



Uncovering Polysubstance Use Patterns in Canadian Youth with Machine Learning on Longitudinal COMPASS Data

Yang Yang¹ · Zahid A. Butt¹ · Scott T. Leatherdale¹ · Helen H. Chen¹

Accepted: 16 August 2023
© The Author(s) 2023

Abstract

Understanding polysubstance use (PSU) patterns and their associated factors among youth is crucial for addressing the complex issue of substance use in this population. This study aims to investigate PSU patterns in a large sample of Canadian youth and explore associated factors using data from COMPASS, a longitudinal health survey of Canadian secondary school students. The study sample consisted of 8824 students from grades 9 and 10 at baseline in 2016/17, followed over 3 years until 2018/19. Leveraging machine learning methods, especially the least absolute shrinkage and selection operator (LASSO) and multivariate latent Markov models, we conducted a comprehensive examination of PSU patterns. Our analyses revealed distinct PSU patterns among Canadian youth, including no-use (C1), alcohol-only (C2), concurrent use of e-cigarettes and alcohol (C3), and poly-use (C4). C1 showed the highest prevalence (60.5%) in 2016/17, declining by 2.4 times over 3 years, while C3 became the dominant pattern (32.5%) in 2018/19. The prevalence of C3 and C4 increased by 2.3 and 4.4 times, respectively, indicating a growing trend of dual and multiple substance use. Risk factors associated with PSU patterns included truancy ($OR_{C2} = 1.67$, 95% CI [1.55, 1.79]; $OR_{C3} = 1.92$, 95% CI [1.80, 2.04]; $OR_{C4} = 2.79$, 95% CI [2.64, 2.94]), having more smoking friends, more weekly allowance, elevated BMI, being older, and attending schools unsupportive in quitting drugs/alcohol. In contrast, not gambling online ($OR_{C2} = 0.22$, 95% CI [-0.16, 0.58]; $OR_{C3} = 0.14$, 95% CI [-0.24, 0.52]; $OR_{C4} = 0.08$, 95% CI [-0.47, 0.63]), eating breakfast, residing in urban areas, and having higher school connectedness were protective factors against a higher-use pattern. This study provides insights for policymakers, educators, and health professionals to design targeted and evidence-based interventions, addressing youth substance use challenges through a comprehensive examination of PSU patterns and influential factors impacting substance use behaviors.

Keywords Polysubstance use · Risk factor · Canadian youth · COMPASS study · LASSO · Latent Markov model

Youth substance use surveillance and prevention efforts in North America (Government of Canada, (n.d.); Johnston et al., 2016) and global wise (Kraus et al., 2015; Ahumada et al., 2019) typically focus on single substance use, potentially overlooking important

user differences among adolescents (Zapert, Snow & Tebes, 2002). Recent evidence from a large sample of Canadian high school students revealed that in 2017/18, 18% reported using two or more substances, 16% reported using only one substance, and 61% reported no substance use in the past month (Zuckermann et al., 2020). Notably, the number of individuals using multiple substances has increased, possibly influenced by the rising popularity of e-cigarette use (Zuckermann et al., 2019) and the high prevalence of cannabis use among youth. With the increasing availability and legalization of non-medical cannabis products in Canada in October 2018 (Zuckermann et al., 2021a; Zuckermann et al., 2021b; Haines-Saah & Fischer, 2021), studies have demonstrated that ever use of cannabis among youth has steadily increased since legalization, going from 30.5% in 2016/17 to 32.4% in 2018/19 (Zuckermann et al., 2021a).

The growing importance of considering e-cigarettes in the discussion of general substance use patterns among youth is becoming evident (Zuckermann et al., 2020; Rothrock et al., 2020; Hughes et al., 2015; Thepthien et al., 2021; Mehra et al., 2019). Despite being a relatively new product, e-cigarettes have rapidly gained popularity among youth, potentially contributing to increased substance use in this population (Zuckermann et al., 2020; Zuckermann et al., 2019; Rothrock et al., 2020; Hughes et al., 2015; Thepthien et al., 2021; Mehra et al., 2019).

While previous research has identified different patterns of youth substance use, many studies have focused mainly on alcohol, cigarette, and cannabis use due to their widespread occurrence. For instance, a study involving Canadian youth between the ages of 12 and 18 in Victoria, British Columbia, identified three patterns: no or low use (63%), concurrent use of alcohol and cannabis (23%), and poly-use including cigarettes, alcohol, cannabis, and other illicit drugs (11%) (Merrin & Leadbeater, 2018). Recent research has highlighted polysubstance use (PSU) patterns involving concurrent and multiple uses of e-cigarettes along with other substances, underscoring the need to consider e-cigarettes in examining multiple substance use (Zuckermann et al., 2020; Rothrock et al., 2020; Hughes et al., 2015; Thepthien et al., 2021; Mehra et al., 2019).

Understanding the factors contributing to PSU patterns among youth is crucial for assessing health risks, identifying intervention opportunities, and evaluating existing policies and practices. Various approaches, including family and peer support, policies, and school interventions, can address the increasing trend of youth PSU. However, limited research exists on the factors influencing the initiation of PSU patterns among youth. Therefore, this study aims to investigate PSU patterns in a large longitudinal sample of Canadian youth and explore associated factors using a latent variable modeling approach. PSU, in this study, refers to the use of cigarettes, e-cigarettes, alcohol, and cannabis. By examining these factors, we aim to gain insights that can inform effective prevention and intervention strategies targeting youth PSU. For the purpose of this study, we specifically focus on a more restricted age range, targeting mid-adolescents in grades 9 and 10. Mid-adolescence is a critical period marked by increased risk-taking behaviors, making it particularly crucial to examine risk factors associated with substance use during this developmental stage (Government of Canada (n.d.); Hale & Viner, 2016).

Methods

Participants and Setting

The COMPASS study is a longitudinal health survey that collects self-reported student- and school-level data on youth health behaviors from a large, convenience sample of Canadian

secondary school students every year starting in 2012/13 (Leatherdale et al., 2014; Reel, Bredin & Leatherdale, 2018). The COMPASS questionnaire (Cq) includes demographic and personal information and students' responses to multiple choice questions about their physical characteristics, behaviors, attitudes towards health and wellness, and academic performance gathered annually (Leatherdale et al., 2014). Participating schools use a separate questionnaire to provide information about school policies and practices related to their students' health behaviors. Additionally, school-level socio-economic status, urbanity, and built environment are collected as supplementary community-level data. Students must actively assent to participate in the COMPASS study, and parental/guardian consent is required using active-information passive-consent protocols (Thompson-Haile & Leatherdale, 2013). The COMPASS study has been granted ethics clearance by the University of Waterloo Office of Research Ethics (ORE 30118).

Data Collection and Preprocessing

For this retrospective cohort study, we used 3 years of linked longitudinal data from the Cq (collected in 2016/17, 2017/18, and 2018/19) before the outbreak of the COVID-19 pandemic. This study included 9307 Canadian students in grades 9 and 10 (including students at secondary I and II in Quebec) in 2016/17 from 76 secondary schools in Alberta, British Columbia, Ontario, and Quebec and followed up through 2018/19. The guidelines of Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) were adhered to in this study.

To prepare the data for analysis, we performed several preprocessing steps, such as cleaning, linking, merging, and analyzing missing data. The flowchart of data preprocessing steps with detailed descriptions was published in another paper (Yang et al., 2022).

Of the 9307 linked samples, present analyses were limited to 8824 students with regular patterns of advancement from one grade to the next in each school year. In 2016/17, about half (49.4%) of the students were in grade 9, 33.4% were in grade 10, and 15.3% were in grades 7–8 (secondary I–II) in Quebec. Approximately, 3/4 was White, 53.6% of the students were female, and 67.2% were from Ontario. Fifty-four percent lived in large urban areas, 31.3% of the students lived in small urban areas, and 46.8% of the students' household incomes were between \$50 and \$75K. The descriptive statistics of the linked samples at three annual waves were published in another paper (Yang et al., 2022).

Measures

The Cq was used to assess substance use, including cigarette, e-cigarette, alcohol, and cannabis use. The Cq includes two questions about cigarette and e-cigarette smoking to determine the prevalence and frequency of use of these substances. The first question asks if the students have tried smoking cigarettes (or e-cigarettes) to distinguish between never use and ever use. A follow-up question asks about the last 30-day use of cigarettes and e-cigarettes, with the responses ranging from “none” to “30 days (every day).” The Cq includes specific measures to assess the consumption of alcohol and cannabis in the last 12 months, with the responses ranging from “never use” to “every day.” The ordinal responses were collapsed into three-category indicators to avoid the sparseness of the observed frequency table. Following the most common categorization for determining the patterns of youth substance use, the initial responses were categorized into “never use,” “occasional use,” and “current use.”

In addition to the substance use measures, this study examined a set of risk and protective factors related to substance use among youth, encompassing a range of individual, social, and environmental factors that have been previously linked to youth substance use. The specific risk factors analyzed in the model included, but were not limited to, the number of smoking friends, the number of classes skipped, and the amount of weekly allowance. Additionally, protective factors, such as school connectedness, mental health (as measured by the Centre for Epidemiologic Studies Depression Scale (CES-D)), were also examined in the model. Please refer to the supplemental materials for a comprehensive list and detailed descriptions of all factors included in the model and analyzed.

