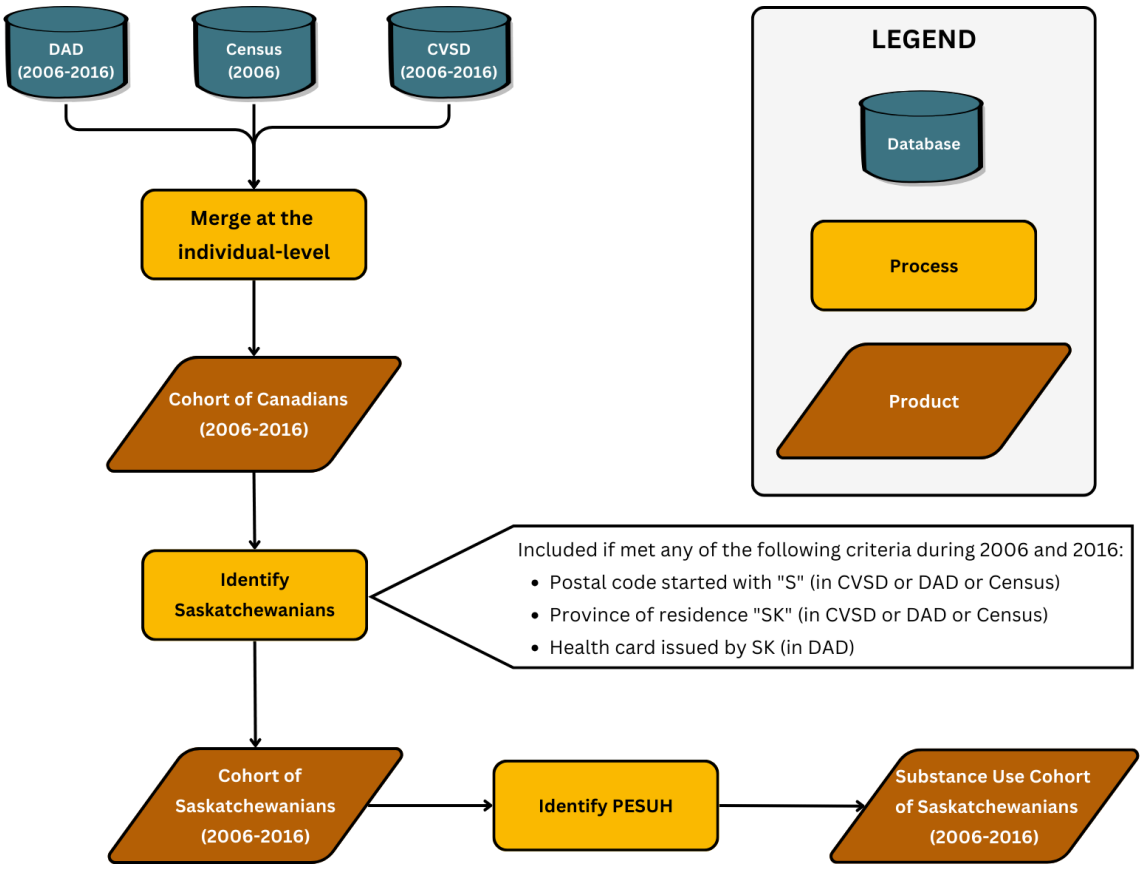
The CanCHEC 2006 includes usual residents of Canada as of the census date, capturing both permanent and non-permanent residents who responded to the 2006 long-form Census,[20] but excluding institutionalized populations (e.g., those living in nursing homes, penitentiaries,

and group homes). [15] As a result of this exclusion criteria, of the 6,463,927 long-form census respondents in 2006, 90.8% (5,871,337 records) were qualified for CanCHEC.[15] These records are linked to 436,407 CVSD[21] records (2006 to 2016) and 8,923,516 DAD[22] records (2000/2001 to 2016/2017).[15] For this study, we linked Census 2006 respondents eligible for CanCHEC to 2006-2016 hospitalization and death data recorded in DAD and CVSD to follow up the cohort prospectively. We excluded the NACRS[23] due to its inconsistent coverage in Saskatchewan between 2006 and 2016.

Cohort creation

Figure 2 details the three-step process of creating a provincial substance use cohort for Saskatchewan from the CanCHEC datasets, that can be operationalized using the do-file provided in Appendix 1. In step one, we created a national cohort by linking sociodemographic and administrative health data. In the second step, we extracted Saskatchewan residents from this national cohort to form the provincial cohort. Step three involved identifying people who experienced substance use harm (PESUH). Cohorts for other provinces can be readily created by extracting their residents instead, or users of the do-file can create a national cohort by skipping the second step. The do-file is fully annotated and includes extensive notes on the inputs and outputs of the cohort and the methodological details that inform it. These detailed annotations are likely to also help researchers who may wish to implement the cohort in other analytic clients, such as RStudio or SAS.

Figure 2. Creating the substance use cohort of Saskatchewanians.



Notes: DAD: Discharge Abstract Dataset. CVSD: Canadian Vital Statistics Death Database. PESUH: People who have experienced substance use harm. SK: Saskatchewan.

1. Creating a national cohort by linking hospitalization, mortality, and sociodemographic data

CanCHEC datasets have key files enabling the linkage of Census data to administrative health data. Appendix 2 illustrates the relationship between CanCHEC 2006's datasets used in this study to link the sociodemographic data available in the Census to hospitalization records in DAD (2006 to 2016) and mortality data in CVSD (2006 to 2016), at the individual level. The final output at this step was a cohort of Canadians whose hospitalization and mortality information were tracked from 2006 to 2016.

2. Creating a provincial cohort (Saskatchewan)

We used the national cohort produced in the previous step to create a provincial cohort that includes people who resided in Saskatchewan at any point in time between 2006 and 2016. We defined a set of criteria to identify these people. We considered a person Saskatchewanian if the postal code for their place of residence started with "S" in CVSD or DAD or Census; or if their province of residence was Saskatchewan in CVSD or DAD or Census; or if their health card was issued by Saskatchewan as recorded in DAD. In other words, a person who lived in British Columbia in 2006 but moved to Saskatchewan in 2008 and thereafter experienced a substance use harm event recorded in the CVSD or DAD would appear in our cohort. There were a small number of records whose postal codes started with "S" but their province of residence was Alberta in the Census records. After further investigation and mapping using QGIS, it was found that these postal codes are related to Onion Lake First Nation, located on the boundary of Saskatchewan and Alberta. We elected to include these individuals because their community centre is located in Saskatchewan, where they primarily receive health care. At the end of this step, individuals who were identified as Saskatchewanians were included in the cohort, and the remaining records were excluded.

3. Identifying people who experienced substance use harm (PESUH)

We used the 10th revision of the International Statistical Classification of Diseases and Related Health program (ICD-10)[24] codes to identify hospitalizations or deaths that happened due to substance use. The case-finding algorithm used to identify SUH events is provided in Appendix 3. This algorithm was developed based on methodologies applied by other researchers.[12,25–28] We did not include opioid-related adverse drug reactions (Y45.0) as we were not interested in the harms resulting from the adverse effects of prescribed medications. It should also be noted that this list only contains harms that are 100% attributable to substance use, and therefore harms partially attributable to substance use (e.g., cancer, stroke) are excluded from this analysis.

After creating the cohort of Saskatchewanians, we identified two subpopulations of people who experienced substance use harm (PESUH) and people who have not experienced substance use harm (NPESUH) within the cohort. People were categorized in the PESUH group if they experienced at least one SUH event (i.e., hospitalization or death) between 2006 and 2016. Anyone who was not identified as PESUH was assigned to the NPESUH group.

Descriptive analysis

We described the cohort of Saskatchewanians and its subgroups, PESUH and NPESUH, by age, sex, income, region, ethnicity, occupation, employment status, education, and number of hospitalizations (See Appendix 4 for more detail on variables.) We used the after-tax household income variable captured by the census and adjusted it to the household size. The

adjusted after-tax household income was then used to create income quintiles at the provincial level. We also plotted the annual rates of PESUH per 100,000 people, from 2006 to 2016, to present a picture of substance use harm trends in Saskatchewan, broken down by substance categories. In calculating the annual rates, we took the number of PESUH in a year as the numerator, and the number of people in the cohort who were alive in that year as the denominator. We applied the CanCHEC weight variable in all the calculations to produce estimates representative of the non-institutional population in Saskatchewan at the time of Census 2006. Moreover, for privacy considerations, we rounded the numerators and denominators to the nearest multiple of five before calculating the rates and percentages, following the Statistics Canada Research Data Centre (RDC) vetting rules.

Data was accessed via the Saskatchewan Research Data Centre on the University of Saskatchewan campus (SKY-RDC).[14] Ethics exemption was obtained from the Saskatchewan Health Authority Research Ethics Board (REB-22-20).

