Development of the Canadian Substance Use Exposure Database (CanSUED): Modeling the prevalence of substance use in Canadian Jurisdictions, 2006 to 2017

Jinhui Zhao (zhaoj@uvic.ca)
University of Victoria https://orcid.org/0000-0003-2166-5370
Tim Stockwell
CISUR, University of Victoria
Justin Sorge
CISUR, University of Victoria
Adam Sherk
CISUR, University of Victoria
John Dorocicz
CISUR, University of Victoria
Jill Fairbank
Canadian Centre on Substance Use and Addiction
Bridget Hall
Canadian Centre on Substance Use and Addiction
Pam Kent
Canadian Centre on Substance Use and Addiction
Scott Macdonald
CISUR, University of Victoria
Sarah Wallingford
Canadian Centre on Substance Use and Addiction
Matthew Young
Canadian Centre on Substance Use and Addiction

Research

Keywords: Alcohol, drug, survey, mixed model, empirical best linear unbiased prediction

Posted Date: February 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-15494/v1

Abstract

Background

Small area and regional estimates of substance use (SU) exposures are increasingly needed to support estimation of the burden of SU-attributable morbidity and mortality. There is also a need to assess SU prevalence for subgroups by place, time and sociodemographic characteristics to plan the efficient delivery of treatment and harm reduction services. However, the data available from national surveys are often insufficient to produce reliable estimates for subgroups because of small sample sizes. There are also often missing estimates for some jurisdictions and some years when no surveys were conducted. We describe new methods which utilize Canadian national survey data of SU, sales, SU attributable hospitalisations and demographic data to develop the Canadian Substance Use Exposure Database (CanSUED). Estimates from this database have been used in the study of Canadian Substance Use Costs and Harm (CSUCH).

Methods

Exposures were estimated for eight substance categories: alcohol, tobacco, opioids, cannabis, cocaine, other central nervous system (CNS) stimulants, other CNS depressants and other psychoactive substances. The design-based direct estimates of SU were based on the Canadian Alcohol and Drug Monitoring Survey (CADUMS) in 2008-2012, the Canadian Tobacco, Alcohol and Drug Survey (CTADS) in 2013, 2015 and 2017, per capita alcohol and tobacco sales, and rates of wholly SU-attributable hospitalisations for all Canadian jurisdictions by age and gender. Multilevel models were used to model the design-based estimates of SU to produce reliable estimates for subgroups when the coefficient of variation (CV) of the estimates were > 33.3% and to predict SU exposure in ten provinces in 2006, 2007, 2014 and 2016 and in the three territories for 2006-2017 by using empirical best linear unbiased prediction (EBLUP).

Results

Direct design-based estimates were based on the surveys from a total of 107,750 Canadians aged 15+. The analyses produced reliable estimates of SU prevalence by year-province-gender-age using mixed models with the EBLUP method. Correlational analyses show that the model-based estimates were significantly related to the design-based estimates produced from both the CADUMS/CTADS and Canadian Community Health Survey. The new model estimates indicate increases in binge drinking, cannabis use, other CNS depressant substance use and cocaine use between 2006 and 2017. Rates of use of opioids and tobacco showed declines. Rates of use of other substances were relatively stable or did not show overall change across the whole time period.

Conclusion

The mixed model-adjusted approaches produced reliable estimates for small areas and age-gender groups and help fill gaps caused by data suppression in local and national surveys. We suggest that these methods provide the most comprehensive and reliable estimates available of Canadian substance use by substance category, year, jurisdiction, age and gender. The methods could also be applied in other countries where similar data are available.

Background

Small area and domain estimates of substance use (SU) exposure are needed to support estimation of the burden of SU-attributable harms such as SU-attributable morbidity and mortality. There is also a need to assess SU prevalence for sub-groups defined by place, time and sociodemographic characteristics, to plan the delivery of prevention, treatment and harm reduction services efficiently. However, sampling methods used in national surveys such as the Canadian Alcohol and Drug Use Monitoring Survey (CADUMS) and the Canadian Tobacco, Alcohol and Drug Survey (CTADS) were designed to produce maximum precision of estimates when reporting at the provincial and national level. The estimates directly produced from such national surveys may be unreliable for population subgroups, sub-regions of smaller size than the provincial level and for when rates of use of a particular substance are low [1]. The available surveys have not always been conducted annually and have not included the three territories in Canada. The aim of this study is thus to present a methodology that overcomes these limitations in available survey estimates by using small area estimation methods supplemented with additional jurisdictions-specific data on SU-attributable hospitalisations, alcohol and tobacco sales and demographic characteristics. In essence, these methods estimate patterns in the data by region, year, age, gender and type of SU and, with large and relevant auxiliary data sets, extrapolate these patterns to create more reliable estimates, especially where data are sparse. This exercise was conducted in order to develop the Canadian Substance Use Exposure Database (CanSUED), an open access database containing information on exposure and prevalence to eight categories of substance use by six-age-sex groups in 13 Canadian provinces and territories from year 2006 to 2017 and that were used in the assessment of Canadian Substance Use Costs and Harm (CSUCH; www.csuch.ca) [2] to overcome problems with missing or suppressed data exposure and prevalence estimates for eight categories of SU alcohol, tobacco, cannabis, opioids, other central nervous system (CNS) depressants (e.g., benzodiazepines, barbiturates), cocaine, other CNS stimulants (e.g., amphetamine, methamphetamine, ecstasy) and other substances (e.g., hallucinogens, inhalants) by age and sex for 2006 to 2017 and for each of Canada's 13 provincial and territorial jurisdictions.

Methods

CADUMS and CTADS each employ a complex sample design, with stratification, multiple stages of selection and unequal probabilities of selection. Using data from complex surveys such as these can present problems as the survey design and selection probabilities could affect estimates of exposure and variance, particularly for small areas and subgroups when samples are small. Statistics Canada apply reporting restrictions on survey-based estimates of prevalence which result in the suppression of estimates with a sample size below 30 or with very high CVs (above 33.3%). Secondly, CADUMS was conducted in ten provinces in 2008–2012 and the CTADS was conducted in ten provinces in 2013, 2015 and 2017. Missing are exposure estimates of SU

prevalence for the provinces in 2006, 2007, 2014 and 2016. Thirdly, there is a need to assess SU of particular populations such as those in the three Territories and also age and sex subgroups regionally in order to plan the delivery of treatment and harm reduction services efficiently.

There are several approaches to small area estimation (SAE) that have been developed and used to produce estimates when reliable estimates cannot be obtained directly from surveys for any of the above reasons [3, 4]. One approach is the composite estimator called empirical best linear unbiased prediction (EBLUP). EBLUP has been used to combine cross-sectional and time-series data [4].

To develop our estimates, we first used the method of direct estimates to estimate SU for six age-sex groups in each of ten provinces for survey years adjusted for survey design effects [5]. Using multilevel models [6] make direct estimates of SU prevalence (i.e., mean annual alcohol consumption, tobacco sales, wholly SU attributable hospitalisations, prevalence of SU and relevant auxiliary data to predict the estimates using the empirical best linear unbiased prediction (EBLUP) approach [3]. These estimates were produced by age-sex groups in provinces/territories for each year between 2006 and 2017. All the estimates were broken down by sex and age groups (15–34, 35–64 and 65+) for ten provinces, three territories and the whole of Canada from 2006 to 2017.

Data sources

Analyses were based on two national surveys: the Canadian Alcohol and Drug Use Monitoring Survey (CADUMS), which was conducted annually from 2008 to 2012 by Health Canada [7–11], and the Canadian Tobacco Alcohol and Drugs Survey (CTADS), which replaced the CADUMS in 2013 and was conducted in 2013, 2015 and 2017 by Statistics Canada [12–14]. The CTADS is conducted every two years. Details on the surveys can be found elsewhere [7–14].

Several auxiliary data sources were used to produce reliable estimates for province-age-sex groups when no reliable estimates can be produced directly from the surveys, or when no survey data are available (i.e., 2006, 2007, 2014 and 2016). These auxiliary data include age-sex population counts, per capita alcohol and tobacco sales data, and counts of wholly SU-attributable hospitalizations. Age-sex population data in provinces/territories over years [15] as well as per capita alcohol consumption data for the provinces/territories across the years 2006–2017 [16] were obtained from Statistics Canada. Annual tobacco sales data by province/territory were obtained from Health Canada [17]. The data of wholly SU (alcohol, cannabis, opioids, CNS depressant, CNS stimulant, cocaine and any other psychoactive substances)-attributable hospitalizations covered 10 provinces and three territories from 2006 to 2017 where obtained from the Canadian Institute for Health Information. Our analyses showed that official provincial per capita alcohol and tobacco consumption rates for the years 2006 to 2017 were significantly correlated with the survey-based estimates of alcohol and tobacco use produced by CADUMS/CTADS (Tables A1, A2 and A3 in Appendix A). A US study conducted also found that per capita sales and per capita self-reported consumption were highly correlated across individual states [18].