Analysis

Before constructing the multivariate latent Markov models (LMM) to examine different patterns of substance use among youth, we applied the least absolute shrinkage and selection operator (LASSO) regression (Tibshirani, 1996), a linear regression method that performs both variable selection and regularization to prevent overfitting. In this study, LASSO regression was applied to select a subset of covariates from the dataset. The goal was to identify the most relevant risk and protective factors associated with youth substance use. The LASSO regression model was implemented using an adaptive version of the coordinate descent fitting algorithm with an ordinal response (Archer, 2015). This approach is suitable for dealing with an ordinal response variable, which was used to categorize substance use patterns.

After applying LASSO regression, we proceeded with the construction of LMM, aiming to identify different subgroups of youth exhibiting distinct patterns of substance use. LMM is a type of latent variable modeling technique that allows for the analysis of sequential data, such as repeated measurements over time. In this study, LMM was used to test the hypothesis that different subgroups of youth tend to exhibit different patterns of substance use over the 3-year longitudinal data. Comprising structural and measurement models, the LMM incorporates latent (unobserved) endogenous factors, such as substance use indicators, to model the conditional probabilities of response variables. Conversely, the measurement models solely involve manifested (observed) endogenous factors, conditioned on the latent status of substance use patterns. The LMM construction was built based on the variables selected through the LASSO regression process. The selection of the optimal number of classes was based on the Bayesian information criterion (BIC), which helps to assess the model's fit and complexity. Additionally, goodness-of-fit measures were utilized to evaluate the accuracy of the fitted models.

To provide a comprehensive understanding of the selection process, we chose only variables with non-zero coefficients after LASSO regression for further analysis and model construction. Multiple imputation (MI) was performed simultaneously with the LMM fitting to address missing data, and coefficients were averaged over five imputed runs to obtain overall coefficients.

To assess the significance of covariates on the effect of class membership for the initial state probabilities, we employed Wald test statistics based on the parameter estimates and standard errors. All p -values were calculated using two-sided tests, with results having $p < .05$ considered statistically significant.

Statistical analyses were conducted using the R language, open-source software for statistical computing and graphics (R Core Team, 2013). We used the LMMest package (Bartolucci et al., 2015) for generalized LMMs. Detailed descriptions of the analysis and LMM construction were published in another paper (Yang et al., 2022).

Results

Prevalence of Substance Use

Overall, the prevalence of “never use” has decreased over time, particularly for alcohol drinking (from 60.6% in 2016/17 to 29.2% in 2018/19). Over 3 years, the prevalence of “occasional use” and “current use” has risen for all substances. In particular, the current use of e-cigarettes (from 6.8% in 2016/17 to 32.3% in 2018/19) and cannabis (from 4.4% in 2016/17 to 16.1% in 2018/19) has increased 4.8 and 3.7 times, respectively.

PSU Patterns

Four patterns/classes of substance use were identified and summarized as follows: no-use (C1), alcohol-only (C2), concurrent use of e-cigarettes and alcohol (C3), and poly-use (C4). The final model classified each class based on the predominant conditional response probabilities of substance use. The conditional response probabilities were generally distinct, indicating a clear heterogeneity between the classes (see Fig. 1). At baseline (2016/17), the initial state probabilities of these use patterns were $P_{C1}=0.589$, $P_{C2}=0.216$, $P_{C3}=0.149$, and $P_{C4}=0.047$.

The Z-scores provide insights into the average substance use for each class compared to the overall average. For cigarette use, C4 shows above-average use, while C1, C2, and C3 have slightly below-average use. Similarly, C4 exhibits above-average e-cigarette use, C3 is slightly above the overall average, and C1 and C2 have slightly below-average use. For alcohol use, C1 has significantly below-average use, C2 and C3 are slightly above the overall average, and C4 has significantly above-average use. Regarding cannabis use, C4 demonstrates significantly above-average use, while C1, C2, and C3 have below-average use. This underscores the diversity in PSU patterns among youth, with distinct classes demonstrating different levels of engagement in cigarette, e-cigarette, alcohol, and cannabis use.

Associated Factors

We estimated the covariates' effects on their initial state probabilities to investigate the factors associated with the diverse patterns of PSU among youth. Table 1 lists the coefficients of all factors with corresponding odds ratios (OR) affecting the initial state probabilities for each class membership under the final model.

Overall, not gambling online for money ($OR_{C2}=0.22$, 95% CI $-0.16-0.58$; $OR_{C3}=0.14$, 95% CI $-0.24-0.52$; $OR_{C4}=0.08$, 95% CI $-0.47-0.63$), eating breakfast ($OR_{C2}=0.80$, 95% CI $0.66-0.94$; $OR_{C3}=0.64$, 95% CI $0.47-0.80$; $OR_{C4}=0.56$, 95% CI $0.25-0.87$), residing in urban areas ($OR_{C2}=0.82$, 95% CI $0.75-0.90$; $OR_{C3}=0.77$, 95% CI $0.68-0.85$; $OR_{C4}=0.66$, 95% CI $0.51-0.81$), having higher school connectedness ($OR_{C2}=0.95$, 95% CI $0.93-0.98$; $OR_{C3}=0.91$, 95% CI $0.89-0.94$; $OR_{C4}=0.82$, 95% CI $0.77-0.86$), and being black (vs. white, $OR_{C2}=0.92$, 95% CI $0.88-0.97$; $OR_{C3}=0.94$, 95% CI $0.89-0.99$; $OR_{C4}=0.96$, 95% CI $0.88-1.03$) consistently positively affected the initial membership in C2 through C4 relative to C1.

On the contrary, risk factors associated with PSU patterns included truancy as measured by number of classes skipped ($OR_{C2}=1.67$, 95% CI $1.55-1.79$; $OR_{C3}=1.92$, 95% CI $1.80-2.04$; $OR_{C4}=2.79$, 95% CI $2.64-2.94$), having more smoking friends ($OR_{C2}=1.35$, 95% CI $1.25-1.46$; $OR_{C3}=1.81$, 95% CI $1.71-1.91$; $OR_{C4}=2.75$, 95% CI $2.64-2.86$), being older ($OR_{C2}=1.32$, 95% CI $1.24-1.40$; $OR_{C3}=1.26$, 95% CI $1.17-1.34$; $OR_{C4}=1.51$,

Fig. 1 Polysubstance use patterns at baseline (2016/17). The patterns are determined by the prominent conditional response probabilities. X-axis: four substance use measures (left to right: cigarette, e-cigarette, alcohol, cannabis); Y-axis: conditional response probability between 0 and 1 (as shown in the caption below the image in a table format, prominent conditional response probabilities are highlighted in bold font); and Z-axis: three categories of never use (0), occasional use (1), and current use (2) for any substance: *Class 1 (C1, No-Use)* C1 corresponds to students who have not used substances, where the probability of never using any substance was > 94.7%. In 2016/17, C1 was the largest class, containing 60.5% of the students. The longitudinal evidence showed that although C1 was prominent in 2016/17, its prevalence had decreased significantly over time (60.5% → 39.0% → 25.0%; 2016/17 → 2017/18: $\Delta_{C1} = -21.5\%$; 2017/18 → 2018/19: $\Delta_{C1} = -14.0\%$). *Class 2 (C2, Alcohol-Only)* C2 was composed of individuals who typically used alcohol with a more significant probability of 89.6%, whereas over 85.5% of probabilities never used the other three substances in this class. As the second largest class consisting of 20.9% of youth in 2016/17, its prevalence had been relatively stable over time, increasing to 24.4% ($\Delta_{C2} = +3.5\%$) in 2017/18 and slightly decreasing to 22.3% ($\Delta_{C2} = -2.1\%$) by 2018/19. *Class 3 (C3, Concurrent Use of E-cigarettes and Alcohol)* Individuals in C3 had larger probabilities of using e-cigarettes and alcohol, 70.6% and 85.9%, respectively. The probabilities of using other substances in this class were not prominent. C3 was the second minor class in 2016/17, containing 14.0% of students. Its prevalence had increased significantly over time (14.0% → 24.9% → 32.5%; 2016/17 → 2017/18: $\Delta_{C3} = +10.9\%$; 2017/18 → 2018/19: $\Delta_{C3} = +7.6\%$), becoming the prominent use pattern by 2018/19. *Class 4 (C4, Poly-Use)*. Individuals in C4 differed from those in C3 by having a greater probability of using multiple substances concurrently. For instance, the conditional response probabilities of current e-cigarette, alcohol, and cannabis use were 74.5%, 79.5%, and 60.1%, respectively. Thus, C4 corresponds to the heavy/current multi-user group. Although C4 had been the minor class for 3 years, its prevalence had increased significantly across the 3 years by 4.4 times (4.6% → 11.7% → 20.2%; 2016/17 → 2017/18: $\Delta_{C4} = +7.1\%$; 2017/18 → 2018/19: $\Delta_{C4} = +8.5\%$), becoming very close to C2 and C1 by 2018/19. To provide clarity and facilitate labeling of the latent classes, we plotted the z-scores of each of the four substances across the four patterns/classes, C1 to C4 (Fig. 2).