Validation analysis

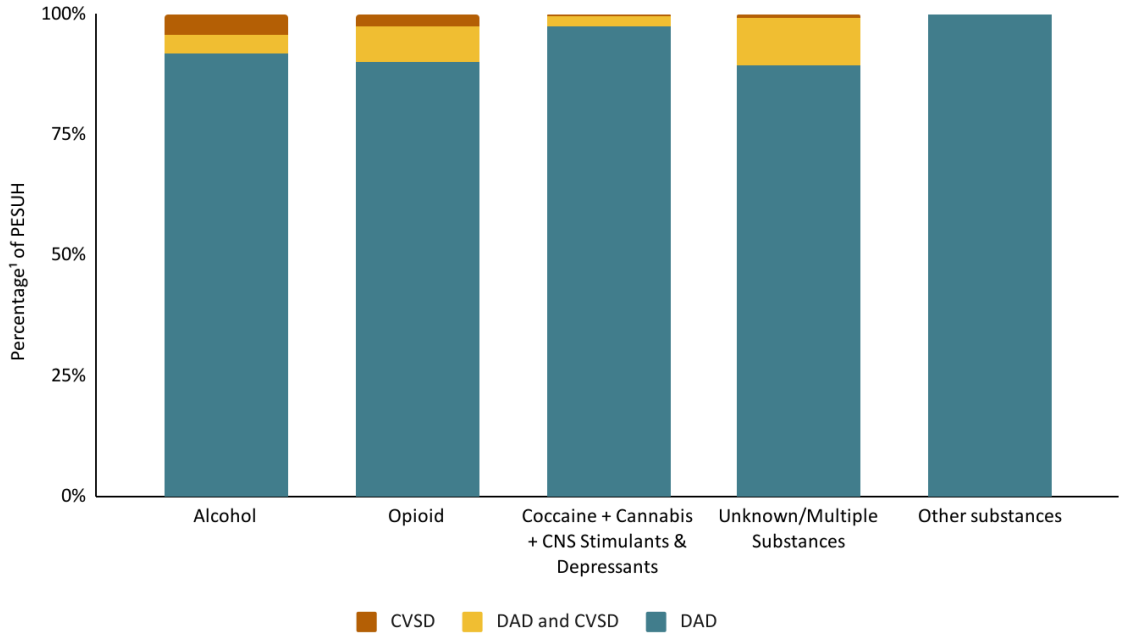
We validated our cohort by comparing its descriptive characteristics with those from other Canadian studies on substance use. In doing so, we profiled the PESUH group using various social determinants of health and compared our results with what has been found by other researchers using different data sources. This validation was crucial, as CanCHEC's dataset structure is fundamentally different compared to those used in other studies. CanCHEC is a federal data repository linking various national datasets at the individual level. The secure research environment provided by RDCs has made it possible for the researchers to access such line-level data. However, Canadian researchers often use provincial health data repositories that are not linked to social information and are limited to only one province.

Results

Identification of people who experienced substance use harm

Out of the cohort population of 228,000, approximately 7,000 individuals (3.07%) experienced at least one SUH event from 2006 to 2016 (PESUH). These individuals were further categorized based on the type of substances recorded as a cause of their hospitalization or death. Notably, across all substance categories, over 95% of these individuals were identified exclusively through their hospitalization records. This indicates that while they were hospitalized due to that substance during the follow-up period, there were no recorded deaths related to that substance. A very small portion of the PESUH in the cohort were identified through both hospitalization and death records and an even smaller fraction solely via CVSD. The data sources used for identifying PESUH are presented in Figure 3.

Figure 3. Identification of people who experienced substance use harm.

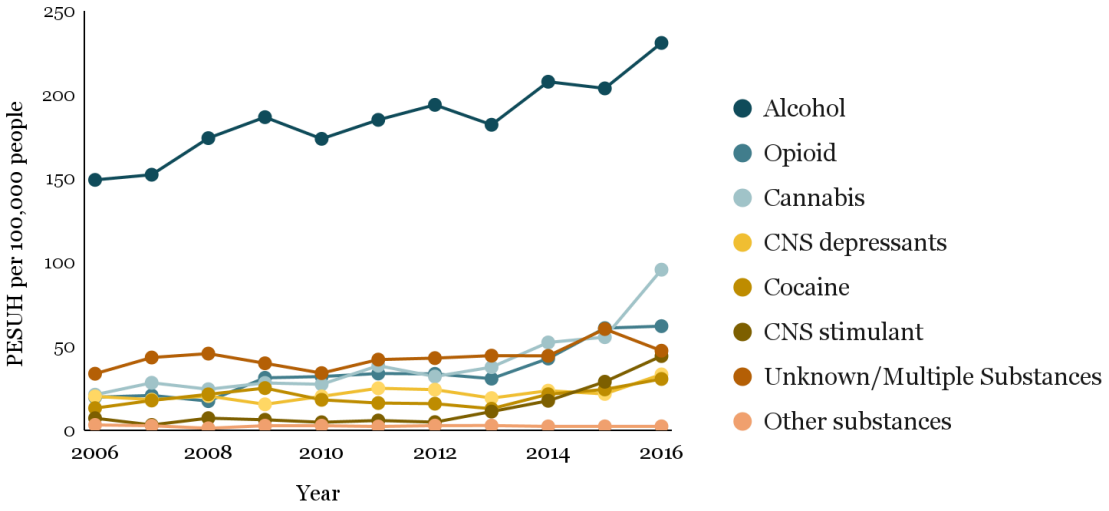


Notes: We aggregated cocaine, cannabis, CNS stimulants, and CNS depressants due to the low numbers of people falling in the CVSD categories. The RDC vetting rules do not allow releasing percentages with a numerator or denominator smaller than 5 (unweighted). PESUH: people who experienced substance use harm. CNS: central nervous system. [1]Numerator: weighted number of PESUH identified by data source; denominator: weighted number of PESUH.

Trends by SUH category, 2006 to 2016

Our findings indicate an increase in rates of PESUH per 100,000 population in Saskatchewan from 2006 to 2016 (Figure 4). This uptrend spanned across most substance categories examined in this study. While there has been steady growth over these years, a substantial jump in the PESUH rate was evident between 2015 (55 per 100,000 population) and 2016 (95 per 100,000 population) for the cannabis category. We found that in Saskatchewan, the rates of PESUH for alcohol and cannabis were 230 and 95 per 100,000 population in 2016, compared to 203 and 55 per 100,000 population in 2015. We also calculated the age-standardized rates to ensure that these trends are not affected by the fact that our cohort is getting older through the 11 years of follow-up. The trends for age-standardized rates were similar to the non-age-standardized rates trends (not reported), meaning that the ageing of the cohort was not a factor here.

Figure 4. Trends of the number of people who have experienced substance use harm in Saskatchewan by substance category, 2006 to 2016.



Notes: The data table on these trends is provided in Appendix 5. PESUH: people who experienced substance use harm. CNS: central nervous system.

Number of hospitalizations

Our analysis indicates that people who have experienced at least one SUH event between 2006 and 2016 were far more likely to be admitted to the hospital during this timeframe. While 44.9% of the PESUH group had 2 to 5 hospital admissions, 40.4% were admitted 6 times or more. In contrast, the rest of the population exhibited lower hospitalization rates, with 31.8% admitted 2 to 5 times and only 11.5% admitted 6 times or more in the same period. However, it should be noted that since a primary criterion for inclusion in the PESUH group was hospitalization due to substance use, some of this gap may be due to this inclusion criterion.

Socio-demographic characteristics

Table 1 provides the socio-demographic characteristics of the entire cohort and its two subgroups in 2006. Overall, modal individuals who experienced substance use harm events during the follow-up period were young (median age of 35), white[1] (56.0%), working (37.6%) or seeking employment (9.5%), held less than a high school degree (39.6%), and resided in urban areas (62.3%) in 2006 (study baseline). However, the PESUH group had an overrepresentation of individuals with less than a high school education (39.6% in PESUH vs. 23.7% in NPESUH), those in the low-income quintile (42.5% in PESUH vs. 19.5% in NPESUH), residents of rural areas (37.7% in PESUH vs. 34.2% in NPESUH), and individuals of Indigenous ethnicity (42.7% in PESUH vs. 13.8% in NPESUH).

Table 1: Socio-demographic characteristics of Census 2006 respondents, by having experience of substance use harm between 2006 and 2016.