Survey sampling and population coverage

The CADUMS was a yearly survey on alcohol and other SU among Canadians initiated in April 2008 by the Controlled Substances and Tobacco Directorate, Health Canada [7–11]. The survey was derived from the Canadian Addiction Survey administered in 2004 and contained questions on substance use (including prescription drug misuse) and associated harms [19]. From 2013, the same SU questions were carried forward into the CTADS [12, 13]. Both the CADUMS and CTADS used random digit dialing to obtain a stratified sample across all 10 provinces with equal representation of subjects each month and based on a two-stage (telephone household, respondent) random sample stratified by province. The surveys used random-digit dialing (RDD) methods via Computer Assisted Telephone Interviewing (CATI). The sampling approach was designed to produce maximum precision of estimates when reporting at the provincial level by sex and the national level by sex and major age groups.

The sampling frame was based on an electronic inventory of all active telephone area codes and exchanges in Canada. Within each of the 10 provincial strata, a random sample of telephone numbers was selected with equal probability in the first stage of selection (i.e., households). Within selected households, one respondent aged 15 years or older who could complete the interview in English or French was chosen. The person who would celebrate his/her birthday next within the household was asked to complete the interview. The surveys covered the population aged 15 years and older in ten provinces and excludes residents of the Yukon, the Northwest Territories and Nunavut, permanent residents of institutions, people living in households without a telephone and people with cell phones only. Some provinces purchased additional cases in some years. The sample size was 16,674 in 2008, 13,082 in 2009, 13,615 in 2010, 10,076 in 2011, 11,090 in 2012, 14,565 in 2013, 15,154 in 2015 and 16,349 in 2017. Each sample represented approximately 26,000,000 Canadians aged 15 years and older. Details of sample sizes for provinces each survey year can be found in Table A4 in Appendix A.

Measures of substance use

The CADUMS and CTADS core content included self-report questions concerning general health and well-being, smoking status, alcohol use and harms, pharmaceutical use, cannabis use and harms, other illicit SU (opioids, cocaine, other CNS stimulants and depressants and harms, alcohol and cannabis and driving, pregnancy and SU, and demographics. The questions on SU are presented in Table A5 in Appendix A. Specific indicators analyzed in this study are described below. These exposure estimates of SU were needed to help estimate the number of SU attributable conditions in the CSUCH the study [2].

Alcohol consumption

Measures of alcohol consumption included in the CADUMS and CTADS and used in CanSUED were prevalence of lifetime abstainers, former drinkers and current drinkers in the population aged 15+, percentage of binge drinkers among current year drinkers aged 15+, and annual litres of ethanol consumed in

the population aged 15+. Lifetime abstainers were defined as those who have consumed no alcohol or less than one standard drink (SD) of alcohol in their lifetime (one SD = 13.6 g or 17.05 ml in Canada). Former drinkers are defined as those who have consumed alcohol within their lifetime but who have not consumed at least one SD of alcohol within the past year. Current drinkers are defined as those who have consumed at least one SD of alcohol in the past year. The quantity and frequency (QF) method [20] was used to estimate total annual litres of alcohol consumption for current drinkers.

Tobacco smoking

Measures of tobacco smoking included the prevalence of lifetime non-smokers, former smokers and current smokers. Lifetime non-smokers were those who smoked less than 100 cigarettes in their lifetime. Former smokers were those who smoked at least 100 cigarettes but did not smoke daily or occasionally. Current smokers are those who smoked daily or occasionally when they were surveyed.

Other substance use

The use of cannabis, opioids (illicit or prescribed pain relievers), other CNS depressants (sedatives, tranquilizers), cocaine, other CNS stimulants (amphetamine, methamphetamines, ecstasy and any other stimulants) and other psychoactive substances (hallucinogens, inhalants, etc.) in the past year was assessed. In addition, some SU-related conditions are causally associated with injection drug use (IDU) and an additional analysis was carried out regarding IDU, which was restricted to SU types with injection as a possible route of administration (opioids, cocaine, other CNS stimulants). The proportions of those reporting use of these substances among those aged 15 years and older in the past 12 months were estimated.

Analytical strategy to estimate substance use exposures

We developed a statistical model to estimate trends and patterns observed across all the available survey data sets so as to allow reliable estimates of suppressed or otherwise missing data. In our analyses, an estimate with a CV of greater than 33.3% was considered unreliable and was modelled using the methods described below. Specifically, we did so by using auxiliary information and borrowing strength from (1) data collected in neighbouring areas (2) data collected at other times (3) exploiting spatial correlation in the data across regions (4) exploiting the temporal correlation of the target variable in each area to indirect estimates of SU prevalence. Indirect estimators borrow strength from other area and/or time periods to increase effective sample size. These indirect estimates were based on implicit or explicit models that provides a link to related areas and/or time periods through supplementary information such as recent census counts or current administrative records related to the variable of interest.

Direct estimates

Direct estimates of self-reported SU were obtained from the surveys with adjustment for design effects due to strata, clustering and disproportionate selection of subjects in the surveys [5]. Direct estimates were based on the CADUMS and CTADS surveys of 107,750 Canadians aged 15+ in the provinces in 2008–2017. There were a total of 600 direct estimates by six age-sex groups in ten provinces in 2008–2017. The estimates in 2014 were produced based on the pooled 2013 and 2015 CTADS and the estimates in 2016 were produced based on the pooled 2015 and 2017 CTADS. Equations for the estimates of per capita alcohol consumption and corresponding standard errors produced directly from the surveys are presented in Box I.

Prevalence of SU, estimated for computing SU attributable hospitalization and death includes proportions $\hat{P}$ of lifetime non-smokers, former smokers and current smokers, lifetime non-drinkers, former drinkers and current drinkers, and other SUs in past year among the population aged 15, and proportion of binge drinkers among current drinkers. The equations for these direct estimates can be found in Box II.

Model-based empirical best linear unbiased prediction

Using the empirical best linear unbiased prediction (EBLUP) method [21, 22], we computed the estimates of per capita alcohol consumption and the SU prevalence for age-sex groups in ten provinces and three territories from 2006 to 2017. These computations were performed for cases in which the design-based direct estimates are unreliable due to small sample sizes or years when no surveys were conducted. The equations of the model-based estimates are presented in Box III.

The standard error of a model-based EBLUP estimate is the square root of the variance $\nabla\hat{m}$-m of the EBLUP estimate and the coefficient of variation of the EBLUP estimate can be computed using the standard error of the EBLUP estimate divided by the EBLUP estimate.

The fixed effect estimates β can be obtained from the mixed models [6, 23]. The mixed model is written as $y = X\beta+Z\gamma+\varepsilon$

where

- y denotes the vector of observed y_i's
- X is the known matrix of x_{ij} ' and the values of explanatory variables x_{ij} can be either regression-type continuous variables or dummy variables indicating class membership
- β is the unknown fixed effects parameter vector

- Z the known design matrix
- γ is the vector of unknown random-effects parameters
- ε is the unobserved vector of independent and identified distributed normal (Gaussian) random variables with mean 0 and variance σ^2.

Statistical analyses were completed using SAS 9.3 [24]. Direct estimates of mean alcohol consumption were produced using the SAS SURVEYMEANS procedure and percentages of substance users and non-users were estimated using the SAS SURVEYFREQ procedure because these procedures analyze sample survey data taking into account the sample design effects [24]. Direct estimates were conducted by age, sex, province/territory and year. The SAS MIXED procedure estimates the fixed-effects parameters and further produce the EBLUP estimates. The SAS MIXED procedure was used to perform multilevel regression of the direct estimates in which province/territories and year are considered as random effects and auxiliary data such as year-province-age-sex population, rates of wholly SU attributable conditions available by age and gender for all 13 jurisdictions by year, annual per capita cigarettes data and litres of alcohol of official sales data at province level as covariates fixed effects [3, 25]. Using the EBLUP method [3] predicts the estimates for all six age-sex groups by years in ten provinces and three territories in 2006–2017.