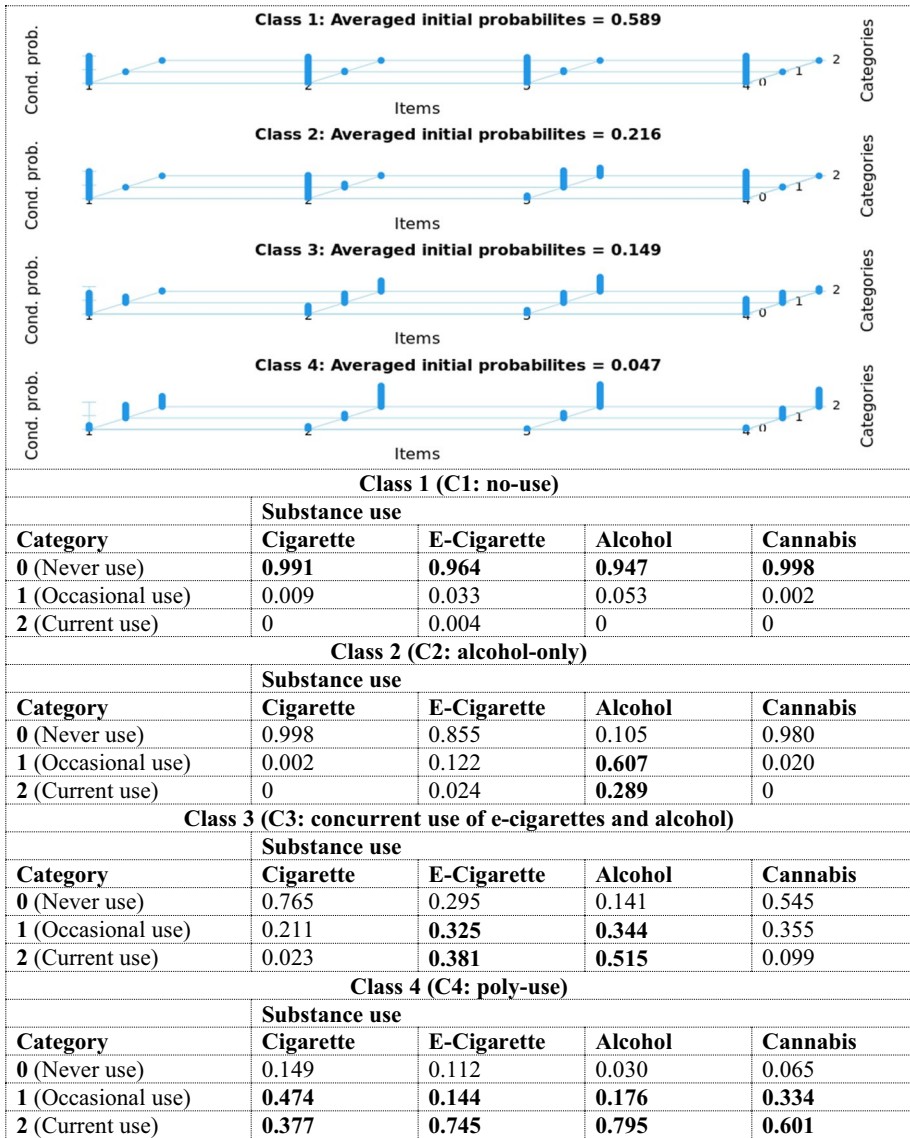
95% CI 1.37–1.65), attending school unsupportive in quitting drugs/alcohol ($OR_{C2} = 1.25$, 95% CI 1.17–1.34; $OR_{C3} = 1.30$, 95% CI 1.21–1.39; $OR_{C4} = 1.43$, 95% CI 1.27–1.60), obtaining more weekly allowance ($OR_{C2} = 1.14$, 95% CI 1.08–1.20; $OR_{C3} = 1.29$, 95% CI 1.22–1.36; $OR_{C4} = 1.34$, 95% CI 1.21–1.46), having elevated BMI ($OR_{C2} = 1.22$, 95% CI 1.16–1.28; $OR_{C3} = 1.20$, 95% CI 1.13–1.27; $OR_{C4} = 1.28$, 95% CI 1.16–1.40), having more physically active friends ($OR_{C2} = 1.21$, 95% CI 1.17–1.26; $OR_{C3} = 1.18$, 95% CI 1.13–1.23; $OR_{C4} = 1.10$, 95% CI 1.01–1.19), and having longer sedentary time ($OR_{C2} = 1.02$, 95% CI 1.01–1.04; $OR_{C3} = 1.07$, 95% CI 1.06–1.08; $OR_{C4} = 1.07$, 95% CI 1.04–1.09). These factors consistently negatively affected the initial membership in C2 through C4, relative to C1.

Sex (female as reference group vs. male, $OR_{C2} = 0.74$, 95% CI 0.60–0.89; $OR_{C3} = 1.34$, 95% CI 1.18–1.50; $OR_{C4} = 0.80$, 95% CI 0.51–1.10) had mixed positive and negative effects on the initial membership in C2 through C4 relative to C1. Except for race/ethnicity, all the factors had significant effects ($p < .05$) for each class.

Discussion

PSU Patterns

Our study, utilizing multivariate LMM modeling on longitudinal health surveys (COMPASS data), reveals four distinct patterns of PSU among Canadian adolescents: no-use (C1), alcohol-only (C2), concurrent use of e-cigarettes and alcohol (C3), and poly-use (C4). Each pattern represents an overarching theme, increasing severity by class. Generally, these patterns are well-separated, with considerably different sizes and good heterogeneity between classes. The longitudinal evidence on PSU patterns suggests an increasing



tendency and frequency of youth substance use over time, demonstrating a successful application of LMM. In this study, we focus on identifying youth PSU patterns and associated factors using LMM modeling technique for unobserved heterogeneity analysis.

Specifically, we observed that C4 consistently demonstrated above-average use of cigarettes, e-cigarettes, alcohol, and cannabis, while C1, C2, and C3 exhibited slightly below-average or significantly below-average use across these substances. The higher levels of substance use observed in C4 are in line with previous research on multiple substance use among adolescents. Studies have shown that individuals in the C4 category, often referred to as “multi-substance users,” are at higher risk of experiencing adverse health outcomes

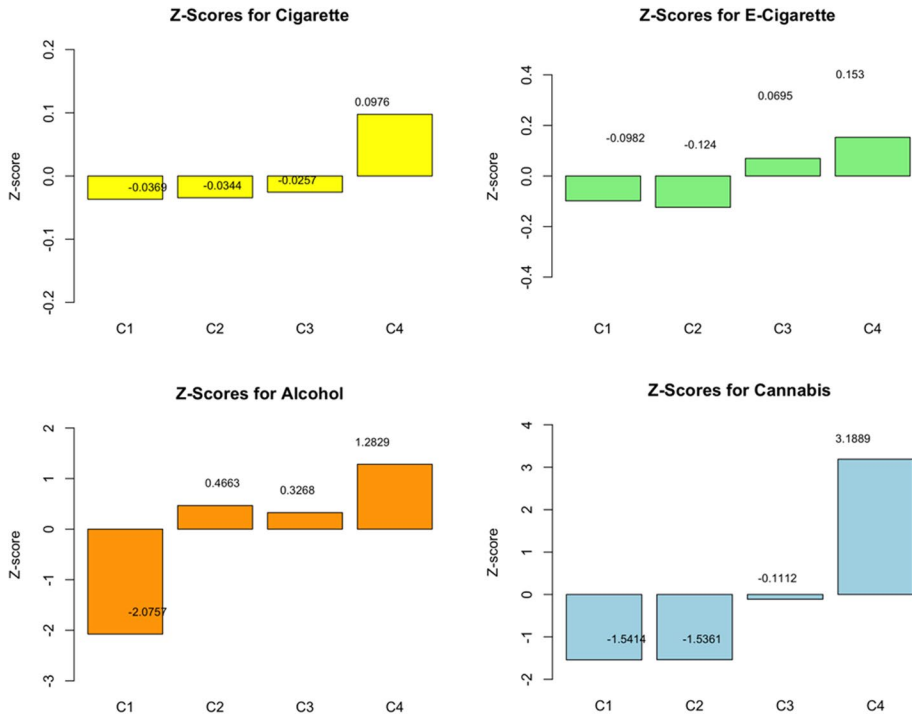


Fig. 2 Z-scores for each substance use across the four classes (2016/17). X-axis: four patterns/classes, from left to right, C1 (no-use), C2 (alcohol-only), C3 (concurrent use of e-cigarettes and alcohol), and C4 (poly-use); Y-axis: Z-scores for each substance use

and psychosocial problems compared to those in other classes. This finding is consistent with the notion that poly-use may indicate more severe or problematic substance use behaviors (Brown et al., 2015; Stockings et al., 2016).

The slight below-average or significantly below-average use of substances in C1, C2, and C3 suggests that these groups may be engaging in less frequent or lower-risk substance use behaviors. C1 exhibiting the lowest levels of substance use may represent a group of individuals who abstain from substance use altogether or have minimal exposure to such behaviors. C2 and C3 may include individuals who engage in substance use, but not at levels considered to be problematic or high risk. These findings align with previous research that identifies distinct subgroups of youth with varying levels of substance involvement, ranging from non-use to moderate or occasional use (Tomczyk et al., 2015; Lee et al., 2020).