[1] In Census, “White” is one of the response categories in the population group question. where people are asked to respond to in the Visible Minority question.[29] This is similar to what CIHI proposes in the “Guidance on the Use of Standards for Race-Based and Indigenous Identity Data Collection and Health Reporting in Canada” report. [30]

Characteristics	Provincial cohort (n ≈ 228,000)	PESUH (n ≈ 7,000)	NPESUH (n ≈ 221,000)
Age at cohort entry			
25 percentile (SE)	18 (0.00)	19 (0.46)	18 (0.00)
50 percentile (SE)	37 (0.19)	35 (0.75)	37 (0.33)
75 percentile (SE)	54 (0.22)	51 (0.61)	54 (0.44)
Sex			
female, % (SE)	51.2 (0.12)	46.8 (0.84)	51.3 (0.13)
male, % (SE)	48.8 (0.12)	53.2 (0.84)	48.7 (0.13)
Income, after-tax household income quintile in SK			
Quintile 1, % (SE)	20.0 (0.10)	42.5 (0.80)	19.5 (0.09)
Quintile 2, % (SE)	20.0 (0.10)	18.5 (0.65)	20.0 (0.11)
Quintile 3, % (SE)	20.0 (0.10)	14.8 (0.59)	20.1 (0.10)
Quintile 4, % (SE)	20.0 (0.10)	12.7 (0.59)	20.1 (0.11)
Quintile 5, % (SE)	20.0 (0.10)	11.4 (0.56)	20.2 (0.11)
Region of residency			
Rural, % (SE)	34.3 (0.12)	37.7 (0.73)	34.2 (0.12)
Urban, % (SE)	65.7 (0.12)	62.3 (0.73)	65.8 (0.12)
Ethnicity			
Black, % (SE)	0.5 (0.02)	0.3 (0.09)	0.5 (0.02)
East Asian, % (SE)	1.0 (0.03)	0.3 (0.09)	1.1 (0.03)
Latin American, % (SE)	0.3 (0.01)	0.1 (0.06)	0.3 (0.01)
South Asian, % (SE)	0.7 (0.02)	0.4 (0.11)	0.7 (0.02)
Southeast Asian, % (SE)	0.7 (0.02)	0.2 (0.07)	0.7 (0.02)
White, % (SE)	82.0 (0.09)	56.0 (0.80)	82.6 (0.09)
Indigenous, % (SE)	14.4 (0.08)	42.7 (0.79)	13.8 (0.08)
Other, % (SE) [1]	0.4 (0.02)	0.1 (0.03)	0.4 (0.02)
Occupation			
Management occupations, % (SE)	4.6 (0.06)	2.6 (0.24)	4.7 (0.05)
Business, finance and administrative occupations, % (SE)	9.3 (0.07)	4.7 (0.39)	9.4 (0.08)
Natural and applied sciences and related occupations, % (SE)	2.5 (0.04)	1.2 (0.21)	2.5 (0.04)
Health occupations, % (SE)	3.7 (0.05)	2.2 (0.27)	3.7 (0.05)
Occupations in social science, education, government service and religion, % (SE)	4.9 (0.06)	2.3 (0.24)	5.0 (0.06)
Occupations in art, culture, recreation and sport, % (SE)	1.4 (0.03)	0.8 (0.15)	1.4 (0.03)
Sales and service occupations, % (SE)	14.7 (0.09)	16.2 (0.62)	14.7 (0.10)
Trades, transport and equipment operators and related occupations, % (SE)	9.5 (0.07)	11.2(0.54)	9.5 (0.08)
Occupations unique to primary industry, % (SE)	7.4 (0.07)	5.6 (0.37)	7.5 (0.07)
Occupations unique to processing, manufacturing and utilities, % (SE)	1.9 (0.04)	2.6 (0.29)	1.9 (0.04)
Not applicable, % (SE)	40.1 (0.13)	50.4 (0.87)	39.9 (0.12)
Employment status			
Not applicable, less than 15 years, % (SE)	19.4 (0.10)	15.5 (0.59)	19.5 (0.10)
Employed, % (SE)	52.0 (0.13)	37.6 (0.85)	52.3 (0.13)

Not in labour force, % (SE)	24.7 (0.10)	37.3 (0.79)	24.4 (0.11)
Unemployed, % (SE)	3.9 (0.05)	9.5 (0.49)	3.8 (0.05)
Educational attainment			
Less than high school, % (SE)	24.0 (0.11)	39.6 (0.80)	23.7 (0.11)
High school graduation certificate or equivalency certificate, % (SE)	21.6 (0.11)	19.8 (0.69)	21.6 (0.10)
Non-university post-secondary certificate or diploma, % (SE)	24.4 (0.11)	21.1 (0.70)	24.5 (0.11)
Bachelor's degree, % (SE)	7.7 (0.07)	2.5 (0.25)	7.8 (0.07)
University certificate or diploma above bachelor level, % (SE)	2.9 (0.04)	1.4 (0.22)	3.0 (0.04)
Not applicable (Institutional residents), % (SE)	19.4 (0.10)	15.5 (0.59)	19.5 (0.10)
Total number of hospitalizations, between 2006 and 2016			
0, % (SE)	37.3 (0.12)	1.9 (0.26)	38.0 (0.13)
1, % (SE)	18.5 (0.09)	12.8 (0.54)	18.6 (0.10)
2 to 5, % (SE)	32.1 (0.12)	44.9 (0.86)	31.9 (0.12)
6 and more, % (SE)	12.1 (0.08)	40.4 (0.83)	11.5 (0.08)

Notes: All percentages are weighted to the whole population of Canadians in 2006, using the CanCHEC cohort weight. Standard errors (SE) are calculated by bootstrapping. The difference between the calculated percentages for PESUH and NPESUH is statistically significant based on the performed t-test. [1] Other includes individuals who were categorized under the "multiple minorities" or "Middle Easterns" groups in the Visible Minority question in Census. PESUH: people who experienced substance use harm during 2006-16; NPESUH: people who did not experience substance use harm. SK: Saskatchewan.

Validation of cohort against previous studies

Despite the limitations of CanCHEC, we found that population estimates for our cohort were broadly comparable to those from other studies.

The proportion of PESUH identified through DAD and CVSD data aligns with Homayra et al.'s findings on the British Columbia cohort. According to a CIHI report, the highest age-standardized hospitalization rates in Saskatchewan were due to cannabis (345 per 100,000) and alcohol (259 per 100,000) use among youth aged 10 to 24 in 2017-2018.[8] Our findings show crude hospitalization rates of 95.6 and 230.9 per 100,000 population for cannabis and alcohol use, respectively, in 2016. The difference in cannabis rates is likely due to age variations in cannabis consumption.[31]

CIHI has also highlighted a higher representation of substance use-related harm in rural areas, with rural age-standardized hospitalization rates being significantly higher than urban rates.[8] This finding is consistent with our results, which show a higher prevalence of substance use harm in rural settings.

In terms of ethnicity, there is a conspicuous overrepresentation of the Indigenous population within PESUH (42.7%) compared to their representation within the entire cohort (14.4%) and the NPESUH group (13.8%). This overrepresentation aligns with studies showing higher hospitalization rates due to opioid poisoning among the Indigenous population[17] and higher reports of poly-substance use among students with Indigenous identity.[32]

More than one-third (42.5%) of PESUH were in the lowest income quintile at the study baseline, similar to CIHI's national report on substance use-related hospitalizations in 2019.[33] Similar trends were found in studies on opioid-related harms, with higher rates

among those in the lowest income quintile.[5,17] Education also inversely correlated with experiencing SUH events, with a higher percentage of PESUH having less than high school education (39.6%) compared to NPESUH (23.7%) and the overall cohort (24.0%). This finding is consistent with studies showing higher hospitalization rates due to opioid poisoning among those without a high school diploma.[17]

Employment status was another significant factor, with PESUH more likely to be unemployed (9.5%) compared to NPESUH (3.8%). Our cohort findings are similar to studies on the socioeconomic profile of individuals who experienced opioid overdoses, where employment rates were low (33.8%).[34] The investigation of hospitalization due to opioid overdose also presents similar findings, with the highest rates of hospitalization due to opioid poisoning among people who were unemployed (17.0 per 100,000).[17] Occupational data also revealed that PESUH were more likely to work in "sales and services," (16.2%), “trades, transport and equipment operators and related occupations” (11.2%), and “ occupations unique to processing, manufacturing and utilities” (2.6%) compared to NPESUH (14.7%, 9.5%, and 1.9%, respectively). This distribution is consistent with findings from studies on opioid poisoning hospitalizations, which also noted higher rates in similar occupational categories.[17,34]

In summary, our cohort findings are validated through their alignment with previous studies, demonstrating similar trends and distributions in substance use harm across different populations and contexts.

Discussion

Strengths of CanCHEC in constructing substance use cohorts

We compared our cohort with two others designed to study substance use harm in British Columbia (BC)[12] and Saskatchewan (SK).[13] The comparison was primarily aimed at understanding the potential advantages and challenges of using CanCHEC, as in our study, versus using provincial data repositories, as was the approach in the BC[12] and SK[13] cohorts, in creating longitudinal cohorts for studying substance use harm.

A standout advantage of using CanCHEC for substance use research lies in its capacity to link socioeconomic and ethnocultural data of individuals to their health records. This allows an in-depth analysis of disparities in substance use harm based on factors such as income, education, occupation, language, self-identified ethnicity, including First Nations, Métis, and Inuit, and immigration status, among others.[15] The availability of this information for cohort participants makes CanCHEC an ideal data source for studying inequalities and intersectionality among social factors in substance use harm. Such comprehensive linkage is not possible with cohorts exclusively using provincial data repositories, as observed with the BC[12] and SK[13] cohorts.