Validity assessments

We conducted several internal validity checks of the model-based EBLUP estimates. First, we compared the EBLUP estimates against the CADUMS/CTADS design-based direct survey estimates of per capita alcohol consumption and the prevalence of SUs for age-sex groups by provinces and years where there were reliable estimates, i.e., CVs <33.3%. We further compared the EBLUP estimates with the prevalence estimates from the Canadian Community Health Survey (CCHS) where the CVs of the estimates by age-sex groups were smaller than 33.3%. The CCHS conducted by Statistics Canada has a large sample size (a total of 984,911 Canadians were surveyed in 2005–2014) but only provided equivalent questions for some key alcohol and tobacco indicators for the Yukon, the Northwest Territories and Nunavut. More details on the CCHSs can be found elsewhere [12, 13, 26, 27]. Bivariate correlation was used to assess the relationship between the EBLUP estimates and the direct estimates; we estimated the Pearson correlation coefficient for each pair of the estimates.

Results

SU varied in Canada over years. The model-based and direct-survey estimates of SU show similar trends between 2006 and 2017 (Figures 1 and 2 and Tables A6 and A7 in Appendix). Linear regression was used to examine the change in the estimates over years. The percentages of former drinkers, current drinkers and binge drinkers significantly increased by 0.03, 0.14 and 0.45 percent points per year on average respectively. The percentages of former smokers and current smokers significantly decreased by 0.06 and 0.61 percent points per year on average respectively. While the percentage of any other SU use (legal or illegal drugs) tended to decrease from 33.86% in 2006 to 30.92% in 2017, the percentages of cannabis users and other CNS depressant users significantly increased by 0.37 and 0.31 percent points (t-test P = 0.0108 and < 0.0001). The percentage of opioids users significantly decreased by 0.95 percentage point per year (t-test P < 0.0001).

Nationwide and province-level estimates of substance use

The CADUMS and CTADS can produce relatively reliable estimates of alcohol, tobacco and other substance at the provincial and national level. Table 1 presents provincial and national mean of SU prevalence estimates. There are 8 nationwide estimates of each SU but no IDU (cannot be estimated or no users reported); all nationwide estimates have CV values of less than 33.3% except IDU use, indicating the estimates were reliable. There were no reliable estimates of IDU at the provincial and national level, i.e., the CVs are larger than 33.3%. Some province-level estimates of cocaine, other CNS stimulants and any other drug use have CVs larger than 33.3%: 43 among 100 cocaine use estimates and 76 among 100 province-level estimates of other CNS stimulant use and 56 among 99 any other psychoactive substance use estimates (no subjects reported use of any other substance in New Brunswick province in 2012). Means of the estimates at the provincial and national levels were similar.

National estimates of substance use by sex-age groups

The study produced direct estimates of alcohol, tobacco and other SU indicators by sex-age groups at national level and the means of the direct estimates are presented in Tables A8 and A9 in Appendix A. All the estimates of alcohol and tobacco use are with the CVs less than 33.3% except one estimate for females aged 15–34 (CV = 35.2%). Some estimates of SU such as cannabis, cocaine, CNS stimulants and any other psychoactive substances have CVs larger than 33.3%, indicating unreliable estimates, even though the estimates were performed at a national level due to small sample sizes and rates of use. For example, there are five estimates of cocaine use prevalence for males aged 65+ but all the estimates are with the CVs of larger than 33.3%.

Model-based EBLUP estimates of substance use by sex-age groups at the provincial level

The analyses on CADUMS and CTADS produced 100 direct estimates for each type of SU measures for each age-sex group in 10 provinces from 2008 to 2017, i.e. 600 direct estimates. The direct estimates in 2014 were based on the combined 2013 CTADS with 2015 CTADS and the direct estimates in 2016 were based on the combined 2015 CTADS with 2017 CTADS. Using mixed models with the EBLUP method produced the 936 estimates for six age-sex groups in ten provinces and three territories from 2006 to 2017.

We summarized means of the direct estimates and EBLUP estimates in Tables 2–5. There are 572 direct estimates of per capita alcohol consumption (the amount of alcohol use in the past year reported by current drinkers divided by the survey sample) in CADUMS and CTADS with the CV estimates being smaller than 33.3% (Table 2). The Pearson correlation coefficient (CC) is 0.97 and its P-value is 0.0001 indicating a highly significant correlation of direct estimates with EBLUP estimates of per capita alcohol consumption. The same analysis on 538 direct estimates with the CVs of small than 33.3% in CCHS and model-based EBLUP estimates also produced a significant correlation of CCHS direct estimates with the EBLUP estimates (CC = 0.68 and P<0.0001). Table 3 presents means of the prevalence estimates of lifetime nondrinking, former drinking, current drinking and binge drinking. In all cases, means were close. Correlation analyses of the direct estimates with CVs <33.3% show that the EBLUP estimates are significantly correlated with the direct estimates in either CADUMS/CTADS (Pearson CCs range from 0.86 to 0.90) or CCHS (Pearson CCs range from 0.91 to 0.93).

Table 4 presents the mean of the prevalence estimates for lifetime nonsmoking, former smoking and current smoking. Means of the EBLUP estimates and direct estimates particularly the percentages of current smokers are close. Correlation analyses suggest that the EBLUP are significantly correlated to the direct estimates in CADUMS/CTADS (Pearson CCs range from 0.85 to 0.93) and CCHS (Pearson CCs range from 0.90 to 0.99).

Table 5 presents the means of the prevalence estimates of other substance use exclusive of alcohol and tobacco which include past year use of any prescribed or illicit drugs, cannabis (marijuana, hashish, hash oil or other cannabis derivatives), other CNS depressants (sedatives and tranquilizers), cocaine, opioids (prescribed or non-prescribed), other CNS stimulants (amphetamines, methamphetamines, ecstasy and other stimulants such as Ritalin, Concerta, Adderall and Dexedrine), IDU and any other psychoactive substance such as hallucinogens and inhalants. No direct estimates using CCHS data were done because only the data on some types of substances were collected in two provinces, SU was surveyed in the samples of two provinces. A few subjects in CADUMS/CTADS reported use of cocaine, other CNS stimulants, IDU and any other psychoactive substance; therefore, the direct estimates of these SUs for age-sex groups were either unreliable or missing. The direct estimates of prevalence of using any prescribed drug or illicit substance were reliable; none of these estimates had CVs larger than 33.3%. Furthermore, the means of the EBLUP and of the direct estimates of any prescribed or illicit substance use were both very close and highly correlated (Pearson CC: 0.79 and P<0.0001). Fewer Canadians aged 65 years or older reported cannabis use in the past year (mean of direct estimates is 2.42% for men and 1.41% for women) but the EBLUP estimates suggest more Canadians aged 65+ used cannabis (mean of the EBLUP estimates is 3.01% for men and 2.79% for women). The EBLUP estimates were highly correlated with the direct estimates of cannabis use in past 12 months (Pearson CC = 0.92 and P<0.0001). Some direct estimates of the prevalence of other CNS depressant use were unreliable, i.e., they had CVs larger than 33.3%. However, the correlation analyses on other CNS depressant use indicated the EBLUP estimates are highly correlated to the direct estimates with the CVs larger than 33.3% (Pearson CC = 0.84 and P<0.0001). Most direct estimates of opioids use (prescribed opioids and heroin) are reliable; the EBLUP and direct estimates of opioids use were very similar. Prevalence of opioid use from the EBLUP estimates were highly correlated with the direct estimates (Pearson CC = 0.71 and P<0.0001).

Additionally, we performed mixed regression analyses of direct estimates and model-based estimates by each type of SU in which the province was treated as a random variable, age and sex were treated as categorical variables and year was treated as a continuous variable to examine the differences in age-sex-province-year of the model-based estimates of SU are consistent with the direct estimates. There were significant differences in the prevalence of past 12 month use for nearly all types of SU, across age and sex groups and variations across provinces and years (Table A10 in Appendix A) but the direct-survey and model-based estimates show very consistent results.

Discussion

Lack of reliable and comprehensive information on SU by age-sex subgroups, region and year has been a persistent problem for policy makers and researchers attempting to describe SU trends, estimate comparative SU harms and costs, plan serve delivery and evaluate impacts of prevention strategies. Small area estimation that borrows too much strength from larger areas or groups is likely to result in diminished local variances. We produced model-based estimates by incorporating consistent and reliable auxiliary variables from Health Canada, the Canadian Institute for Health Information and Statistics Canada at the provincial and territorial levels to supplement this lack of information.