According to a systematic review of youth substance use patterns from 70 studies conducted by Halladay et al., (2020), the average number of use patterns was four, including low use, single or dual use, moderate multi-use, and high multi-use (Halladay et al., 2020).¹² Our research identifies four patterns (C1 to C4) using novel analytical methods, aligning with their findings. Concerning e-cigarettes and other substance use among youth and young adults (Rothrock et al., 2020; Hughes et al., 2015; Thepthien et al., 2021; Mehra et al., 2019), studies identify the concurrent use of e-cigarettes with alcohol, cannabis, and tobacco products. Notably, alcohol and e-cigarette co-users among adolescents are representative in North America and other countries (Rothrock et al., 2020; Hughes et al., 2015; Thepthien et al.,

Table 1 Factors associated with polysubstance use patterns at baseline (2016/17, REF: C1)

	Subgroup		
	C2	C3	C4
Intercept			
β (beta coefficient)	-0.84	-0.54	-0.44
Odds ratios	0.43 (0.23 to 0.64)***	0.58 (0.45 to 0.71)***	0.65 (0.58 to 0.71)***
Urbanity: "rural" (REF) vs. "small urban" vs. "medium urban" vs. "large urban"			
Effect of urbanity on the probability of initial membership in C2 through C4, relative to C1 in 2016/17			
OR _{C2} = 0.82 can be interpreted as, in 2016/17, students who reported residing in small urban were 0.82 times less likely to start in the alcohol-only (C2) subgroup relative to the no-use (C1) subgroup than were those who reported residing in rural areas, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("large urban" vs. "medium urban" vs. "small urban" vs. "rural" (REF))			
OR _{C3} = 0.77 can be interpreted as, in 2016/17, students who reported residing in small urban were 0.77 times less likely to start in the concurrent use of e-cigarettes and alcohol (C3) subgroup relative to the no-use (C1) subgroup than were those who reported residing in rural areas, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("large urban" vs. "medium urban" vs. "small urban" vs. "rural" (REF))			
OR _{C4} = 0.66 can be interpreted as, in 2016/17, students who reported residing in small urban were 0.66 times less likely to start in the multi-use (C4) subgroup relative to the no-use (C1) subgroup than were those who reported residing in rural areas, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("large urban" vs. "medium urban" vs. "small urban" vs. "rural" (REF))			
β (beta coefficient)	-0.20	-0.27	-0.42
Odds ratios	0.82 (0.75 to 0.90)***	0.77 (0.68 to 0.85)***	0.66 (0.51 to 0.81)***
Grade: every one-grade increase (grades 7–10 at baseline)			
Effect of grade/age on the probability of initial membership in C2 through C4, relative to C1 in 2016/17			
OR _{C2} = 1.32 can be interpreted as, in 2016/17, an individual who's in grade 8 was 1.32 times more likely to start in the alcohol-only (C2) subgroup relative to the no-use (C1) subgroup than was an individual who's in grade 7, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons (grade 10 vs. grade 9 vs. grade 8 vs. grade 7)			
OR _{C3} = 1.26 can be interpreted as, in 2016/17, an individual who's in grade 8 was 1.26 times more likely to start in the concurrent use of e-cigarettes and alcohol (C3) subgroup relative to the no-use (C1) subgroup than was an individual who's in grade 7, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons (grade 10 vs. grade 9 vs. grade 8 vs. grade 7)			
OR _{C4} = 1.51 can be interpreted as, in 2016/17, an individual who's in grade 8 was 1.51 times more likely to start in the poly-use (C4) subgroup relative to the no-use (C1) subgroup than was an individual who's in grade 7, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons (grade 10 vs. grade 9 vs. grade 8 vs. grade 7).			
β (beta coefficient)	0.28	0.23	0.41
Odds ratios	1.32 (1.24 to 1.40)***	1.26 (1.17 to 1.34)***	1.51 (1.37 to 1.65)***
Race/ethnicity: "White" (REF) vs. "Black" vs. "Asian" vs. "Aboriginal (First Nations, Métis, Inuit)" vs. "Latin American/Hispanic" vs. "Other"			
Effect of race/ethnicity on the probability of initial membership in C2 through C4, relative to C1 in 2016/17			
OR _{C2} = 0.92 can be interpreted as, in 2016/17, students who reported being black were 0.92 times less likely to start in the alcohol-only (C2) subgroup relative to the no-use (C1) subgroup than were those who reported being white, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("Other" vs. "Latin American/Hispanic" vs. "Aboriginal (First Nations, Métis, Inuit)" vs. "Asian" vs. "Black" vs. "White" (REF))			
OR _{C3} = 0.94 can be interpreted as, in 2016/17, students who reported being black were 0.94 times less likely to start in the concurrent use of e-cigarettes and alcohol (C3) subgroup relative to the no-use (C1) subgroup than were those who reported being white, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("Other" vs. "Latin American/Hispanic" vs. "Aboriginal (First Nations, Métis, Inuit)" vs. "Asian" vs. "Black" vs. "White" (REF))			
OR _{C4} = 0.96 can be interpreted as, in 2016/17, students who reported being black were 0.96 times less likely to start in the multi-use (C4) subgroup relative to the no-use (C1) subgroup than were those who reported being white, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("Other" vs. "Latin American/Hispanic" vs. "Aboriginal (First Nations, Métis, Inuit)" vs. "Asian" vs. "Black" vs. "White" (REF))			
β (beta coefficient)	-0.08	-0.06	-0.04
Odds ratios	0.92 (0.88 to 0.97)**	0.94 (0.89 to 0.99)*	0.96 (0.88 to 1.03)+++

Table 1 (continued)

	Subgroup		
	C2	C3	C4
Weekly allowance: "I do not know how much money I get each week" (REF) vs. "zero" vs. "\$1-\$20" vs. "\$21-\$100" vs. "\$100+"			
Effect of weekly allowance on the probability of initial membership in C2 through C4, relative to C1 in 2016/17			
OR _{C2} = 1.14 can be interpreted as, in 2016/17, an individual who had \$100+ weekly allowance was 1.14 times more likely to start in the alcohol-only (C2) subgroup relative to the no-use (C1) subgroup than was an individual who had \$21-\$100 weekly allowance, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("100+" vs. "\$21-\$100" vs. "\$1-\$20" vs. "zero" vs. "I do not know how much money I get each week" (REF))			
OR _{C3} = 1.29 can be interpreted as, in 2016/17, an individual who had \$100+ weekly allowance was 1.29 times more likely to start in the concurrent use of e-cigarettes and alcohol (C3) subgroup relative to the no-use (C1) subgroup than was an individual who had \$21-\$100 weekly allowance, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("100+" vs. "\$21-\$100" vs. "\$1-\$20" vs. "zero" vs. "I do not know how much money I get each week" (REF))			
OR _{C4} = 1.34 can be interpreted as, in 2016/17, an individual who had \$100+ weekly allowance was 1.34 times more likely to start in the poly-use (C4) subgroup relative to the no-use (C1) subgroup than was an individual who had \$21-\$100 weekly allowance, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("100+" vs. "\$21-\$100" vs. "\$1-\$20" vs. "zero" vs. "I do not know how much money I get each week" (REF))			
β (beta coefficient)	0.13	0.25	0.29
Odds ratios	1.14 (1.08 to 1.20)***	1.29 (1.22 to 1.36)***	1.34 (1.21 to 1.46)***
# of physically active friends: "none" (REF) vs. "1 friend" vs. "2 friends" vs. "3 friends" vs. "4 friends" vs. "5 friends or more"			
Effect of # of physically active friends on the probability of initial membership in C2 through C4, relative to C1 in 2016/17			
OR _{C2} = 1.21 can be interpreted as, in 2016/17, an individual who had one physically active friend was 1.21 times more likely to start in the alcohol-only (C2) subgroup relative to the no-use (C1) subgroup than was an individual who had none physically active friend, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("5 friends or more" vs. "4 friends" vs. "3 friends" vs. "2 friends" vs. "1 friend" vs. "none" (REF))			
OR _{C3} = 1.18 can be interpreted as, in 2016/17, an individual who had one physically active friend was 1.18 times more likely to start in the concurrent use of e-cigarettes and alcohol (C3) subgroup relative to the no-use (C1) subgroup than was an individual who had none physically active friend, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("5 friends or more" vs. "4 friends" vs. "3 friends" vs. "2 friends" vs. "1 friend" vs. "none" (REF))			
OR _{C4} = 1.10 can be interpreted as, in 2016/17, an individual who had one physically active friend was 1.10 times more likely to start in the poly-use (C4) subgroup relative to the no-use (C1) subgroup than was an individual who had none physically active friend, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("5 friends or more" vs. "4 friends" vs. "3 friends" vs. "2 friends" vs. "1 friend" vs. "none" (REF))			
β (beta coefficient)	0.19	0.17	0.10
Odds ratios	1.21 (1.17 to 1.26)***	1.18 (1.13 to 1.23)***	1.10 (1.01 to 1.19)**
Eating breakfast: "No" (REF) vs. "Yes"			
Effect of eating breakfast on the probability of initial membership in C2 through C4, relative to C1 in 2016/17			
OR _{C2} = 0.80 can be interpreted as, in 2016/17, students who reported eating breakfast were 0.80 times less likely to start in the alcohol-only (C2) subgroup relative to the no-use (C1) subgroup than were those who reported not eating breakfast, assuming that all the other variables were held constant			
OR _{C3} = 0.64 can be interpreted as, in 2016/17, students who reported eating breakfast were 0.64 times less likely to start in the concurrent use of e-cigarettes and alcohol (C3) subgroup relative to the no-use (C1) subgroup than were those who reported not eating breakfast, assuming that all the other variables were held constant			
OR _{C4} = 0.56 can be interpreted as, in 2016/17, students who reported eating breakfast were 0.56 times less likely to start in the multi-use (C4) subgroup relative to the no-use (C1) subgroup than were those who reported not eating breakfast, assuming that all the other variables were held constant			
β (beta coefficient)	-0.22	-0.45	-0.59
Odds ratios	0.80 (0.66 to 0.94)*	0.64 (0.47 to 0.80)***	0.56 (0.25 to 0.87)**