CanCHEC is a national data source allowing for inter-provincial comparison, which enables researchers to conduct comparative policy analysis[35] concerning substance use countrywide. Additionally, provincial cohorts are particularly beneficial for provinces without well-developed data infrastructures. For example, in Saskatchewan, researchers do not have access to CVSD. Moreover, the consistent linkage methodology used to create cohorts in CanCHEC provides the opportunity to examine trends over time both within and between CanCHEC cycles.[15] CanCHEC also includes an annual historical postal code file

from 1981 onward providing information on where the cohort members live year after year, which can be used to examine the impact of moving patterns on substance use harm. It should also be noted that CanCHEC administrative data updates continuously, enabling a prolonged follow-up of the individuals through time. Therefore, the current cohort created in this study may be extended to cover additional years in the foreseeable future.

Although some populations are less likely to be fully captured by CanCHEC, there are others it is likely to capture more fully. For instance, it is likely to excel at portraying substance use behaviours amongst the middle class,[36] which tends to be more likely to be employed and have higher education. Therefore, CanCHEC provides a great opportunity to investigate the characteristics of people who are outside the typical focus of substance use literature. A notable potential opportunity provided by CanCHEC is that it can also track the health outcomes of individuals who were teenagers or younger during a given census, providing potential insights into substance use and social mobility in their adolescence and young adulthood. The prospective follow-up of CanCHEC includes very thorough coverage of young, vulnerable populations whose households completed the census before they started to use substances, ensuring a comprehensive dataset. Although CanCHEC does not provide a complete picture of people experiencing substance use harm in Canada, our findings align with patterns seen in the most vulnerable populations. This highlights CanCHEC as a valuable data source for studying the social determinants of substance use harm and informing evidence-based policymaking.

Limitations of CanCHEC in constructing substance use cohorts

There are limitations in using CanCHEC to investigate substance use harm. Notably, we are not able to identify substance use harms that did not result in hospitalization or death, using only CanCHEC. This is because of the absence of certain databases, such as pharmacy, community-based services, clinics, and physician billing data. These databases are usually accessible through provincial repositories. In their work on creating a provincial cohort for British Columbia, Homayra et al. (2021) identified approximately half of all non-opioid non-alcohol substance use disorder cases (49.9%) using physician billing records exclusively. They also found approximately 30% of alcohol use disorder cases and more than 10% of opioid use disorder cases using only physician billing records. Moreover, Homayra et al. identified approximately 5% of opioid use disorder cases using only Pharmanet. However, there were not many cases of alcohol use disorder or non-opioid non-alcohol substance use disorder cases identified using Pharmanet alone.[12] Considering these findings, we anticipate that more cases would have been identified if physician billing records were also linked to CanCHEC. Nevertheless, it is important to note that our focus in this study is substance use harms and not substance use diagnosis in general, and only a portion of physician billing records indicates harm.

Moreover, since CanCHEC excludes the institutional population[2] (e.g., those living in nursing homes, penitentiaries, group homes) at baseline,[15] the created cohort is representative of the non-institutional population living in Canada at the time of the census. This makes the cohort population younger and healthier than the Canadian population. Substance use harm events occurring in correctional settings and following release are well documented and particularly prevalent immediately following release,[37–46] when many individuals are without an address or are residing in a group home so are not likely to be included with

[2] CancCHEC considers individuals institutionalized if they do not have any other residency address in Canada and they have been in an institution for not less than six months.

CanCHEC data. A recent scoping review also suggests that prescription misuse is a growing concern among older adults in Canada,[47] suggesting that harm events in nursing homes may also be missed in the CanCHEC data. Group homes, or sober-living houses, also offer a supportive living environment for many people seeking abstinence-based recovery or treatment options.[48,49] However, since a return to use is common for almost half of people engaging in recovery programs for substance use,[50] it is unlikely related harm events are captured by CanCHEC in these settings. Including institutional populations would increase the total number and frequency of harm events identified. On the other hand, the BC cohort effectively covers these missed populations. Out of 162,099 individuals in the BC cohort, 13,154 (8.1%) were homeless at least once during the study period from 1 April 2009 to 31 March 2017. Therefore, research focusing on this vulnerable population might benefit from using the BC cohort approach to investigate substance use harm.[12] However, it should be noted that the data infrastructure in BC that allows such an approach is not available in most provinces and territories in Canada, as is the case for Saskatchewan.

In general, census data quality reports indicate that a small proportion of Canadians is missed in any given census.[15] These people are more likely to be young, mobile, low-income, homeless, or Indigenous.[15] Given that these groups of the population are more likely to have experienced substance use harm,[5,17,51,52] related estimates produced by CanCHEC are likely to be underestimated compared to the Canadian population. Although the linkage error in CanCHEC is minimal, it is present,[15] and we do not know exactly how this affects our findings which could result in either an underestimation or overestimation of mortality and hospitalization rates attributed to substance use. Additionally, it is important to note that socio-demographic characteristics are only collected at the time of census, potentially not capturing shifts in an individual's socioeconomic status or identification throughout their life. Provincial substance use reporting includes our population. Our findings suggest that CanCHEC captures the vast majority of that population. Therefore, caution must be used when making inferences from CanCHEC, especially when generalizing to the broader population. The cohort created by CanCHEC is not appropriate for every research question due to this limitation.

Our cohort of Saskatchewanians might include individuals with hospitalization or mortality records over those without. This is due to our approach regarding identifying Saskatchewanians, where in addition to the census, we used DAD (2006-2016) and CVSD (2006-2016) data to find people who had an indication of residency in Saskatchewan. We adopted this approach to capture the greatest breadth of substance use harm events possible in Saskatchewan, even if respondents did not live in Saskatchewan at the time of the 2006 census. If required, this potential source of bias can be removed by making appropriate modifications to the second step in the do-file.

Finally, CanCHEC is a longitudinal database, tracking health outcomes of Census and NHS respondents both prospectively and retrospectively. In longitudinal designs, researchers compute two types of weights: longitudinal weight for studying trajectories over time, and cross-sectional weight for making inferences at specific points in time. While CanCHEC provides longitudinal weights, cross-sectional weights are absent and need to be calculated to ensure estimates reflect the population at a given time. Our research team is currently developing cross-sectional weights for each cycle of CanCHEC to facilitate pooled analyses for cross-sectional inferences.

Increase in cannabis-related PESUH rate in 2016

The increase seen in the rates of PESUH in the cannabis category in the final year of this study might be in part attributed to evolving national and provincial policies on substance use. Studies have shown an increase in reporting cannabis use among youth in Canada after its legalization in 2018.[29] A potential explanation for the spike in cannabis-related hospitalizations in pre-legalization years may be the evolving related policies throughout the countries in the pre-legalization period. For instance, in 2013, Dr. Fern Stockdale Winder's appointment as a Commissioner resulted in the *Working Together for Change: A 10-Year Mental Health and Addictions Action Plan for Saskatchewan* report by December 2014.[53] This report paved the way for increased mental health and substance use service awareness and destigmatization, with the recommendations being incorporated into various ministries' agendas for a more comprehensive approach to mental health and substance use challenges.[54] Nevertheless, further research should seek the underlying factors driving this trend.

Conclusion

The primary aim of this study was to develop a STATA do-file that links sociodemographic data to health outcomes to create provincial and national substance use cohorts, facilitating further research on substance use harm at national, provincial, and local levels. We validated our cohort by comparing our descriptive findings with those from other Canadian studies on substance use. An ancillary objective was to underscore the potential CanCHEC holds in studying substance use, especially by bridging the gap in data concerning the linkage of social determinants of health and administrative health data. We hope this study can serve as a precursor to many more that use CanCHEC to examine the social determinants of substance use harm events using linked social and health administrative data. Future studies could focus on small-area analysis and multivariate analyses to understand the geographic distribution and social drivers of substance use harm in Canada. Furthermore, researchers can adapt the STATA do-file developed by our research team to build their own cohorts using CanCHEC and investigate their research questions related to substance use (See Appendix 1).

Considering the findings of the current study, there is a clear need for greater research on possible changes in recording substance use hospitalizations by ICD codes in Canada. Understanding these changes is crucial in interpreting the rates of substance use harm. Further, policy impacts on substance use-related stigma and help-seeking behaviours warrant exploration, given their profound influence on substance use harm events statistics.

Acknowledgments: This research was funded by the Saskatchewan Health Research Foundation (SHRF). We are grateful to the Saskatchewan Research Data Centre analyst, Ruben Mercado, for all the help he provided in accessing data and vetting the results. We also thank Donica Janzen, a Senior Researcher at the Health Quality Council, and Raadiya Malam, a researcher and policy analyst at the Canadian Substance Use Costs and Harms (CSUCH), for the valuable information on substance use case-finding algorithms. The authors thank Carol Spencer, our patient family partner, for taking the time to give our research team excellent feedback and insightful comments. We also thank Daniel Yupanqui, Bong Soo Kim, and Shubrandu Sanjoy for their helpful comments throughout the project.

Statement of Conflict of Interest: The authors declare that they have no conflict of interest.