In this study, we used the model-based EBLUP method to produce estimate of SU exposure by six sex-age groups for each of 13 provinces and territories from 2006 to 2017 based on two Canadian national surveys on alcohol, tobacco, cannabis, opioids, cocaine and other categories of SU. The EBLUP method is used because the sample sizes are small for sex-age groups in survey years, no surveys were conducted in years 2006–2007, 2014 and 2016, and the national surveys did not cover the population in three territories. We were able to demonstrate that incorporating available and relevant auxiliary datasets such as rates of SU specific hospitalisation, alcohol and tobacco sales data and population demographics resulted in more reliable and comprehensive estimates across year end jurisdiction than are currently available only from available self-report surveys. Specifically, our methods generated reliable estimates of levels of substance use across all major categories of legal and illegal SU covering years missed in national survey data, jurisdictions not usually covered (i.e. the three territories) and types of substance where prevalence of use is low.

Our results show that the EBLUP estimator had smaller estimated means square error compared to the design-based direct estimator. Our analyses show that the model-based estimates had smaller ranges than design-based direct estimates from CADUMS/CTADS and CCHS and tend to smooth out the group variations in SU estimates in year-province-sex-age groups. This is to be expected since small area statistical models generalize population characteristics and always tend to smooth the final prediction of population outcomes and underestimate the true ranges of EBLUP estimates.

Our validation analysis showed that the model-based EBLUP estimates demonstrated both high internal consistency with the design-based direct estimates with the CVs of less than 33.3% and high consistency with reliable external design-based direct estimates. The main validation results empirically confirmed that the model-based EBLUP approach could provide reliable and sensible estimates of SU indicators in population using the nationwide province-based SU monitoring surveys.

The CCHS is a cross-sectional survey that collects information related to health status, health care utilization and health determinants for Canadians. It surveys a larger sample of respondents and is designed to provide reliable estimates at the health region level under provinces. Our analyses show that more reliable estimates (mainly alcohol and tobacco use), i.e., more estimates with the CVs of less than 33.3% at the year-province-sex-age groups were produced from CCHS than that produced from CADUMS/CTADS. Correlation analyses also show that the model-based estimates of alcohol and tobacco exposures were significantly correlated with the design-based direct estimates from CCHS.

The methodology detailed in this study was developed for the comprehensive estimation of SU costs and harms in the context of Canada for the years 2007 to 2017 [28]. Reliable SU prevalence and exposure information is critical to any project estimating healthcare and economic loss of production costs caused by population SU. In turn, SU harms and costs provide information for policymakers and program officials to develop effective policies and strategies to help various populations live healthy, active and rewarding lives.

Several limitations must be noted in the estimates of SU. There are limited auxiliary data particularly for illicit and 'other' SU categories. For 'other' SU we first modeled the direct estimates from the self-report surveys and produced the model-based estimates by age and gender groups in each provinces/territories from 2006 to 2017 which were used as covariates for the model-based estimates of specific SU. This might have affected the reliability of any SU estimates, especially where data are sparse or missing.

Conclusion

We have employed these methodologies to create a comprehensive Canadian Substance Use Exposure Database (CanSUED) by three age and two sex groups covering all 13 provinces and territories for 2006 to 2017. We recommend CanSUED for use by researchers and policymakers whose goal is to better understand SU in Canada, to recognize trends and patterns in SU or to estimate the harms and costs caused by SU in society. The analyses presented show that the application of statistical techniques designed to deal with sparse and missing data through the inclusion of relevant auxillary data result in more robust estimates of SU prevalence.

Declarations

Ethics approval and consent to participate

The estimates of substance use in this study involve only secondary analysis of existing survey (public files) and administrative data obtained from Statistics Canada and Health Canada. All datasets are publicly available and human subjects are anonymous.

Consent for publication

Not applicable.

Availability of data and materials

The public use microdata files (PUMFS) of 2008-2012 CADUMS and 2013, 2015 and 2017 CTADS are available at the Research Data Centre (RDC) from any of Canadian universities. The alcohol and tobacco sales data and population data are available by accessing to the websites of Health Canada and Statistics Canada (see the list of references).

Competing interests

The authors declare that they have no competing interests.

Funding

The estimates of substance use in this study were conducted in the project of "Canadian Substance Use Costs and Harms (CSUCH)". The research was supported by a research grant from the Health Canada, Substance Use and Addictions Program to the Canadian Centre on Substance Use and Addiction.

Authors' contributions

All the authors contributed to the design of the study. JZ, TS and JS carried out the data work and performed the data analysis. JZ, TS and AS drafted the manuscript. All authors contributed to the interpretation of the results and revising the manuscript for intellectual content. All authors read and approved the final version of the manuscript.

Acknowledgements

Not applicable.

References

1. Zhao J, Martin G, Macdonald S, Stockwell T: **The prevalence of alcohol and illicit drug use in British Columbia**. In. Victoria: CARBC. Available from URL: http://www.carbc.ca/Portals/0/AOD/GeneralPopulationSurveys/CADUMSReport0809.pdf. Accessible 16 March 2012; 2011.

2. Canadian Substance Use Costs and Harms Scientific Working Group: **Estimating Canadian substance use costs and harms** In. Ottawa, ON: the Canadian Institute for Substance Use Research and the Canadian Centre on Substance Use and Addiction; 2018.
3. Rao JNK, Molina I: **Small area estimation**, 2 edn. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2015.
4. Singh AC, Mantel HJ, Thomas BW: **Time series EBLUPs for small areas using survey data**. *Survey Methodology* 1994, **20**(1):11.
5. Lewis TH: **Complex survey data analysis with SAS** Boca Raton: CRC Press; 2017.
6. Brown H, Prescott R: **Applied Mixed Models in Medicine**. Hoboken, NJ: John Wiley & Sons Ltd.; 2006.
7. Health Canada: **Canadian Alcohol and Drug Use Monitoring Survey 2008: Microdata User Guide**. In. Ottawa: Health Canada. Available from URL: http://prod.library.utoronto.ca/datalib/codebooks/cstdli/cadums/2008/cadums-technical-guide-2008-final-eng.pdf. Accessible 14 December 2010; 2009.
8. Health Canada: **Canadian alcohol and drug use monitoring survey 2009: microdata user guide**. In. Ottawa: Health Canada. Available from URL: http://datalib.chass.utoronto.ca/codebooks/cstdli/cadums09/cadums%20final%20technical%20guide%202009%20-en.pdf. Accessible 29 March 2011; 2010.
9. Health Canada: **Canadian alcohol and drug use monitoring survey 2010: microdata user guide**. In. Ottawa: Health Canada. Available from URL: odesi.ca/documentation/CADUMS/2010/cadums_2010_gid_eng.pdf. Accessible 18 June 2013.; 2011.
10. Health Canada: **Canadian Alcohol and Drug Use Monitoring Survey 2011: microdata user guide**. In. Ottawa: Health Canada; 2012.
11. Health Canada: **Canadian alcohol and drug use monitoring survey 2012: microdata user guide**. In. Ottawa: Health Canada. Available from URL: http://odesi1.scholarsportal.info/documentation/CADUMS/2012/CADUMS_User_Guide_2012_ENG.pdf. Accessible 3 February 2018; 2013.
12. Statistics Canada: **Canadian Tobacco, Alcohol and Drugs Survey (CTADS), Annual 2013 - User Guide**. In. Ottawa: Statistics Canada; 2013.
13. Statistics Canada: **Canadian Tobacco, Alcohol and Drugs Survey (CTADS), Annual 2015 - User Guide**. In. Ottawa: Statistics Canada; 2015.
14. Statistics Canada: **Canadian Tobacco, Alcohol and Drugs Survey (CTADS), Annual 2017 - User Guide**. In. Ottawa: Statistics Canada; 2017.
15. Statistics Canada: **Table 051-0001 Estimates of population, by age group and sex for July 1, Canada, provinces and territories annual**. In. Edited by Canada S. Ottawa: Statistics Canada; 2017.
16. Statistics Canada: **Table 183-0023 - Sales and per capita sales of alcoholic beverages by liquor authorities and other retail outlets, by value, volume, and absolute volume, annual**. In. Edited by Canada S. Ottawa: Statistics Canada; 2017.
17. Health Canada: **Wholesale sales data: cigarette and fine-cut sales, province and territory, 1980-2014**. In., January 12, 2016 edn. Ottawa: Health Canada; 2016.
18. Smith PF, Remington PL, Williamson DF, Anda RF: **A comparison of alcohol sales data with survey data on self-reported alcohol use in 21 states**. *Am J Public Health* 1990, **80**(3):309-312.
19. Canadian Centre on Substance Abuse: **Canadian Addiction Survey 2004: Microdata eGuide**. In. Ottawa: CCSA; 2005.
20. Stockwell T, Chikritzhs T: **International guide for monitoring alcohol consumption and related harm**. In. Geneva: World Health Organization. Available from URL: http://apps.who.int/iris/bitstream/10665/66529/1/WHO_MSD_MSB_00.4.pdf. Accessible 13 December 2016; 2000.
21. Henderson CR: **Applications of linear models in animal breeding**. Guelph, ON: University of Guelph; 1994.
22. Harville DA: **BLUP (best linear unbiased prediction), and beyond**. In: *Advances in statistical methods for genetic improvement of livestock. Volume Advanced series in agricultural sciences*, edn. Edited by Hammond DGaK. Berlin: Springer-Verlag; 1990: 239-276.
23. Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O: **SAS for mixed models, second edition**, Second edition edn. Cary, NC: SAS Institute Inc.; 2006.
24. SAS Institute: **SAS/STAT 9.3 user's guide**. Cary, NC: SAS Institute Inc. ; 2011.
25. Little RC, Stroup WW, Freund RJ: **SAS for linear models**, 4 edn. Cary, NC: SAS Institute Inc.; 2002.
26. Statistics Canada: **Canadian community health survey (CCHS) annual component user guide 2014 and 2013-2014 microdata files**. In. Ottawa: Statistics Canada; 2015.
27. Statistics Canada: **Canadian community health survey (CCHS) annual component user guide 2012 and 2012-2012 microdata files**. In. Ottawa: Statistics Canada; 2013.
28. Canadian Substance Use Costs and Harms Scientific Working Group: **Canadian substance use costs and harms (2007-2014). (Prepared by the Canadian Institute for Substance Use Research and the Canadian Centre on Substance Use and Addiction.)**. In. Ottawa, ON: Canadian Centre on Substance Use and Addiction; 2018.