Table 1 (continued)

	Subgroup		
	C2	C3	C4
# of smoking friends: “none” (REF) vs. “1 friend” vs. “2 friends” vs. “3 friends” vs. “4 friends” vs. “5 or more friends”			
Effect of # of smoking friends on the probability of initial membership in C2 through C4, relative to C1 in 2016/17			
OR _{C2} = 1.35 can be interpreted as, in 2016/17, an individual who had one smoking friend was 1.35 times more likely to start in the alcohol-only (C2) subgroup relative to the no-use (C1) subgroup than was an individual who had none smoking friend, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons (“5 or more friends” vs. “4 friends” vs. “3 friends” vs. “2 friends” vs. “1 friend” vs. “none” (REF))			
OR _{C3} = 1.81 can be interpreted as, in 2016/17, an individual who had one smoking friend was 1.81 times more likely to start in the concurrent use of e-cigarettes and alcohol (C3) subgroup relative to the no-use (C1) subgroup than was an individual who had none smoking friend, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons (“5 or more friends” vs. “4 friends” vs. “3 friends” vs. “2 friends” vs. “1 friend” vs. “none” (REF))			
OR _{C4} = 2.75 can be interpreted as, in 2016/17, an individual who had one smoking friend was 2.75 times more likely to start in the poly-use (C4) subgroup relative to the no-use (C1) subgroup than was an individual who had none smoking friend, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons (“5 or more friends” vs. “4 friends” vs. “3 friends” vs. “2 friends” vs. “1 friend” vs. “none” (REF))			
β (beta coefficient)	0.30	0.59	1.01
Odds ratios	1.35 (1.25 to 1.46)***	1.81 (1.71 to 1.91)***	2.75 (2.64 to 2.86)***
Support quit drug/alcohol: “Very supportive” (REF) vs. “Supportive” vs. “Unsupportive” vs. “Very unsupportive”			
Effect of support quit drug alcohol on the probability of initial membership in C2 through C4, relative to C1 in 2016/17			
OR _{C2} = 1.25 can be interpreted as, in 2016/17, an individual whose school was supportive to quit drug/alcohol was 1.25 times more likely to start in the alcohol-only (C2) subgroup relative to the no-use (C1) subgroup than was an individual whose school was very supportive, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons (“Very unsupportive” vs. “Unsupportive” vs. “Supportive” vs. “Very supportive” (REF))			
OR _{C3} = 1.30 can be interpreted as, in 2016/17, an individual whose school was supportive to quit drug/alcohol was 1.30 times more likely to start in the concurrent use of e-cigarettes and alcohol (C3) subgroup relative to the no-use (C1) subgroup than was an individual whose school was very supportive, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons (“Very unsupportive” vs. “Unsupportive” vs. “Supportive” vs. “Very supportive” (REF))			
OR _{C4} = 1.43 can be interpreted as, in 2016/17, an individual whose school was supportive to quit drug/alcohol was 1.43 times more likely to start in the poly-use (C4) subgroup relative to the no-use (C1) subgroup than was an individual whose school was very supportive, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons (“Very unsupportive” vs. “Unsupportive” vs. “Supportive” vs. “Very supportive” (REF))			
β (beta coefficient)	0.22	0.26	0.36
Odds ratios	1.25 (1.17 to 1.34)***	1.30 (1.21 to 1.39)***	1.43 (1.27 to 1.60)***
Sex: “Female” (REF) vs. “Male”			
Effect of sex on the probability of initial membership in C2 through C4, relative to C1 in 2016/17			
OR _{C2} = 0.74 can be interpreted as assuming all the other variables were held constant, in 2016/17, females were 1.35 times (OR _{C2} = 1/0.74) more likely to engage in C2 than males relative to C1			
OR _{C3} = 1.34 can be interpreted as assuming all the other variables were held constant, in 2016/17, males were 1.34 times more likely to start in C3 than females relative to C1			
OR _{C4} = 0.80 can be interpreted as assuming all the other variables were held constant, in 2016/17, females were 1.25 times (OR _{C4} = 1/0.80) more likely to engage in C4 than males relative to C1			
β (beta coefficient)	-0.30	0.29	-0.22
Odds ratios	0.74 (0.60 to 0.89)**	1.34 (1.18 to 1.50)**	0.80 (0.51 to 1.10)++

2021). Our results support the evidence of dual and multi-use of e-cigarettes with other substances, highlighting the importance of considering e-cigarettes in the analysis of substance use patterns. Moreover, it is alarming that the current use of e-cigarettes has increased significantly over time, from 6.8% in 2016/17 to 32.3% in 2018/19, by 4.8 times.

Table 1 (continued)

	Subgroup		
	C2	C3	C4
# of classes skipped (truancy): "0 classes" (REF) vs. "1 or 2 classes" vs. "3 to 5 classes" vs. "6 to 10 classes" vs. "11 to 20 classes" vs. "More than 20 classes"			
Effect of # of classes skipped on the probability of initial membership in C2 through C4, relative to C1 in 2016/17			
OR _{C2} = 1.67 can be interpreted as, in 2016/17, an individual who reported skipping 1 or 2 classes was 1.67 times more likely to start in the alcohol-only (C2) subgroup relative to the no-use (C1) subgroup than was an individual who reported zero classes skipped, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("More than 20 classes" vs. "11 to 20 classes" vs. "6 to 10 classes" vs. "3 to 5 classes" vs. "1 or 2 classes" vs. "0 classes" (REF))			
OR _{C3} = 1.92 can be interpreted as, in 2016/17, an individual who reported skipping 1 or 2 classes was 1.92 times more likely to start in the concurrent use of e-cigarettes and alcohol (C3) subgroup relative to the no-use (C1) subgroup than was an individual who reported zero classes skipped, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("More than 20 classes" vs. "11 to 20 classes" vs. "6 to 10 classes" vs. "3 to 5 classes" vs. "1 or 2 classes" vs. "0 classes" (REF))			
OR _{C4} = 2.79 can be interpreted as, in 2016/17, an individual who reported skipping 1 or 2 classes was 2.79 times more likely to start in the multi-use (C4) subgroup relative to the no-use (C1) subgroup than was an individual who reported zero classes skipped, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("More than 20 classes" vs. "11 to 20 classes" vs. "6 to 10 classes" vs. "3 to 5 classes" vs. "1 or 2 classes" vs. "0 classes" (REF))			
β (beta coefficient)	0.51	0.65	1.03
Odds ratios	1.67 (1.55 to 1.79)***	1.92 (1.80 to 2.04)***	2.79 (2.64 to 2.94)***
BMI category: "Not Stated" (REF) vs. "Underweight" vs. "Healthy Weight" vs. "Overweight" vs. "Obese"			
Effect of BMI category on the probability of initial membership in C2 through C4, relative to C1 in 2016/17			
OR _{C2} = 1.22 can be interpreted as, in 2016/17, an individual who reported being obese was 1.22 times more likely to start in the alcohol-only (C2) subgroup relative to the no-use (C1) subgroup than was an individual who reported being overweight, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("Obese" vs. "Overweight" vs. "Healthy Weight" vs. "Underweight" vs. "Not Stated" (REF))			
OR _{C3} = 1.20 can be interpreted as, in 2016/17, an individual who reported being obese was 1.20 times more likely to start in the concurrent use of e-cigarettes and alcohol (C3) subgroup relative to the no-use (C1) subgroup than was an individual who reported being overweight, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("Obese" vs. "Overweight" vs. "Healthy Weight" vs. "Underweight" vs. "Not Stated" (REF))			
OR _{C4} = 1.28 can be interpreted as, in 2016/17, an individual who reported being obese was 1.28 times more likely to start in the multi-use (C4) subgroup relative to the no-use (C1) subgroup than was an individual who reported being overweight, assuming that all the other variables were held constant. Same ORs apply to all categorical comparisons ("Obese" vs. "Overweight" vs. "Healthy Weight" vs. "Underweight" vs. "Not Stated" (REF))			
β (beta coefficient)	0.20	0.18	0.25
Odds ratios	1.22 (1.16 to 1.28)***	1.20 (1.13 to 1.27)***	1.28 (1.16 to 1.40)**
School connectedness: every one-unit increase in the score (ranging from 6 to 24)			
Effect of school connectedness on the probability of initial membership in C2 through C4, relative to C1 in 2016/17			
OR _{C2} = 0.95 can be interpreted as, in 2016/17, every one-unit increase in the school connectedness, students were 0.95 times less likely to start in the alcohol-only (C2) subgroup relative to the no-use (C1) subgroup, assuming that all the other variables were held constant			
OR _{C3} = 0.91 can be interpreted as, in 2016/17, every one-unit increase in the school connectedness, students were 0.91 times less likely to start in the concurrent use of e-cigarettes and alcohol (C3) subgroup relative to the no-use (C1) subgroup, assuming that all the other variables were held constant			
OR _{C4} = 0.82 can be interpreted as, in 2016/17, every one-unit increase in the school connectedness, students were 0.82 times less likely to start in the multi-use (C4) subgroup relative to the no-use (C1) subgroup, assuming that all the other variables were held constant			
β (beta coefficient)	-0.05	-0.09	-0.20
Odds ratios	0.95 (0.93 to 0.98)**	0.91 (0.89 to 0.94)***	0.82 (0.77 to 0.86)***