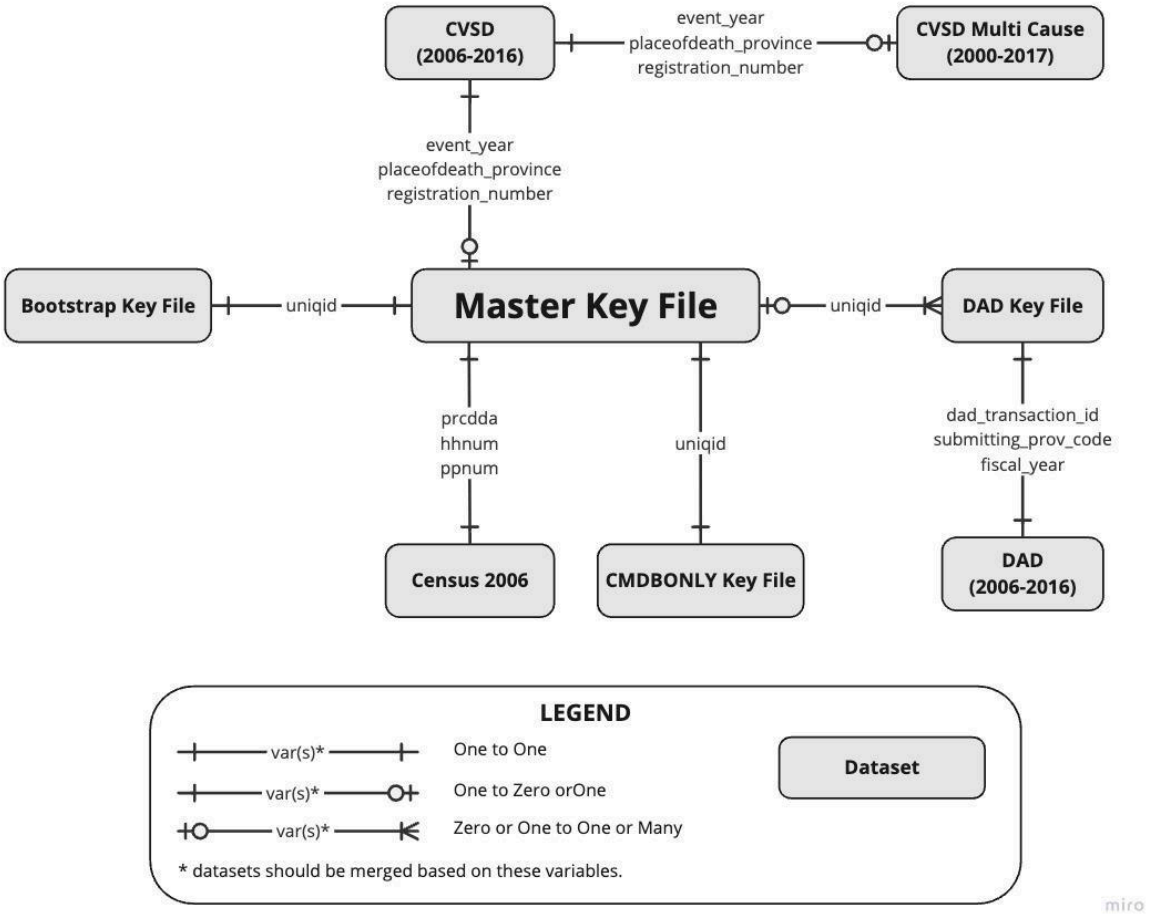
Ethics Statement: This study (REB-22-20) was deemed exempt by the University of Saskatchewan Research Ethics Board.

Data Availability Statement: Data used in this study are not publicly available due to confidentiality reasons. Researchers can request access to data by using the Microdata Access Portal of the Statistics Canada Research Data Centre program.

Supplementary Appendices

Appendix 1. STATA do-file. (Uploaded separately)

Appendix 2. CanCHEC 2006 Data Structure.



Note: This figure illustrates the relationship between different datasets available within CanCHEC 2006. The datasets shown in this figure only include those used in the current study and not all CanCHEC 2006 datasets. Data structures of different cycles of CanCHEC are different, and this diagram solely represents the cycle of 2006.

Appendix 3. Substance use case finding algorithm.

Database	Substance use category	Substance use case finding algorithm[1]	
		Diagnosis ICD-10-CA codes	**Diagnosis type**
DAD	Alcohol	F10.–, T51.–, E24.4, G31.2, G62.1, G72.1, I42.6, K29.2, K70.–, K85.2, K86.0, O35.4–, O99.3–[2], Q86.0, R78.0, X45, X65, Y15, P04.3	(M), (1), (2), (W), (X), (Y) or (9)
	Opioids	F11.–, T40.0, T40.1, T40.2–, T40.3, T40.4–, T40.6 O99.3–[2], P96.1	(M), (1), (2), (W), (X), (Y) or (9)
	Cannabis	F12.–, T40.7, O99.3–[2]	(M), (1), (2), (W), (X), (Y) or (9)
	Other CNS depressants	F13.–, T42.3, T42.4, T42.6, T42.7, O99.3–[2]	(M), (1), (2), (W), (X), (Y) or (9)
	Cocaine	F14.–, T40.5, O99.3–[2]	(M), (1), (2), (W), (X), (Y) or (9)
	Other CNS stimulants	F15.–, T43.6, O99.3–[2]	(M), (1), (2), (W), (X), (Y) or (9)
	Unknown and multiple substances	F19.–, T43.8, T43.9, O99.3–[2], X41, X42, X61, X62, Y11, Y12[3], O35.5	(M), (1), (2), (W), (X), (Y) or (9)
	Other substances[4]	F16.–, F18.–, F55, T40.8, T40.9, O99.3–[2]	(M), (1), (2), (W), (X), (Y) or (9)
		Underlying cause of death ICD-10-CA codes	**Multiple cause of death ICD-10-CA codes**
CVSD	Alcohol	F10.–, E24.4, G31.2, G62.1, G72.1, I42.6, K29.2, K70.–, K85.2, K86.0, O35.4–, Q86.0, R78.0, X45, X65, Y15, P04.3	--
		X40-X44, X60-X64, X85, Y10-Y14	At least one comorbidity is T51.–.
	Opioids	F11.–, P96.1	--
		X41-X44, X61-X64, X85, Y11-Y14	At least one comorbidity is T40.0 (Opium), T40.1 (Heroin), T40.2 (Natural and semisynthetic opioids), T40.3 (Methadone), T40.4 (Synthetic opioids, other than methadone), or T40.6 (Other and unspecified narcotics).
	Cannabis	F12.–	--
		X40-X44, X60-X64, X85, Y10-Y14	At least one comorbidity is T40.7.
	Other CNS depressants	F13.–	--
		X40-X44, X60-X64, X85, Y10-Y14	At least one comorbidity is T42.3, T42.4, T42.6, T42.7.
	Cocaine	F14.–	--
		X40-X44, X60-X64, X85, Y10-Y14	At least one comorbidity is T40.5
	Other CNS stimulants	F15.–	--
		X40-X44, X60-X64, X85 Y10-Y14	At least one comorbidity is T43.6.
	Unknown and multiple substances	F19.–, X41, X42, X61, X62, Y11, Y12, O35.5	--
		X40-X44, X60-X64, X85, Y10-Y14	At least one comorbidity is T43.8, T43.9.
	Other substances	F16.–, F18.–, F55	--
		X40-X44, X60-X64, X85, Y10-Y14	At least one comorbidity is T40.8, T40.9.

Notes: [1] Conditions listed in both columns should be met in order for an event to be considered a SUH event. For more information on the variables and their operationalization, please refer to Appendix 3 and do-file. [2] Include only if F10–16 or F18–19 as diagnosis type (3) is in the same abstract. [3] Include X41, X61 and Y11 if neither T42.– nor T43.– are in the same abstract; include X42, X62 and Y12 if T40.– is not in the same abstract. [4] Includes hallucinogens, solvents, and abuse of non-dependence-producing substances. DAD: discharge abstract database; CVSD: Canadian vital statistics death database; ICD: international classification of diseases; CNS: central nervous system.

Appendix 4. List of variables in the final derived 2006 CanCHEC cohort.