Table

Table 1. Mean estimates of substance use prevalence at the provincial level and national level for 2008-2017 from the CADUMS and CTADS

Substance use indicators	Provincial level						National level					
	N †	Mean	SD	Min	Max	n ‡	N †	Mean	SD	Min	Max	n ‡
Alcohol indicators												
Per capita alcohol in litres	100	2.78	0.33	2.11	3.68	100	10	2.87	0.18	2.63	3.21	10
Lifetime non-drinker (%)	100	10.24	1.84	6.43	14.39	100	10	10.26	0.92	8.76	11.71	10
Former drinker (%)	100	14.44	2.46	9.39	22.75	100	10	14.25	0.67	13.25	15.38	10
Current drinker (%)	100	75.32	2.99	68.15	82.58	100	10	75.49	0.70	74.61	76.85	10
Binge drinker (%) ††	100	29.66	4.49	20.23	41.89	100	10	29.20	2.97	25.31	32.81	10
Smoking indicators												
Lifetime non-smoker (%)	100	54.99	4.63	46.41	64.82	100	10	55.40	1.58	53.03	57.72	10
Former smoker (%)	100	27.53	2.85	21.06	32.41	100	10	27.41	0.63	26.59	28.27	10
Current smoker (%)	100	17.49	3.36	10.17	25.06	100	10	17.19	1.84	14.65	19.90	10
Other SU indicators ‡‡												
Cannabis (%)	100	10.89	2.55	6.81	23.03	100	10	11.02	1.57	9.00	14.16	10
Opioids (%)	100	16.15	3.84	6.49	23.75	100	10	15.97	3.53	11.82	21.66	10
Other CNS depressants (%)	100	10.49	2.50	6.49	20.18	100	10	10.10	0.90	8.68	11.47	10
Cocaine (%)	100	1.12	0.60	0.21	3.93	43	10	1.28	0.52	0.68	2.41	10
Other CNS stimulants (%)	100	1.98	0.93	0.62	5.62	76	10	2.11	0.66	1.30	3.40	10
Any other psychoactive substances (%)	99	1.13	0.76	0.08	3.98	56	10	1.13	0.49	0.65	2.30	10
IDU (%)	21	0.13	0.11	0.02	0.55	0	9	0.02	0.02	0.00	0.07	0

Note: SU: Substance use, CNS: Central nervous system, IDU: Injection drug use, and SD: Standard deviation. † N: 10 estimates in each of ten provinces and Canada from 2008 to 2017. ‡ n: Number of estimates with CV < 33.3% among N provincial estimates and N national estimates. †† % of binge drinkers among current drinkers. ‡‡ % past year users among Canadians aged 15+.

Table 2. Mean of direct and model-based estimates of per capita alcohol consumption in litres by six sex-age groups in provinces and territories in 2006-2017

Sex-age	CADUMS/CTADS direct				CADUM/CTADS model				CCHS direct				Pearson †	Pearson ††
	N	Mean	SD	n ‡	N	Mean	SD	n ‡	N	Mean	SD	n ‡	CC‡‡, P	CC‡‡, P
Male														
15-34	100	4.24	1.12	96	156	4.42	1.32	156	90	3.41	1.54	90	0.90, 0.0001	0.17, 0.1088
35-64	100	3.95	0.73	100	156	4.12	1.16	156	90	3.15	1.39	90	0.83, 0.0001	0.29, 0.0051
65+	100	3.37	1.08	93	156	3.56	1.20	156	90	2.31	1.14	90	0.94, 0.0001	0.42, 0.0001
Female														
15-34	100	1.83	0.78	93	156	1.93	0.50	156	90	1.47	0.60	88	0.93, 0.0001	0.28, 0.0079
35-64	100	1.73	0.41	100	156	1.81	0.48	156	90	1.42	0.56	90	0.91, 0.0001	0.46, 0.0001
65+	100	1.38	0.69	90	156	1.41	0.58	156	90	0.86	0.45	90	0.96, 0.0001	0.46, 0.0001
All estimates	600	2.75	1.41	572	936	2.88	1.52	936	540	2.10	1.40	538	0.97, 0.0001	0.68, 0.0001

Note: † Pearson correlation analysis includes direct estimates in CADUMS/CTADS with CV < 33.3%. †† Pearson correlation analysis includes direct estimates in CCHS with CV < 33.3%. ‡ Number of estimates with CV < 33.3% among N estimates. ‡‡ Correlation coefficient.

Table 3. Mean of direct and model-based estimates of percentages of lifetime non-drinkers, former drinkers, current drinkers and binge drinkers by six sex-age groups in provinces and territories in 2006-2017

Sex	Age	CADUMS/CTADS direct				CADUM/CTADS model				CCHS direct *				Pearson †	Pearson ††
		N	Mean	SD	n ‡	N	Mean	SD	n ‡	N	Mean	SD	n ‡	CC‡‡, P	CC‡‡, P
Lifetime non-drinker															
Male	15-34	100	12.89	3.86	78	156	12.98	1.52	156	90	11.02	2.50	89	0.59, 0.0001	0.45, 0.0001
	35-64	100	5.42	2.36	59	156	5.73	0.88	156	90	5.20	0.81	84	0.50, 0.0001	0.61, 0.0001
	65+	99	7.30	3.21	30	156	8.58	2.55	156	90	8.71	2.41	84	0.71, 0.0001	0.81, 0.0001
Female	15-34	100	15.10	4.40	92	156	15.11	1.97	156	90	12.23	3.18	90	0.68, 0.0001	0.75, 0.0001
	35-64	100	8.38	3.05	93	156	8.24	1.42	156	90	7.95	1.95	89	0.66, 0.0001	0.87, 0.0001
	65+	100	17.38	7.71	87	156	16.75	3.46	156	90	17.53	4.04	90	0.82, 0.0001	0.92, 0.0001
All estimates		599	11.09	6.20	439	936	11.23	4.51	936	540	10.44	4.71	526	0.86, 0.0001	0.91, 0.0001
Former drinker															
Male	15-34	100	7.20	2.76	49	156	7.16	0.92	156	90	5.77	1.36	85	0.49, 0.0003	0.45, 0.0001
	35-64	100	13.10	3.26	98	156	12.85	1.50	156	90	11.00	1.83	90	0.64, 0.0001	0.65, 0.0001
	65+	100	22.53	5.72	93	156	23.03	3.12	156	90	20.28	4.72	90	0.68, 0.0001	0.73, 0.0001
Female	15-34	100	10.16	3.21	77	156	10.40	1.21	156	90	8.10	1.08	90	0.32, 0.0036	0.14, 0.1808
	35-64	100	15.54	3.94	100	156	15.88	1.94	156	90	14.16	1.93	90	0.63, 0.0001	0.54, 0.0001
	65+	100	24.93	5.54	97	156	24.59	2.56	156	90	24.18	3.04	90	0.68, 0.0001	0.71, 0.0001
All estimates		600	15.58	7.62	514	936	15.65	6.67	936	540	13.92	7.03	535	0.89, 0.0001	0.95, 0.0001
Current drinker															
Male	15-34	100	79.91	5.28	100	156	79.87	1.95	156	90	83.09	3.77	90	0.52, 0.0001	0.36, 0.0005
	35-64	100	81.48	4.08	100	156	81.42	1.76	156	90	83.81	2.15	90	0.52, 0.0001	0.45, 0.0001
	65+	100	70.24	6.41	100	156	68.39	3.81	156	90	71.17	6.33	90	0.72, 0.0001	0.70, 0.0001
Female	15-34	100	74.75	5.76	100	156	74.49	2.02	156	90	79.60	3.84	90	0.53, 0.0002	0.32, 0.0018
	35-64	100	76.07	4.71	100	156	75.88	2.06	156	90	77.91	3.23	90	0.62, 0.0001	0.47, 0.0001
	65+	100	57.68	9.14	100	156	58.66	4.44	156	90	58.75	5.87	90	0.82, 0.0001	0.79, 0.0001
All estimates		600	73.35	9.97	600	936	73.12	8.21	936	540	75.64	9.90	540	0.88, 0.0001	0.92, 0.0001
Binge drinker															
Male	15-34	100	47.94	8.95	100	156	48.33	3.83	156	90	48.22	5.74	90	0.58, 0.0001	0.63, 0.0001
	35-64	100	34.20	7.79	100	156	34.82	4.55	156	90	32.11	6.61	90	0.75, 0.0001	0.72, 0.0001
	65+	100	20.27	7.19	80	156	20.47	3.32	156	90	13.35	4.73	83	0.61, 0.0001	0.54, 0.0001