Table 1 (continued)

	Subgroup		
	C2	C3	C4
Sedentary time: every 1-h increase			
Effect of sedentary time on the probability of initial membership in C2 through C4, relative to C1 in 2016/17			
OR _{C2} = 1.02 can be interpreted as, in 2016/17, every 1-h increase in sedentary time, an individual was 1.02 times more likely to start in the alcohol-only (C2) subgroup relative to the no-use (C1) subgroup, assuming that all the other variables were held constant			
OR _{C3} = 1.07 can be interpreted as, in 2016/17, every 1-h increase in sedentary time, an individual was 1.07 times more likely to start in the concurrent use of e-cigarettes and alcohol (C3) subgroup relative to the no-use (C1) subgroup, assuming that all the other variables were held constant			
OR _{C4} = 1.07 can be interpreted as, in 2016/17, every 1-h increase in sedentary time, an individual was 1.07 times more likely to start in the poly-use (C4) subgroup relative to the no-use (C1) subgroup, assuming that all the other variables were held constant			
β (beta coefficient)	0.00	0.00	0.00
Odds ratios	1.02 (1.01 to 1.04)*	1.07 (1.06 to 1.08)***	1.07 (1.04 to 1.09)***
Gambling online: “Yes” (REF) vs. “No”			
Effect of gambling online on the probability of initial membership in C2 through C4, relative to C1 in 2016/17			
OR _{C2} = 0.22 can be interpreted as, in 2016/17, students who reported not gambling online for money for the last 30 days were 0.22 times less likely to start in the alcohol-only (C2) subgroup relative to the no-use (C1) subgroup than were those who reported gambling online, assuming that all the other variables were held constant			
OR _{C3} = 0.14 can be interpreted as, in 2016/17, students who reported not gambling online for money for the last 30 days were 0.14 times less likely to start in the concurrent use of e-cigarettes and alcohol (C3) subgroup relative to the no-use (C1) subgroup than were those who reported gambling online, assuming that all the other variables were held constant			
OR _{C4} = 0.08 can be interpreted as, in 2016/17, students who reported not gambling online for money for the last 30 days were 0.08 times less likely to start in the multi-use (C4) subgroup relative to the no-use (C1) subgroup than were those who reported gambling online, assuming that all the other variables were held constant			
β (beta coefficient)	-1.55	-1.98	-2.54
Odds ratios	0.22 (-0.16 to 0.58)***	0.14 (-0.24 to 0.52)***	0.08 (-0.47 to 0.63)***

The same ORs apply to all categorical comparisons. Interpretations were given to ORs highlighted in bold. Ninety-five percent confidence intervals are in brackets

*** $p < .00001$; ** $p < .001$; * $p < .05$; +++the result is *not* significant at $p < .05$

C1 no-use, C2 alcohol-only, C3 concurrent use of e-cigarettes and alcohol, C4 poly-use

Associated Factors

Evidence suggests that factors such as sex, race, early onset of alcohol drinking, and academic achievements in secondary school can influence youth substance use patterns (Lanza, Patrick & Maggs, 2010).²⁶ Lanza et al., (2010) found that, when examining alcohol, cigarette, and cannabis use, males are 4.5 times more likely to be in the highest use group than females relative to non-users. In contrast, females are more likely to smoke cigarettes or binge drink than their male peers. Several studies among US youth have found that female students are at a higher risk of using multiple substances (Silveira et al., 2019; Cranford, McCabe & Boyd, 2013). Similarly, our study reveals sex-related differences in substance use patterns among youth. We found that in 2016/17, females were 1.35 times more likely to engage in alcohol use (C2) compared to males, while males were 1.34 times more likely to initiate concurrent e-cigarette and alcohol use (C3) relative to females. Additionally, females exhibited a 1.25 times higher likelihood of engaging in poly-use (C4) compared to males. Our findings highlight the complex interplay between sex and

substance use behaviors, emphasizing the need for tailored interventions and prevention strategies to address the specific needs of male and female adolescents.

Our study reveals the impact of race/ethnicity on substance use patterns among youth. In 2016/17, students who reported being black were 0.92 times less likely to start in the alcohol-only (C2) subgroup relative to the no-use (C1) subgroup, compared to their white counterparts. Similarly, black students were 0.94 times less likely to start in the concurrent use of e-cigarettes and alcohol (C3) subgroup compared to white students and 0.96 times less likely to start in the multi-use (C4) subgroup relative to the no-use (C1) subgroup. These findings are consistent with prior research, which has also shown that black students are less likely to engage in multiple substance use compared to white students (Connell, Gilreath & Hansen, 2009; Gilreath et al., 2014). Additionally, studies focusing on Asian, Hispanic, and other ethnic students have consistently found lower risk of using more than three substances (Zuckermann et al., 2020; Lanza, Patrick & Maggs, 2010; Banks et al., 2017; Keyes et al., 2022). Our results demonstrate similar protective effects, indicating that Latin American/Hispanic students are less likely to engage in higher-use patterns compared to Asians, who, in turn, have a lower likelihood of engaging in higher-use patterns than black students. These findings highlight the importance of considering race/ethnicity as a potential factor in understanding substance use disparities among youth and may help inform targeted prevention efforts and interventions to address substance use behaviors in different racial and ethnic groups.

In addition to sex and race/ethnicity differences, we contribute to the addiction literature with multifaceted factors and their estimates examined. To our knowledge, no other research includes all the variables investigated in our study to evaluate the association with the initial membership of PSU patterns, summarizing the positive, negative, or mixed effects. For example, peer effect has been identified to correlate with multi-use (Tomczyk, Isensee & Hanewinkel, 2016; Tomczyk, Hanewinkel & Isensee, 2015). Our results show that students with more smoking friends are 2.75 times more likely to start in C4 than their peers who reported fewer smoking friends relative to C1. Likewise, having more physically active friends was 1.21 times more likely to engage in C2 relative to C1, demonstrating the significant impact of peer influence on youth substance use. Older students are more likely to use multiple substances (Zuckermann et al., 2020; Zuckermann et al., 2019). Some studies have found that students who are in higher-use classes are more likely to have access to more spending money (Zuckermann et al., 2020; Valente, Cogo-Moreira & Sanchez, 2017), which is consistent with our results that having more weekly allowance is associated with engaging in a higher-use class ($OR_{C2} = 1.14$, 95% CI 1.08 – 1.20; $OR_{C3} = 1.29$, 95% CI 1.22 – 1.36; $OR_{C4} = 1.34$, 95% CI 1.21 – 1.46). Our results suggest that students with elevated BMI are 1.28 times more likely to engage in C4 than C1. In addition to these individual-level factors, environmental factors such as urbanity and school policies supporting anti-smoking/drugs are important. Our study reveals that students attending a school unsupportive to quitting drugs/alcohol are 1.43 times more likely to start in C4 than those attending a supportive school relative to C1. Our results agree with Silveira et al. (2019), who reported that youth living in non-urban areas are more likely to engage in multi-use of substances (Silveira et al., 2019).

Underage gambling, a problematic behavior, is positively associated with substance use (Barnes et al., 2009). Our study reveals that students not gambling online are less likely ($OR_{C4} = 0.08$) to start in C4 than otherwise, relative to C1. Concerning the correlation between youth substance use and health behaviors, e.g., attitudes towards nutrition, Isralowitz and Trostler (1996) found that substance users are more likely to have unhealthy eating habits, such as skipping breakfast or not eating three meals a day, which puts them at a higher risk for negative health outcomes (Isralowitz and Trostler, 1996). Eating breakfast, a proxy for healthy behaviors, is picked up from our results concerning personal health and its protective effects against a

higher-use class. Low social connectedness (White, Walton & Walker, 2015), e.g., school connectedness or engagement, is a risk factor for multi-use among youth. Some studies have shown that having a lower level of school connection is associated with an increased likelihood of multi-use of substances, consistent with our findings (Zuckermann et al., 2020; Jongenelis et al., 2019).