Name of variable in original dataset[1]	Final variable name	Variable label	Notes
uniqid	uniqid	unique ID (CanCHEC 2006)	Used in linking CanCHEC components based on Appendix A.
prcdda	da_06	dissemination area (CanCHEC 2006)	Used in linking CanCHEC components based on Appendix A.
hhnum	hhnum	key for household table (CanCHEC 2006)	Used in linking CanCHEC components based on Appendix A.
ppnum	ppnum	key for person table (CanCHEC 2006)	Used in linking CanCHEC components based on Appendix A.
pp_id	pp_id	alternative key for census person table (CanCHEC 2006)	Used in linking CanCHEC components based on Appendix A.
event_year	year_death	year when death occurred (CanCHEC 2006)	Used in linking CanCHEC components based on Appendix A.
placeofdeath_province	placeofdeath_province	province where death occurred (CanCHEC 2006)	Used in linking CanCHEC components based on Appendix A.
registration_number	registration_number	death registration number (CanCHEC 2006)	Used in linking CanCHEC components based on Appendix A.
canchecw2	w_cohort	CanCHEC 2006 weight (CanCHEC 2006)	CanCHEC 2006 weight.
dad_transaction_id	dad_transaction_id	DAD transaction ID (CanCHEC 2006)	Used in linking CanCHEC components based on Appendix A.
submitting_prov_code	submitting_prov_code	submitting province code (DAD)	Used in linking CanCHEC components based on Appendix A.
inst_code	inst_code	institution number (DAD)	Code assigned to a reporting facility by a provincial/ territorial ministry of health identifying the facility and the level of care of the data submitted in DAD.
fiscal_year	year_hosp	fiscal year of hospitalization (DAD)	The fiscal year of hospitalization, in DAD.
health_card_prov_code	hcard_pr	province issuing the health card (DAD)	Province issuing the health card of the patients, recorded in DAD.
patient_postal_code	resdn_pcode_dad	patient postal code (DAD)	Patient postal code recorded in DAD.
derived variable[2]	age_hosp	age at the time of hospitalization (DAD)	Age at the time of hospitalization in year, as recorded in DAD. This derived variable is based on the age_units and age_code variables in DAD.
gender_code	sex_hosp	sex of patient (DAD)	Sex of patiend recorded in the "gender_code" variable in DAD.
admission_date	admission_date	admission date (DAD)	Admission date in DAD.
discharge_date	discharge_date	discharge date (DAD)	Discharge date in DAD.
diag_cluster_`i'	diag_cluster_`i'	diagnosis cluster `i' (DAD)	This is a set of 25 variables, referring to the diagnosis cluster of an abstract in DAD. Diagnosis cluster is an alphabetic character which indicates association between diagnoses.
diag_prefix_`i'	diag_prefix_`i'	diagnosis prefix `i' (DAD)	This is a set of 25 variables, referring to the diagnosis prefix of an abstract in DAD. Diagnosis prefix is a one-character-long code which provides additional information on the ICD code to which it is assigned.
diag_code_`i'	diag_code_`i'	diagnosis code `i' (DAD)	This is a set of 25 variables, referring to the ICD-10 codes of the diagnosis for a patient record in DAD.
diag_type_`i'	diag_type_`i'	diagnosis type `i' (DAD)	This is a set of 25 variables, referring to the diagnosis type of an abstract in DAD. Diagnosis type is an alpha or numeric character used to describe the significance of a diagnosis or condition for a patient record in the DAD.
total_los_days	total_los_days	total length of stay (DAD)	This is the total length of stay as recorded in DAD.
acute_los_days	acute_los_days	acute length of stay (DAD)	This is the acute length of stay in hospital as recorded in DAD.
alc_los_days	alc_los_days	alternative level of care length of stay (DAD)	This is the alternative level of care length of stay in hospital, as recorded in DAD.
sex	sex_death	sex of deceased (CVSD)	This variable indicates the sex of deceased as available in CVSD.
derived variable	age_death	age of deceased (CVSD)	This is the age of deceased in year. This variable is derived form the age_value and age_code variables in CVSD.
residence_province_3digit	resdn_pr_cvsd	usual residence of deceased: province (CVSD)	This is the usual residence of deceased (province) available in CVSD.
residence_postalcode	resdn_pcode_cvsd	usual residence of deceased: postal code (CVSD)	This is the usual residence of deceased (postal code) available in CVSD.
death_cause_4digits	icd_underly_cause	ICD for underlying cause of death (CVSD)	This variable contains the ICD-10 codes of the underlying cause of death available in CVSD.

ra_mc_`i'	icd_contr_cause_`i'	ICD for contributing cause of death `i' (CVSD)	This is a set of 20 variables that refer to the ICD-10 codes of the contributing causes of death. These variables are available in the Multi Cause datasets associated with CVSD.
age	age_cens	age (census 2006)	Refers to the age at last birthday (as of the census reference date, May 16, 2006). This variable is derived from Date of birth.
cma	cma_cens	census metropolitan area or census agglomeration of current residence (2006)	Area consisting of one or more neighbouring municipalities situated around a major urban core. A census metropolitan area must have a total population of at least 100,000 of which 50,000 or more live in the urban core. A census agglomeration must have an urban core population of at least 10,000.
sex	sex_cens	sex (census 2006)	Refers to the gender of the respondent in Census 2006.
derived variable	hhatinc_adj	household after-tax income adjusted for household size (census 2006)	The after-tax income of a household is the sum of the after-tax incomes of all members of that household, adjusted for the household size. This variable is derived fromn the household after-tax income (hhinc_at) and household size (nunits) variables in Census 2006.
fsaname	fsa_cens	forward sortation area (census 2006)	The first three characters of the postal code identify the forward sortation area (FSA). FSAs are associated with a postal facility from which mail delivery originates. The average number of households served by an FSA is approximately 8,000, but the number can range from zero to more than 60,000 households.
pcd	cd_cens	census division of current residence (census 2006)	This is the census division of current residence in census 2006.
pcsd	csd_cens	census subdivision of current residence (census 2006)	This is the census subdivision of current residence in Census 2006.
pop	pop_csd	population size group of current census subdivision of residence (census 2006)	This is the population size group of current census subdivision of residence in Census 2006.
pr	pr_cens	province or territory of current residence (census 2006)	This is the province variable in Census 2006, referring to the province or territory of current residence of the respondent.
ruindfg	rufg	rural, urban classification (census 2006)	This is "ruindfg" variable in Census 2006, referring to the 2006 rural or urban classification of the CSD where the person usually resided on May 16, 2005, one year prior to Census Day.
rusize	rusize	rural/urban size code (census 2006)	This is the rural/urban size code variable in Census 2006.
compw2	w_cens	composite weight (perswt + occwtp -1) (census 2006)	This is the Composite Weight (PERSWT + OCCWTP-1) variable in Census (compw2).
derived variable	edu	education (census 2006)	This derived variable is based on "Highest certificate, diploma or degree" variable (hcdd) in Census 2006.
nocsbrd	occ	labour market activities : occupation broad categories (census 2006)	This is the Broad National Occupational Classification variable in Census 2006.
derived variable	race	ethnicity (census 2006)	This derived variable is based on the Visible minority variable (dvismin) in Census 2006.
derived variable	emp	employment status (census 2006)	This derived variable is based on the Labour force activity variable (lf71) in Census 2006.
derived variable	resdn_sk_dad_p	resident of Saskatchewan based on DAD	This variable indicates that if a cohort member was identified as a resident of Saskatchewan based on DAD.
derived variable	resdn_sk_cvsd_p	resident of Saskatchewan based on CVSD	This variable indicates that if a cohort member was identified as a resident of Saskatchewan based on CVSD.
derived variable	resdn_sk_cens_p	resident of Saskatchewan based on census 2006	This variable indicates that if a cohort member was identified as a resident of Saskatchewan based on census 2006.
derived variable	resdn_sk_p	person residing in SK at some point between 2006 and 2016	This variable indicates that if a cohort member was identified as a person residing in Saskatchewan at some point between 2006 and 2016, based on DAD or CVSD or Census 2006.
derived variable	hosp_e	hospitalization event	This variable indicates a hospitalization event due to any reason.
derived variable	hosp_p	hospitalized person	This variable indicates if a person was hospitalized at least once, between 2006 and 2016, due to any reason.
derived variable	death_e	death event	This variable indicates a death event.
derived variable	death_p	deceased person	This variable indicates if a person died between 2006 and 2016.
derived variable	hosp_alc_e	hospitalization event due to alcohol	This variable indicates a hospitalization event due to alcohol.