Female	15-34	100	33.66	6.28	100	156	33.74	2.08	156	90	21.56	6.88	87	0.38, 0.0001	0.70, 0.0001
	35-64	100	19.78	4.72	100	156	19.86	2.11	156	90	8.77	4.66	80	0.67, 0.0001	0.67, 0.0001
	65+	99	9.07	5.17	32	156	9.14	2.00	156	88	2.09	1.86	32	0.22, 0.2120	0.83, 0.0001
All estimates		599	27.52	14.32	512	936	27.73	13.11	936	538	21.32	16.30	462	0.90, 0.0001	0.93, 0.0001

Note: * Estimates for combined lifetime non-drinkers and former drinkers in CCHS. † Pearson correlation analysis includes direct estimates in CADUMS/CTADS with CV < 33.3%. †† Pearson correlation analysis includes direct estimates in CCHS with CV < 33.3%. ‡ Number of estimates with CV < 33.3%. ‡‡ Correlation coefficient.

****

Table 4. Mean of direct and model-based estimates of percentages of lifetime non-smokers, former smokers, and current smokers by six sex-age groups in provinces and territories in 2006-2017

Sex	Age	CADUMS/CTADS direct				CADUM/CTADS model				CCHS direct *				Pearson †	Pearson ††
		N	Mean	SD	n ‡	N	Mean	SD	n ‡	N	Mean	SD	n ‡	CC‡‡, P	CC‡‡, P
Lifetime non-smoker															
Male	15-34	100	67.45	5.86	100	156	63.64	4.02	156	90	59.36	4.67	90	0.63, 0.0001	0.56, 0.0001
	35-64	100	46.04	6.19	100	156	44.19	3.58	156	90	39.27	4.93	90	0.80, 0.0001	0.73, 0.0001
	65+	100	33.25	7.32	98	156	34.44	3.50	156	90	29.80	5.67	90	0.60, 0.0001	0.73, 0.0001
Female	15-34	100	73.90	6.67	100	156	75.29	4.22	156	90	65.35	5.03	90	0.72, 0.0001	0.77, 0.0001
	35-64	100	52.32	5.86	100	156	52.69	3.41	156	90	48.64	6.30	90	0.81, 0.0001	0.76, 0.0001
	65+	100	54.34	5.25	100	156	50.54	2.54	156	90	55.82	4.29	90	0.28, 0.0048	0.27, 0.0088
All estimates		600	54.55	14.75	598	936	53.47	13.63	936	540	49.71	13.19	540	0.93, 0.0001	0.90, 0.0001
Former smoker															
Male	15-34	100	10.42	3.08	58	156	13.99	1.34	156	90	11.65	2.58	84	0.19, 0.1989	0.23, 0.0370
	35-64	100	32.96	4.63	100	156	35.03	1.91	156	90	33.78	3.55	90	0.57, 0.0001	0.70, 0.0001
	65+	100	55.99	5.85	100	156	54.12	2.83	156	90	58.35	4.89	90	0.43, 0.0001	0.72, 0.0001
Female	15-34	100	10.33	3.11	85	156	7.19	1.40	156	90	12.64	2.07	89	0.18, 0.0829	0.45, 0.0001
	35-64	100	29.78	3.65	100	156	28.79	1.74	156	90	29.48	3.90	90	0.70, 0.0001	0.74, 0.0001
	65+	100	36.35	5.03	100	156	40.65	1.93	156	90	34.68	4.10	90	0.21, 0.0355	0.24, 0.0191
All estimates		600	29.30	16.51	543	936	29.96	15.93	936	540	30.10	16.12	533	0.94, 0.0001	0.99, 0.0001
Current smoker															
Male	15-34	100	22.13	5.34	99	156	22.37	3.18	156	90	28.98	4.27	90	0.65, 0.0001	0.48, 0.0001
	35-64	100	21.01	4.81	100	156	20.77	2.84	156	90	26.95	3.77	90	0.67, 0.0001	0.41, 0.0001
	65+	100	10.76	3.99	65	156	11.44	1.96	156	90	11.85	2.73	82	0.65, 0.0001	0.23, 0.0300
Female	15-34	100	15.77	4.86	94	156	17.52	3.36	156	90	21.01	3.78	90	0.80, 0.0001	0.70, 0.0001
	35-64	100	17.90	4.07	99	156	18.52	2.91	156	90	21.88	3.76	90	0.75, 0.0001	0.68, 0.0001
	65+	99	9.41	3.28	80	156	8.80	1.73	156	90	9.50	2.30	85	0.47, 0.0001	0.54, 0.0001
All estimates		599	16.17	6.52	537	936	16.57	5.59	936	540	20.20	8.03	527	0.85, 0.0001	0.90, 0.0001

Note: * Estimates for combined lifetime non-drinkers and former drinkers in CCHS. † Pearson correlation analysis includes direct estimates in CADUMS/CTADS with CV < 33.3%. †† Pearson correlation analysis includes direct estimates in CCHS with CV < 33.3%. ‡ Number of estimates with CV < 33.3%. ‡‡ Correlation coefficient.

Table 5. Mean of direct and model-based estimates of percentages of other SU users by six sex-age groups in provinces and territories in 2006-2017