This study investigates cross-sectional and longitudinal evidence on PSU patterns from multi-dimensional COMPASS data. The factors contributing to youth substance use are complex, involving individual characteristics, peer influence, and environmental components. The diversity of student characteristics, modifiable individual-level risk factors, and the influence of family, friends, school environment, and community-level factors can all contribute to the heterogeneity of substance use patterns. Factors such as eating habits (Israelowitz & Trostler, 1996), gambling behaviors (Barnes et al., 2009), BMI (Hammami et al., 2019), weekly allowances (Zuckermann et al., 2020; Valente, Cogo-Moreira & Sanchez, 2017), peer influence (Tomczyk, Isensee & Hanewinkel, 2016; Tomczyk, Hanewinkel & Isensee, 2015), truancy (Henry & Thornberry, 2010), school programs/policies supporting anti-drugs/alcohol, and school connectedness (White, Walton & Walker, 2015; Su, Supple & Kuo, 2018) are modifiable, which can have multi-dimensional interventions tailored to address these risk factors collectively. In comparison, grade/age and sex are non-modifiable, so interventions can target at-risk groups, such as higher grades and females who tend to be more likely to engage in alcohol and multi-use. Differences in substance use patterns at the population level may significantly affect the mental and behavioral health of youth. This can be further conceptualized as the ability of different school settings to prevent addictive behaviors.

Implications

Our study findings have significant implications for public health and health policy, providing insights that can guide interventions and prevention strategies to reduce the prevalence of substance use behaviors among youth. Understanding the use patterns and associated risk factors can help identify at-risk groups and tailor interventions to address the specific needs of different subgroups. This targeted approach is crucial for reducing the negative consequences of substance abuse and addiction, benefiting both individual youth and society as a whole. Decision makers should consider the diverse associations between substance use patterns and multiple health-related behaviors to design school environments, policies, and procedures that improve the health behaviors of youth.

The variation in PSU patterns observed in our study highlights the importance of considering heterogeneity among youth when developing prevention and intervention strategies. Tailored approaches that address the unique risk profiles and needs of each class can be more effective in reducing substance use and associated harms (Johnston et al., 2016; Ramo et al., 2012). Our findings emphasize the importance of targeted and evidence-based prevention and intervention strategies to address the unique needs of each subgroup. For instance, preventive efforts targeting C4 individuals may focus on harm reduction strategies, early intervention, and providing support for those at higher risk of experiencing substance-related problems.

Moreover, this study underscores the significance of early detection and screening for substance use behaviors among youth. Identifying and addressing substance use patterns at an early stage can prevent escalation to more problematic use and related negative consequences (Volkow & Baler, 2014). Schools, healthcare providers, and community organizations play a crucial role in offering preventive measures and support to adolescents, especially those identified as being at higher risk of engaging in multiple substance use.

Demographic differences were found to impact the initial state probabilities of diverse class membership at baseline (2016/17). Students in higher grades, living in Alberta (Zuckermann et al., 2020), coming from rural areas in Canada, or being obese, had a higher probability of starting in higher-use classes (C2 to C4) relative to C1 compared to those with other demographic characteristics. Therefore, provincial and federal governments should collaborate to develop specialized preventive programs tailored to the unique needs of different regions and demographics, particularly in areas with higher prevalence rates.

It is important to highlight that in this study, the risk and protective factors were not hand-picked from the relevant literature. Instead, a rigorous variable selection process using LASSO regression was employed, ensuring that the factors included in the model are supported by empirical evidence. This data-driven approach enhances the validity and robustness of our findings and provides transparency in the factors' inclusion. Employing advanced machine learning techniques, such as LASSO and multivariate latent variable modeling, represents a significant advancement in understanding the complexity of PSU patterns and associated factors among youth. This cross-disciplinary approach showcases the power of data science in addressing real-world public health challenges. By combining cutting-edge techniques with a comprehensive examination of PSU patterns, our study provides a more nuanced analysis of the diverse pathways and interactions between different substances during adolescence.

Study Limitations

This study has certain limitations. First, it is preferable to have external validation data to evaluate our model performance. Yet, the data collection cycle from 2019/20 onward is challenging due to the COVID-19 pandemic. Second, caution should be exercised in interpreting and generalizing the results of the COMPASS study due to the non-random convenience sampling method used for many participating schools. As a result, the school samples may not fully represent the entire population and may exclude certain Canadian provinces (Yang et al., 2022). Third, our analyses are confined to the variables available in the Cq and are subjected to self-reported bias. Fourth, as with any large-scale health survey, non-responses introduce missing values. We apply MI techniques to impute missing values. Despite this, stakeholders agree that the COMPASS methodologies are reliable in handling the delicate balance between data accuracy and participant anonymity in longitudinal studies focused on youth health behaviors (Battista et al., 2019). Moreover, it is important to note that our findings are in line with previous research examining patterns of substance use among youth, which further supports the validity and generalizability of our results (Esser et al., 2021; Steinhoff et al., 2022). However, there may be other factors not included in our analysis that could influence substance use patterns among youth, such as peer influence, family dynamics, and environmental factors (Marschall-Lévesque et al., 2017; Russell et al., 2019). Future research may consider incorporating these additional factors to provide a more comprehensive understanding of youth substance use behaviors. Furthermore, one review indicates a significant decline in the prevalence of youth substance use during the pandemic (Layman et al., 2022). However, it is crucial to continue monitoring and surveilling youth substance use in the post-pandemic years to ensure a comprehensive understanding of trends and potential changes in behavior.

Conclusion

Our findings shed light on the prevalence and complexity of PSU among youth, emphasizing the significance of understanding the heterogeneity of substance use patterns and associated factors during this critical developmental period. By recognizing unique risk factors associated with specific subgroups, targeted interventions can be developed to address the multifaceted nature of substance use behaviors among youth effectively. This comprehensive approach contributes valuable insights to the existing literature, enhancing our ability to design evidence-based prevention strategies and improve public health outcomes for adolescents. This approach effectively deepened our understanding of this pressing public health challenge and has the potential to inform the development of effective, evidence-based strategies to address it. Ultimately, our data-driven analysis paves the way for more effective and tailored interventions to address the challenges of substance use among youth. Understanding the complex relationship between substance use patterns and multifaceted impact factors empowers us to protect the well-being of the younger generation and foster a healthier, more resilient society.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11469-023-01139-2>.

Data, Materials, and/or Code Availability COMPASS study data is available upon request through the completion and approval of an online form: <https://uwaterloo.ca/compass-system/information-researchers/data-usage-application>. The datasets and code used during the current study are available from the corresponding author upon reasonable request.

Author Contributions YY, HC, ZB, and SL conceived the research questions and designed the study. YY performed all analyses, wrote the first draft, and developed the tables and figures. SL provided data access and helped study design and manuscript edit. HC, ZB, and SL supervised this study. All authors reviewed drafts. All authors contributed to and have approved the final draft.

Funding We acknowledge the receipt of a PhD scholarship from the Applied Health Sciences (AHS) at the University of Waterloo awarded to YY and a Microsoft AI for Good grant awarded to HC to support this research. The COMPASS study has been supported by a bridge grant from the CIHR Institute of Nutrition, Metabolism and Diabetes (INMD) through the “Obesity – Interventions to Prevent or Treat” priority funding awards (OOP-110788; awarded to SL), an operating grant from the CIHR Institute of Population and Public Health (IPPH) (MOP-114875; awarded to SL), a CIHR project grant (PJT-148562; awarded to SL), a CIHR bridge grant (PJT-149092; awarded to SL), and by a research funding arrangement with Health Canada (#1617-HQ-000012; contract awarded to SL), a CIHR-Canadian Centre on Substance Abuse (CCSA) team grant (OF7 B1-PCPEGT 410-10-9633; awarded to SL), a project grant from the CIHR Institute of Population and Public Health (IPPH) (PJT-180262; awarded to SL). The COMPASS-Quebec project additionally benefits from funding from the Ministère de la Santé et des Services sociaux of the province of Québec and the Direction régionale de santé publique du CIUSSS de la Capitale-Nationale.

Declarations

Ethics Approval The COMPASS study has been granted ethics clearance by the University of Waterloo Office of Research Ethics (ORE 30118).

Consent to Participate Informed consent was obtained from all participants for being included in the COMPASS host study.