derived variable	hosp_opi_e	hospitalization event due to opioid	This variable indicates a hospitalization event due to opioid.
derived variable	hosp_can_e	hospitalization event due to cannabis	This variable indicates a hospitalization event due to cannabis.
derived variable	hosp_cnsdep_e	hospitalization event due to other CNS depressants	This variable indicates a hospitalization event due to other CNS depressants.
derived variable	hosp_coc_e	hospitalization event due to cocaine	This variable indicates a hospitalization event due to cocaine.
derived variable	hosp_cnsstim_e	hospitalization event due to other CNS stimulants	This variable indicates a hospitalization event due to other CNS stimulants.
derived variable	hosp_unkmult_e	hospitalization event due to unknown and multiple substances	This variable indicates a hospitalization event due to unknown and multiple substances.
derived variable	hosp_other_e	hospitalization event due to other substances	This variable indicates a hospitalization event due to other substances.
derived variable	hosp_su_e	hospitalization event due to substance use	This variable indicates a hospitalization event due to substance use.
derived variable	hosp_alc_p	person hospitalized due to alcohol	This variable indicates if a person was hospitalized at least once due to alcohol, between 2006 and 2016.
derived variable	hosp_opi_p	person hospitalized due to opioids	This variable indicates if a person was hospitalized at least once due to opioids, between 2006 and 2016.
derived variable	hosp_can_p	person hospitalized due to cannabis	This variable indicates if a person was hospitalized at least once due to cannabis, between 2006 and 2016.
derived variable	hosp_cnsdep_p	person hospitalized due to other CNS depressants	This variable indicates if a person was hospitalized at least once due to other CNS depressants, between 2006 and 2016.
derived variable	hosp_coc_p	person hospitalized due to cocaine	This variable indicates if a person was hospitalized at least once due to cocaine, between 2006 and 2016.
derived variable	hosp_cnsstim_p	person hospitalized due to other CNS stimulants	This variable indicates if a person was hospitalized at least once due to other CNS stimulants, between 2006 and 2016.
derived variable	hosp_unkmult_p	person hospitalized due to unknown and multiple substances	This variable indicates if a person was hospitalized at least once due to unknown and multiple substances, between 2006 and 2016.
derived variable	hosp_other_p	person hospitalized due to other substances	This variable indicates if a person was hospitalized at least once due to other substances, between 2006 and 2016.
derived variable	hosp_su_p	person hospitalized due to substance use	This variable indicates if a person was hospitalized at least once due to substance use, between 2006 and 2016.
derived variable	death_alc_e	death event due to alcohol	This variable indicates death event due to alcohol.
derived variable	death_opi_e	death event due to opioids	This variable indicates death event due to opioids.
derived variable	death_can_e	death event due to cannabis	This variable indicates death event due to cannabis.
derived variable	death_cnsdep_e	death event due to other CNS depressants	This variable indicates death event due to other CNS depressants.
derived variable	death_coc_e	death event due to cocaine	This variable indicates death event due to cocaine.
derived variable	death_cnsstim_e	death event due to other CNS stimulants	This variable indicates death event due to other CNS stimulants.
derived variable	death_unkmult_e	death event due to unknown and multiple substances	This variable indicates death event due to unknown and multiple substances.
derived variable	death_other_e	death event due to other substances	This variable indicates death event due to other substances.
derived variable	death_su_e	death event due to substance use	This variable indicates death event due to substance use.
derived variable	death_alc_p	person died due to alcohol	This vairable indicates a person who died due to alcohol, between 2006 and 2016.
derived variable	death_opi_p	person died due to opioids	This vairable indicates a person who died due to opioids, between 2006 and 2016.
derived variable	death_can_p	person died due to cannabis	This vairable indicates a person who died due to cannabis, between 2006 and 2016.
derived variable	death_cnsdep_p	person died due to other CNS depressants	This vairable indicates a person who died due to other CNS depressants, between 2006 and 2016.
derived variable	death_coc_p	person died due to cocaine	This vairable indicates a person who died due to cocaine, between 2006 and 2016.

derived variable	death_cnsstim_p	person died due to other CNS stimulants	This vairable indicates a person who died due to other CNS stimulants, between 2006 and 2016.
derived variable	death_unkmult_p	person died due to unknown and multiple substances	This vairable indicates a person who died due to unknown and multiple substances, between 2006 and 2016.
derived variable	death_other_p	person died due to other substances	This vairable indicates a person who died due to other substances, between 2006 and 2016.
derived variable	death_su_p	person died due to substance use	person died due to substance use
derived variable	pesuh	person who experienced substance use harm	Person who experienced substance use harm between 2006 and 2016. Substance use harm is defined as hospitalization or death due to substance use.

Notes: [1] All variables are listed in lower case, independet of how they are saved in the original dataset. [2] A *derived variable* is a variable that is generated or recoded using variables in CanCHEC 2006. DAD: discharge abstract database; CVSD: Canadian vital statistics death database; CanCHEC: Canadian census health and environment cohorts; CNS: central nervous system; CSD: census subdivision; ICD: international classification of diseases.

Appendix 5. Rate of people who experienced substance use harm per 100,000 in Saskatchewan weighted to CanCHEC 2006.

	Substance use harm category, rate per 100,000 (SE)							
Year	Alcohol	Opioids	Cannabis	Other CNS depressants	Cocaine	Other CNS stimulants	Unknown/ multiple substances	Other substances[1]
2006	149.2 (9.81)	19.5 (3.53)	21.0 (3.42)	20.0 (3.62)	13.0 (2.70)	7.0 (2.00)	33.5 (3.91)	3.0 (1.30)
2007	152.3 (9.58)	20.6 (3.38)	28.1 (4.13)	18.1 (3.40)	17.6 (3.59)	3.0 (1.02)	43.2 (5.10)	2.5 (1.00)
2008	174.1 (9.74)	17.2 (2.79)	24.3 (3.16)	20.2 (3.34)	21.3 (4.15)	7.0 (1.86)	45.5 (5.38)	1.0 (0.34)
2009	186.6 (10.56)	31.1 (4.25)	28.0 (3.63)	15.3 (2.80)	25.0 (4.38)	6.1 (1.92)	39.8 (5.13)	2.5 (0.98)
2010	173.7 (9.77)	31.9 (4.30)	27.2 (3.64)	20.0 (3.37)	18.0 (3.98)	4.6 (1.83)	33.9 (4.36)	2.6 (1.19)
2011	185.0 (10.84)	33.7 (4.40)	38.3 (5.11)	24.9 (4.11)	16.1 (3.37)	5.7 (1.67)	42.0 (4.99)	2.1 (1.16)
2012	193.9 (10.28)	33.5 (4.61)	31.9 (4.71)	24.0 (3.99)	15.7 (2.80)	4.7 (1.60)	42.9 (4.46)	2.6 (1.25)
2013	182.1 (10.34)	30.6 (4.24)	37.5 (4.54)	19.0 (3.29)	12.7 (2.61)	11.0 (2.85)	44.3 (4.81)	2.6 (1.23)
2014	207.7 (11.42)	42.6 (5.86)	52.2 (6.25)	23.4 (4.18)	21.3 (3.75)	17.4 (3.50)	44.2 (5.89)	2.1 (1.01)
2015	203.8 (11.84)	60.8 (6.61)	55.4 (5.65)	21.5 (3.76)	24.2 (4.17)	28.8 (4.30)	60.2 (6.53)	2.2 (1.01)
2016	230.9 (12.39)	61.9 (6.63)	95.6 (7.39)	33.1 (4.25)	30.4 (3.76)	44.1 (6.06)	47.3 (5.47)	2.2 (0.93)

Notes: [1] Other substances include hallucinogens, solvents, abuse of non-dependence-producing substances.

Abbreviations

BC	British Columbia
CanCHEC	Canadian Census Health and Environment Cohorts
CCR	Canadian Cancer Registry
CIHI	Canadian Institute for Health Information
CRA	Canada Revenue Agency
CSUCH	Canadian Substance Use Costs and Harms
CVSD	Canadian Vital Statistics Death Database
DAD	Discharge Abstract Database
ICD	International Classification of Diseases
MSP	Medical Services Plan
NACRS	National Ambulatory Care Reporting System
NAICS	North American Industry Classification System
NHS	National Household Survey
NOC	National Occupational Classification
NPESUH	People who have Not Experienced Substance Use Harm
PESUH	People who Experienced Substance Use Harm
RDC	Research Data Centre
REB	Research Ethics Board
SHRF	Saskatchewan Health Research Foundation
SK	Saskatchewan
SKY-RDC	Saskatchewan Research Data centre
SUH	Substance Use Harm

References

[1] Link BG, Phelan J. Social conditions as fundamental causes of disease. J Health Soc Behav 1995;Spec No:80–94.

[2] Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. J Health Soc Behav 2010;51 Suppl:S28–40.

[3] Brownson RC, Chriqui JF, Stamatakis KA. Understanding evidence-based public health policy. Am J Public Health 2009;99:1576–83.

[4] Anderson-Baron J, Karekezi K, Koziel J, McCurdy A. Saskatchewan Policy Analysis Case Report: Canadian Harm Reduction Policy Project. Canadian Research Initiative in Substance Misuse; 2017.

[5] Alsabbagh M, Cooke M, Elliott SJ, Chang F, Shah N-U-H, Ghobrial M, et al. Stepping up to the Canadian opioid crisis: a longitudinal analysis of the correlation between socioeconomic status and population rates of opioid-related mortality, hospitalization and emergency department visits (2000-2017). Health Promotion & Chronic Disease Prevention in Canada: Research, Policy & Practice 2022;42.

[6] Kerr T. Public health responses to the opioid crisis in North America. J Epidemiol Community Health 2019;73:377–8.

[7] Galea S, Vlahov D. Social determinants and the health of drug users: socioeconomic status, homelessness, and incarceration. Public Health Rep 2002;117 Suppl 1:S135–45.

[8] CIHI. Hospital stays for harm caused by substance use among youth age 10 to 24. Canadian Institute for Health Information; 2019.

[9] Anderson M, Revie CW, Quail JM, Wodchis W, de Oliveira C, Osman M, et al. The effect of socio-demographic factors on mental health and addiction high-cost use: a retrospective, population-based study in Saskatchewan. Can J Public Health 2018;109:810–20.

[10] Anderson M, Revie CW, Stryhn H, Neudorf C, Rosehart Y, Li W, et al. Defining “actionable” high- costhealth care use: results using the Canadian Institute for Health Information population grouping methodology. Int J Equity Health 2019;18:171.

[11] Wammes JJG, van der Wees PJ, Tanke MAC, Westert GP, Jeurissen PPT. Systematic review of high-cost patients’ characteristics and healthcare utilisation. BMJ Open 2018;8:e023113.

[12] Homayra F, Pearce LA, Wang L, Panagiotoglou D, Sambo TF, Smith N, et al. Cohort profile: The provincial substance use disorder cohort in British Columbia, Canada. Int J Epidemiol 2021;49:1776.