Sex	Age	CADUMS/CTADS direct				CADUM/CTADS model				Pearson †	
		N	Mean	SD	n ‡	N	Mean	SD	n ‡	CC‡‡	P-value
Cannabis										0.92	0.0001
Male	15-34	100	26.37	6.02	100	156	25.90	2.82	156	0.66	0.0001
	35-64	100	10.67	3.66	100	156	10.46	2.12	156	0.69	0.0001
	65+	76	2.42	1.81	5	156	3.01	1.05	156	0.86	0.0553
Female	15-34	100	16.98	4.89	93	156	16.78	1.72	156	0.57	0.0001
	35-64	100	5.26	2.82	81	156	4.96	1.87	156	0.76	0.0001
	65+	55	1.41	1.70	1	156	2.79	0.66	156	n/a	
Opioids										0.71	0.0001
Male	15-34	100	13.66	4.77	89	156	15.49	3.59	156	0.71	0.0001
	35-64	100	16.37	4.26	97	156	17.32	3.70	156	0.69	0.0001
	65+	100	14.90	5.25	78	156	15.36	3.69	156	0.58	0.0001
Female	15-34	100	17.25	5.38	95	156	17.59	3.67	156	0.68	0.0001
	35-64	100	17.74	4.48	100	156	19.08	3.74	156	0.82	0.0001
	65+	100	15.05	4.92	88	156	16.70	3.85	156	0.71	0.0001
Other CNS depressants										0.84	0.0001
Male	15-34	100	5.14	2.07	29	156	5.06	1.10	156	0.28	0.1307
	35-64	100	8.36	2.40	90	156	8.26	0.98	156	0.62	0.0001
	65+	99	10.31	3.90	54	156	10.17	1.08	156	0.36	0.0060
Female	15-34	100	8.80	3.59	67	156	8.54	1.93	156	0.53	0.0001
	35-64	100	13.76	3.53	99	156	13.46	2.35	156	0.70	0.0001
	65+	100	18.20	6.04	92	156	17.51	3.83	156	0.68	0.0001
Cocaine										0.22	0.2050
Male	15-34	97	4.02	2.56	20	156	2.48	0.39	156	-0.17	0.4538
	35-64	84	0.91	0.75	1	156	2.42	0.46	156	n/a	
	65+	7	0.40	0.27	0	156	2.23	0.14	156	n/a	
Female	15-34	88	1.92	1.13	11	156	1.34	0.19	156	-0.10	0.7533
	35-64	53	0.41	0.44	1	156	1.28	0.17	156	n/a	
	65+	3	1.78	1.59	0	156	1.22	0.06	156	n/a	
Other CNS stimulants										0.98	0.0001
Male	15-34	100	6.06	3.61	50	156	5.18	2.72	156	0.97	0.0001
	35-64	93	1.04	0.79	2	156	2.07	0.89	156	n/a	
	65+	42	1.01	1.17	0	156	2.71	0.33	156	n/a	
Female	15-34	97	3.97	2.65	34	156	3.01	1.77	156	0.99	0.0001
	35-64	88	0.71	0.57	3	156	1.38	0.57	156	1.00	0.0001
	65+	35	0.60	0.37	0	156	0.98	0.52	156	n/a	
Any other psychoactive substance										1.00	0.0001
Male	15-34	88	3.79	2.88	27	156	2.81	0.65	156	1.00	0.0001
	35-64	53	0.92	0.68	0	156	2.46	0.49	156	n/a	
	65+	5	0.18	0.08	0	156	2.78	0.11	156	n/a	
Female	15-34	83	1.82	1.41	6	156	2.16	0.47	156	1.00	0.0001

	35-64	30	0.26	0.26	0	156	1.82	0.16	156	n/a
	65+	0	0.00	.	0	156	1.78	0.12	156	n/a
IDU										n/a
Male	15-34	7	0.96	0.93	0	156	0.81	0.39	156	n/a
	35-64	4	0.42	0.24	0	156	0.78	0.38	156	n/a
	65+	0	.	.	0	156	0.75	0.38	156	n/a
Female	15-34	6	0.67	0.60	0	156	0.76	0.39	156	n/a
	35-64	2	0.16	0.13	0	156	0.78	0.38	156	n/a
	65+	0	.	.	0	156	0.81	0.37	156	n/a

Note: † Pearson correlation analysis includes direct estimates in CADUMS/CTADS with CV < 33.3%. ‡ Number of estimates with CV < 33.3%. ‡‡ Correlation coefficient.

Figures

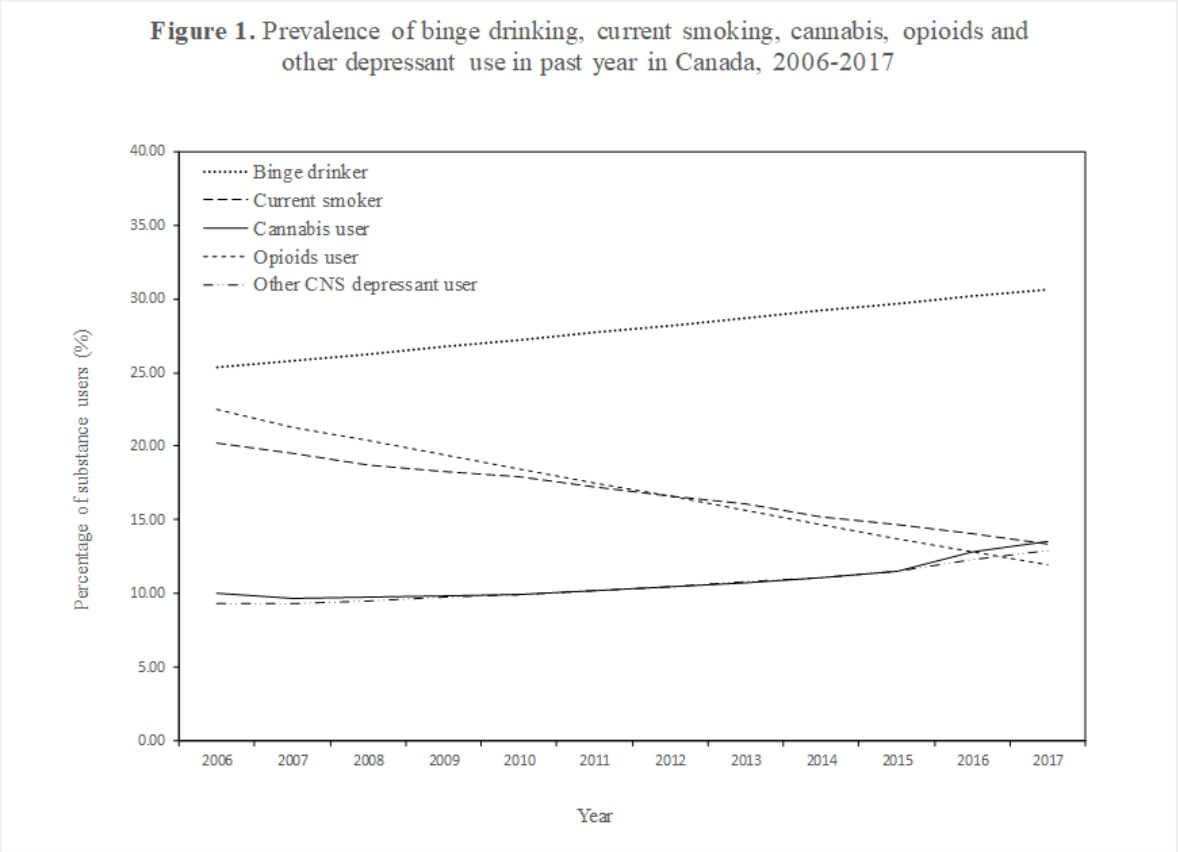

Figure 1

Prevalence of binge drinking, current smoking, cannabis, opioids and other depressant use in past year in Canada, 2006-2017

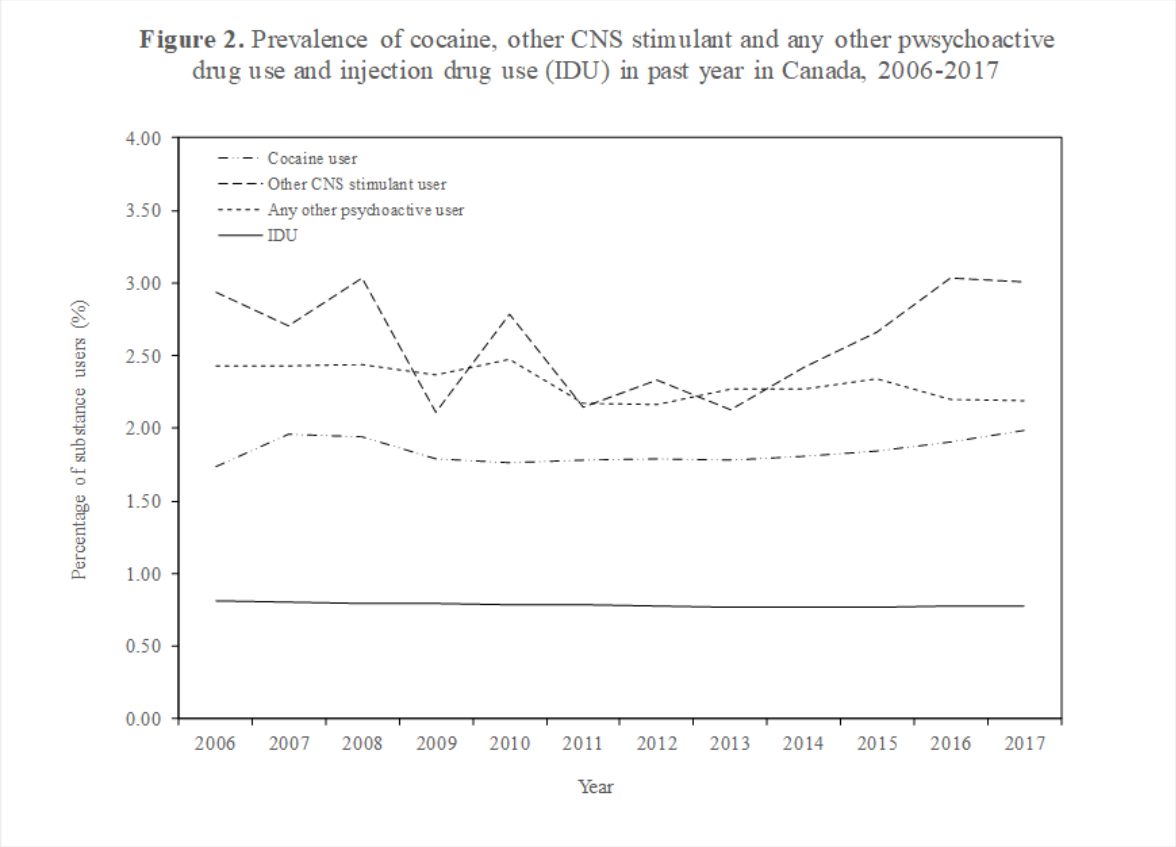

Figure 2

Prevalence of cocaine, other CNS stimulant and any other pwsychoactive drug use and injection drug use (IDU) in past year in Canada, 2006-2017