[13] Orisatoki R, Quail J, Osman M, Teare G, Schwandt M, Neudorf C. Concurrent Mental Health and Substance Use Disorders among Frequent Emergency Department Users in Saskatchewan, Canada. Canadian Journal of Addiction 2017;8:11.

[14] Government of Canada, Statistics Canada. Research Data Centres. Statistics Canada 2019. https://www.statcan.gc.ca/en/microdata/data-centres (accessed January 17, 2022).

[15] Tjepkema M, Christidis T, Bushnik T, Pinault L. Cohort profile: The Canadian Census Health and Environment Cohorts (CanCHECs). Health Rep 2019;30:18–26.

[16] Statistics Canada. Canadian Census Health and Environment Cohorts (CanCHECs). Statistics Canada 2019. https://www.statcan.gc.ca/en/microdata/data-centres/data/canchec (accessed September 28, 2023).

[17] Carrière G, Garner R, Sanmartin C. Social and economic characteristics of those experiencing hospitalizations due to opioid poisonings. Health Rep 2018;29:23–8.

[18] Petrie D, Allanson P, Gerdtham UG. Accounting for the dead in the longitudinal analysis of income-related health inequalities. J Health Econ 2011;30:1113–23.

[19] Vena JE, Sultz HA, Carlo GL, Fiedler RC, Barnes RE. Sources of bias in retrospective cohort mortality studies: a note on treatment of subjects lost to follow-up. J Occup Med 1987;29:256–61.

[20] Statistics Canada. 2006 Census of Population. Statistics Canada 2001. https://www12.statcan.gc.ca/census-recensement/2006/index-eng.cfm (accessed January 24, 2023).

[21] Statistics Canada. Canadian Vital Statistics - Death database (CVSD). Statistics Canada 2020. https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3233 (accessed

December 12, 2020).
[22] CIHI. Discharge Abstract Database metadata (DAD). Canadian Institute for Health Information 2021. https://www.cihi.ca/en/discharge-abstract-database-metadata-dad (accessed August 17, 2021).
[23] CIHI. National Ambulatory Care Reporting System metadata (NACRS). Canadian Institute for Health Information n.d. https://www.cihi.ca/en/national-ambulatory-care-reporting-system-metadata-nacrs (accessed September 6, 2022).
[24] World Health Organization (WHO). International Statistical Classification of Diseases and Related Health Problems, Tenth Revision. World Health Organization (WHO); 2010.
[25] CIHI. Hospital Stays for Harm Caused by Substance Use — Appendices to Indicator Library. Canadian Institute for Health Information; 2022.
[26] CIHI. Hospital Stays for Harm Caused by Substance Use [Indicator]. Canadian Institute for Health Information n.d. https://www.cihi.ca/en/indicators/hospital-stays-for-harm-caused-by-substance-use (accessed July 6, 2023).
[27] CIHI. Opioid-Related Harms in Canada. Canadian Institute for Health Information; 2018.
[28] Canadian Substance Use Costs and Harms Scientific Working Group. Canadian substance use costs and harms 2007–2020. (Prepared by the Canadian Institute for Substance Use Research and the Canadian Centre on Substance Use and Addiction.); 2023.
[29] Statistics Canada. 2006 Census : Visible Minority Population and Population Group Reference Guide. Statistics Canada 2008. https://www12.statcan.gc.ca/census-recensement/2006/ref/rp-guides/visible_minority-minorites_visibles-eng.cfm (accessed December 14, 2023).
[30] CIHI. Guidance on the Use of Standards for Race-based and Indigenous Identity Data Collection and Health Reporting in Canada. Canadian Institute for Health Information; 2022.
[31] Government of Canada. Canadian Cannabis Survey 2022: Summary. Government of Canada 2022. https://www.canada.ca/en/health-canada/services/drugs-medication/cannabis/research-data/canadian-cannabis-survey-2022-summary.html (accessed June 24, 2024).
[32] Zuckermann AME, Williams G, Battista K, de Groh M, Jiang Y, Leatherdale ST. Trends of poly-substance use among Canadian youth. Addict Behav Rep 2019;10:100189.
[33] CIHI. Unintended consequences covid-19: Impacts on harms caused by substance use. Canadian Institute for Health Information; 2021.
[34] Carrière G, Sanmartin C, Garner R. Understanding the socioeconomic profile of people who experienced opioid overdoses in British Columbia, 2014 to 2016. Health Reports 2021.
[35] Imbeau LM, Landry R, Milner H, Petry F, Crete J, Forest P-G, et al. Comparative Provincial Policy Analysis: A Research Agenda. Can J Polit Sci 2000;4:779–804.
[36] Statistics Canada. Guide to the Census of Population, 2021, Chapter 12 – Sampling and weighting for the long form. Statistics Canada 2022. https://www12.statcan.gc.ca/census-recensement/2021/ref/98-304/2021001/chap12-eng.cfm (accessed October 10, 2023).
[37] Butler A, Croxford R, Bodkin C, Akbari H, Bayoumi AM, Bondy SJ, et al. Burden of opioid toxicity death in the fentanyl-dominant era for people who experience incarceration in Ontario, Canada, 2015-2020: a whole population retrospective cohort study. BMJ Open 2023;13:e071867.
[38] Palis H, Zhao B, Young P, Korchinski M, Greiner L, Nicholls T, et al. Stimulant use disorder diagnosis and opioid agonist treatment dispensation following release from prison: a cohort study. Subst Abuse Treat Prev Policy 2022;17:77.
[39] McLeod KE, Timler K, Korchinski M, Young P, Milkovich T, McBride C, et al. Supporting people leaving prisons during COVID-19: perspectives from peer health mentors. Int J Prison Health 2021;17:206–16.
[40] Pijl EM, Bourque S, Martens M, Cherniwchan A. Take-Home Naloxone Kit Distribution: A Pilot Project Involving People Who Use Drugs and Who Are Newly Released from a Correctional Facility. Canadian Journal of Criminology and Criminal Justice 2017;59:559–71.
[41] Groot E, Kouyoumdjian FG, Kiefer L, Madadi P, Gross J, Prevost B, et al. Drug Toxicity Deaths

after Release from Incarceration in Ontario, 2006-2013: Review of Coroner's Cases. PLoS One 2016;11:e0157512.
[42] Waddell EN, Baker R, Hartung DM, Hildebran CJ, Nguyen T, Collins DM, et al. Reducing overdose after release from incarceration (ROAR): study protocol for an intervention to reduce risk of fatal and non-fatal opioid overdose among women after release from prison. Health Justice 2020;8:18.
[43] Cooper JA, Onyeka I, Cardwell C, Paterson E, Kirk R, O'Reilly D, et al. Record linkage studies of drug-related deaths among adults who were released from prison to the community: a scoping review. BMC Public Health 2023;23:826.
[44] Binswanger IA, Nowels C, Corsi KF, Glanz J, Long J, Booth RE, et al. Return to drug use and overdose after release from prison: a qualitative study of risk and protective factors. Addict Sci Clin Pract 2012;7:3.
[45] Binswanger IA, Stern MF, Deyo RA, Heagerty PJ, Cheadle A, Elmore JG, et al. Release from prison--a high risk of death for former inmates. N Engl J Med 2007;356:157–65.
[46] Joudrey PJ, Khan MR, Wang EA, Scheidell JD, Edelman EJ, McInnes DK, et al. A conceptual model for understanding post-release opioid-related overdose risk. Addict Sci Clin Pract 2019;14:17.
[47] Abayateye F, Fornssler B, Feng C, D'Arcy C, Alphonsus K. Prescription drug misuse among adults in Canada: a scoping review. J Subst Use 2023;28:305–14.
[48] Polcin DL, Mericle AA, Braucht GS, Wittman FD. Moving Social Model Recovery Forward: Recent Research on Sober Living Houses. Alcohol Treat Q 2023;41:173–86.
[49] Bobak TJ, Arjmand N, O'Brien C, Islam M, Jason LA. A Pilot Program Focusing on Perceptions and Acceptance of an Intervention on Medication-Assisted Treatments for Recovery Home Residents. Alcohol Treat Q 2023;41:386–93.
[50] McQuaid RJ, Malik A, Moussouni K, Baydack N, Stargardter M, Morrisey M. Life in Recovery from Addiction in Canada. Canadian Centre on Substance Use and Addiction; 2017.
[51] Quayum S, Hunter P, Rivier J, Cooper I, Baker N. Report on addiction, substance use and homelessness. Government of Canada; 2022.
[52] Urbanoski KA. Need for equity in treatment of substance use among Indigenous people in Canada. CMAJ 2017;189:E1350–1.
[53] Winder FS. Working Together for Change: A 10-Year Mental Health and Addictions Action Plan for Saskatchewan. 2014.
[54] Mental Health and Addictions Action Plan. Government of Saskatchewan n.d. https://www.saskatchewan.ca/government/health-care-administration-and-provider-resources/saskatchewan-health-initiatives/mental-health-and-addictions-action-plan (accessed September 21, 2023).