Box I. Equations of computing per capita alcohol consumption and its variance, standard error and coefficient of variation

Annual per capita alcohol consumption is computed using the equation below:

$$\hat{\bar{y}} = \left(\sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} y_{hij} \right) / \sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij}$$

The Taylor series linearization method is used to estimate the variance $\widehat{V}(\hat{\bar{y}})$ of the mean $\hat{\bar{y}}$ using the formula below

$$\hat{V}(\hat{\bar{y}}) = \frac{n_h(1-f_h)}{n_h - 1} \sum_{i=1}^{n_h} (e_{hi\cdot} - \hat{e}_{h\cdot\cdot})^2$$

$$e_{hi\cdot} = \sum_{j=1}^{n_h} w_{hij} \left(y_{hij} - \hat{\bar{y}} \right) / \sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij}$$

$$\hat{e}_{h\cdot\cdot} = \left(\sum_{i=1}^{n_h} e_{hi\cdot} \right) / n_h$$

where

- $h = 1, 2, \ldots, H$ is the stratum number, with a total of H strata
- $i = 1, 2, \ldots, n_h$ is the cluster index within stratum h
- $j = 1, 2, \ldots, m_{hi}$ is the unit index within cluster i of stratum h
- w_{hij} denotes the sampling weight for unit j in cluster i of stratum h
- $y_{hij} = \left(y_{hij}^{(1)}, y_{hij}^{(2)}, \ldots y_{hij}^{(p)} \right)$ are the observed values of the analysis variables for unit j in cluster i of stratum h , including both the values of numerical variables and the values of indicator variables for levels of categorical variables. $p = 1, 2, \ldots, p$ is the analysis variable number, with a total of P variables
- f_h is the sampling rate f_h for stratum h , with is used in Taylor series variance estimation, is the fraction of first-stage units (PSUs) selected for the sample.

The standard error $StdErr$ of the mean $\hat{\bar{y}}$ is the square root of the estimated variance $\hat{V}(\hat{\bar{y}})$ that can be computed using the equation as follows.

$$StdErr(\hat{\bar{y}}) = \sqrt{\hat{V}(\hat{\bar{y}})}$$

The coefficient of variation $cv(\hat{\bar{y}})$ are which is the ratio of the standard error of the mean $StdErr(\hat{\bar{y}})$ to the estimated mean $\hat{\bar{y}}$ can be computed using the equation as follows:

$$cv(\hat{\bar{y}}) = StdErr(\hat{\bar{y}}) / \hat{\bar{y}}$$

Figure 3

Box I. Equations of computing per capita alcohol consumption and its variance, standard error and coefficient of variation

Box II. Equations of computing the prevalence of substance uses, and its variance, standard error and coefficient of variation

Prevalence $\hat{\bar{P}}$ of SU are computed using the equation:

$$\hat{\bar{P}} = \left(\sum_{h=1}^{H}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} n_{hij}W_{hij}\right) / \left(\sum_{h=1}^{H}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} N_{hij}W_{hij}\right)$$

Taylor series linearization method is used to calculate the variance $\hat{V}\left(\hat{\bar{P}}\right)$ of the proportion estimate $\hat{\bar{P}}$ and that can be expressed as

$$\hat{V}\left(\hat{\bar{P}}\right) = \frac{n_h(1-f_h)}{n_h-1}\sum_{i=1}^{n_h}\left(g^{hi}-\bar{g}^{h}\right)^2$$

$$g^{hi} = \left(\sum_{j=1}^{m_{hi}}\left(n_{hij}-\hat{\bar{P}}N_{hij}\right)w_{hij}\right) / \left(\sum_{h=1}^{H}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}} N_{hij}w_{hij}\right)$$

$$\bar{g}^{h} = \left(\sum_{i=1}^{n_h} g^{hi}\right)/n_h$$

where

- $h = 1, 2, \ldots, H$ is the stratum number, with a total of H strata
- $i = 1, 2, \ldots, n_h$ is the cluster index within stratum h
- $j = 1, 2, \ldots, m_{hi}$ is the unit index within cluster i of stratum h
- n_{hij} is the number of non-users, former users or current users in each age-sex group
- N_{hij} is the sample size in each age-sex group
- W_{hij} denotes the sampling weight for unit j in cluster i of stratum h
- f_h is the sampling rate f_h for stratum h , with is used in Taylor series variance estimation, is the fraction of first-stage units (PSUs) selected for the sample.

The standard error $StdErr\left(\hat{\bar{P}}\right)$ of the proportion $\hat{\bar{P}}$ is the square root of the estimated variance $\hat{V}\left(\hat{\bar{P}}\right)$ and that can be computed using the equation as follows.

$$StdErr\left(\hat{\bar{P}}\right) = \sqrt{\hat{V}\left(\hat{\bar{P}}\right)}$$

The coefficient of variation $cv\left(\hat{\bar{P}}\right)$ are which is the ratio of the standard error of the mean $StdErr\left(\hat{\bar{P}}\right)$ to the estimated mean $\hat{\bar{P}}$ can be computed using the equation as follows:

$$cv\left(\hat{\bar{P}}\right) = StdErr\left(\hat{\bar{P}}\right)/\hat{\bar{P}}$$

Figure 4

Box II. Equations of computing the prevalence of substance uses, and its variance, standard error and coefficient of variation

Box III. Equations for computing EBLUP estimates and mixed model used to model the direct estimates

The formula of EBLUP estimates is

$$\hat{m} = X_m\hat{\beta} + \hat{C}_m\hat{V}^{-1}\left(y - X\hat{\beta}\right)$$

and the estimated prediction variance $\hat{V}_{(\hat{m}-m)}$ is as follows:

$$\hat{V}_{(\hat{m}-m)} = \hat{V}_m - \hat{C}_m\hat{V}^{-1}\hat{C}_m^T + \left[X - \hat{C}_m\hat{V}^{-1}X\right]\left(X^T\hat{V}^{-1}X\right)^{-}\left[X - \hat{C}_m\hat{V}^{-1}X\right]^T$$

where

- m represents a hypothetical realization of a missing data vector with associated design matrix X_m
- y is the observed dependent variable y
- X is the observed explanatory variables X
- $\hat{V}$ is the covariance matrix of y.
- $\hat{C}_m$ is the model-based covariance matrix between m and the observed data y
- $\hat{V}_m$ is the model-based variance matrix of m. Using hats (^) to indicates estimates
- T stands for transpose of a matrix
- $\hat{\beta} = \beta_1, \ldots, \beta_p$ are unknown fixed-effects parameters to be estimated in the mixed model.

The standard error of an EMBLUP estimate is the square root of the variance $\hat{V}_{(\hat{m}-m)}$ of the EMBLUP estimate and the coefficient of variation of the EMBLUP estimate can be computed using the standard error of the EMBLUP estimate divided by the EMBLUP estimate.

The mixed model is written as $y = X\beta + Z\gamma + \varepsilon$

where

- y denotes the vector of observed y_i's
- X is the known matrix of x_{ij} ' and the values of explanatory variables x_{ij} can be either regression-type continuous variables or dummy variables indicating class membership
- β is the unknown fixed effects parameter vector
- Z the known design matrix
- γ is the vector of unknown random-effects parameters
- ε is the unobserved vector of independent and identified distributed normal (Gaussian) random variables with mean 0 and variance σ^2.

Figure 5

Box III. Equations for computing EBLUP estimates and mixed model used to model the direct estimates

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Zhaoetal2020CanSUEDPHMadditionalfiles.docx