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons

licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahumada, G., Areneda, J. C., Clarke, P., et al. (2019). *Report on drug use in the Americas*.
- Archer, K. J. (2015). *glmpathcr: An R package for ordinal response prediction in high-dimensional data settings*.
- Banks, D. E., Rowe, A. T., Mpofo, P., & Zapolski, T. C. (2017). Trends in typologies of concurrent alcohol, marijuana, and cigarette use among US adolescents: An ecological examination by sex and race/ethnicity. *Drug and alcohol dependence, 179*, 71–77.
- Barnes, G. M., Welte, J. W., Hoffman, J. H., & Tidwell, M. C. O. (2009). Gambling, alcohol, and other substance use among youth in the United States. *Journal of studies on alcohol and drugs, 70*(1), 134–142.
- Bartolucci, F., Farcomeni, A., Pandolfi, S., & Pennoni, F. (2015). LMest: An R package for latent Markov models for categorical longitudinal data. <https://arxiv.org/abs/1501.04448>.
- Battista, K., Qian, W., Bredin, C., & Leatherdale, S. T. (2019). Student data linkage over multiple years. *COMPASS Tech Rep Ser., 6*, 1–10.
- Brown, S. A., Brumback, T. Y., Tomlinson, K., Cummins, K., Thompson, W. K., Nagel, B. J., De Bellis, M. D., Hooper, S. R., Clark, D. B., Chung, T., & Hasler, B. P. (2015). The National Consortium on Alcohol and NeuroDevelopment in Adolescence (NCANDA): A multisite study of adolescent development and substance use. *Journal of studies on alcohol and drugs, 76*(6), 895–908.
- Connell, C. M., Gilreath, T. D., & Hansen, N. B. (2009). A multiprocess latent class analysis of the co-occurrence of substance use and sexual risk behavior among adolescents. *Journal of studies on alcohol and drugs, 70*(6), 943–951.
- Cranford, J. A., McCabe, S. E., & Boyd, C. J. (2013). Adolescents' nonmedical use and excessive medical use of prescription medications and the identification of substance use subgroups. *Addictive behaviors, 38*(11), 2768–2771.
- Esser, M. B., Pickens, C. M., Guy, G. P. Jr., & Evans, M. E. (2021). Binge drinking, other substance use, and concurrent use in the US, 2016–2018. *American Journal of Preventive Medicine, 60*(2), 169–178.
- Gilreath, T. D., Astor, R. A., Estrada, J. N., Johnson, R. M., Benbenishty, R., & Unger, J. B. (2014). Substance use among adolescents in California: A latent class analysis. *Substance Use & Misuse, 49*(1-2), 116–123.
- Government of Canada. Summary of the Canadian Student Tobacco, Alcohol and Drugs Survey 2018–19. <https://www.canada.ca/en/health-canada/services/canadian-student-tobacco-alcohol-drugs-survey/2018-2019-summary.html>
- Haines-Saah, R. J., & Fischer, B. (2021). Youth cannabis use and legalization in Canada—Reconsidering the fears, myths and facts three years in. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 30*(3), 191.
- Hale, D. R., & Viner, R. M. (2016). The correlates and course of multiple health risk behaviour in adolescence. *BMC public health, 16*, 1–12.
- Halladay, J., Woock, R., El-Khechen, H., Munn, C., MacKillop, J., Amlung, M., Ogrodnik, M., Favotto, L., Aryal, K., Noori, A., & Kiflen, M. (2020). Patterns of substance use among adolescents: A systematic review. *Drug and alcohol dependence, 216*, 108222.
- Hammami, N., Chaurasia, A., Bigelow, P., & Leatherdale, S. T. (2019). A gender-stratified, multilevel latent class assessment of chronic disease risk behaviours' association with body mass index among youth in the COMPASS study. *Preventive medicine, 126*, 105758.
- Henry, K. L., & Thornberry, T. P. (2010). Truancy and escalation of substance use during adolescence. *Journal of studies on alcohol and drugs, 71*(1), 115–124.
- Hughes, K., Bellis, M. A., Hardcastle, K. A., McHale, P., Bennett, A., Ireland, R., & Pike, K. (2015). Associations between e-cigarette access and smoking and drinking behaviours in teenagers. *BMC public health, 15*, 1–9.

- Isralowitz, R. E., & Trostler, N. (1996). Substance use: Toward an understanding of its relation to nutrition-related attitudes and behavior among Israeli high school youth. *Journal of adolescent health, 19*(3), 184–189.
- Johnston, L. D., O'Malley, P. M., Miech, R. A., Bachman, J. G., & Schulenberg, J. E. (2016). *Monitoring the future national survey results on drug use, 1975-2015: Overview, key findings on adolescent drug use*. Institute for Social Research.
- Jongenelis, M., Pettigrew, S., Lawrence, D., & Rikkers, W. (2019). Factors associated with poly drug use in adolescents. *Prevention Science, 20*, 695–704.
- Keyes, K. M., Kreski, N. T., Ankrum, H., Cerdá, M., Chen, Q., Hasin, D. S., Martins, S. S., Olfson, M., & Miech, R. (2022). Frequency of adolescent cannabis smoking and vaping in the United States: Trends, disparities and concurrent substance use, 2017–19. *Addiction, 117*(8), 2316–2324.
- Kraus, L., Håkan, L., Guttormsson, U., et al. (2015). *Results from the European School Survey Project on Alcohol and Other Drugs. ESPAD* www.emcdda.europa.eu
- Lanza, S. T., Patrick, M. E., & Maggs, J. L. (2010). Latent transition analysis: Benefits of a latent variable approach to modeling transitions in substance use. *Journal of drug issues, 40*(1), 93–120.
- Layman, H. M., Thorisdottir, I. E., Halldorsdottir, T., Sigfusdottir, I. D., Allegrante, J. P., & Kristjansson, A. L. (2022). Substance use among youth during the COVID-19 pandemic: A systematic review. *Current psychiatry reports, 24*(6), 307–324.
- Leatherdale, S. T., Brown, K. S., Carson, V., Childs, R. A., Dubin, J. A., Elliott, S. J., Faulkner, G., Hammond, D., Manske, S., Sabiston, C. M., & Laxer, R. E. (2014). The COMPASS study: A longitudinal hierarchical research platform for evaluating natural experiments related to changes in school-level programs, policies and built environment resources. *BMC public health, 14*(1), 1–7.
- Lee, H., Yang, K., Palmer, J., Kameg, B., Clark, L., & Greene, B. (2020). Substance use patterns among adolescents: A latent class analysis. *Journal of the American Psychiatric Nurses Association, 26*(6), 586–594.
- Marschall-Lévesque, S., Castellanos-Ryan, N., Parent, S., Renaud, J., Vitaro, F., Boivin, M., Tremblay, R. E., & Séguin, J. R. (2017). Victimization, suicidal ideation, and alcohol use from age 13 to 15 years: Support for the self-medication model. *Journal of Adolescent Health, 60*(4), 380–387.
- Mehra, V. M., Keethakumar, A., Bohr, Y. M., Abdullah, P., & Tamim, H. (2019). The association between alcohol, marijuana, illegal drug use and current use of E-cigarette among youth and young adults in Canada: Results from Canadian Tobacco, Alcohol and Drugs Survey 2017. *BMC public health, 19*, 1–10.
- Merrin, G. J., & Leadbeater, B. (2018). Do classes of polysubstance use in adolescence differentiate growth in substances used in the transition to young adulthood? *Substance use & misuse, 53*(13), 2112–2124.
- R Core Team, R. (2013). R: A language and environment for statistical computing.
- Ramo, D. E., Liu, H., & Prochaska, J. J. (2012). Tobacco and marijuana use among adolescents and young adults: A systematic review of their co-use. *Clinical psychology review, 32*(2), 105–121.
- Reel, R., Bredin, C., Battista, K., & Leatherdale, S. T. (2018). COMPASS year 5 and 6 school recruitment and retention. *Compass Tech Rep Ser, 5*(1), 1–10.
- Rothrock, A. N., Andris, H., Swetland, S. B., Chavez, V., Isaak, S., Pagane, M., Romney, J., & Rothrock, S. G. (2020). Association of E-cigarettes with adolescent alcohol use and binge drinking-drunkenness: A systematic review and meta-analysis. *The American Journal of Drug and Alcohol Abuse, 46*(6), 684–698.
- Russell, B. S., Simpson, E., Flannery, K. M., & Ohannessian, C. M. (2019). The impact of adolescent substance use on family functioning: The mediating role of internalizing symptoms. *Youth & Society, 51*(4), 504–528.
- Silveira, M. L., Green, V. R., Iannaccone, R., Kimmel, H. L., & Conway, K. P. (2019). Patterns and correlates of polysubstance use among US youth aged 15–17 years: Wave 1 of the Population Assessment of Tobacco and Health (PATH) Study. *Addiction, 114*(5), 907–916.
- Steinhoff, A., Bechtiger, L., Ribeaud, D., Eisner, M. P., Quednow, B. B., & Shanahan, L. (2022). Polysubstance use in early adulthood: Patterns and developmental precursors in an urban cohort. *Frontiers in Behavioral Neuroscience, 15*, 797473.
- Stockings, E., Hall, W. D., Lynskey, M., Morley, K. I., Reavley, N., Strang, J., Patton, G., & Degenhardt, L. (2016). Prevention, early intervention, harm reduction, and treatment of substance use in young people. *The Lancet Psychiatry, 3*(3), 280–296.
- Su, J., Supple, A. J., & Kuo, S. I. C. (2018). The role of individual and contextual factors in differentiating substance use profiles among adolescents. *Substance use & misuse, 53*(5), 734–743.

- Thepthien, B. O., Tinn, C. S., Ofuchi, T., & Kim, B. (2021). An analysis of e-cigarette and polysubstance use patterns of adolescents in Bangkok, Thailand. *Tobacco induced diseases, 19*, 88.
- Thompson-Haile, A., & Leatherdale, S. T. (2013). Student-level data collection procedures. *COMPASS Technical Report Series* (vol. 1, no. 5). University of Waterloo. Available at: www.compass.uwaterloo.ca
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267–288.
- Tomczyk, S., Hanewinkel, R., & Isensee, B. (2015). Multiple substance use patterns in adolescents—A multilevel latent class analysis. *Drug and alcohol dependence, 155*, 208–214.
- Tomczyk, S., Isensee, B., & Hanewinkel, R. (2016). Latent classes of polysubstance use among adolescents—A systematic review. *Drug and alcohol dependence, 160*, 12–29.
- Valente, J. Y., Cogo-Moreira, H., & Sanchez, Z. M. (2017). Gradient of association between parenting styles and patterns of drug use in adolescence: A latent class analysis. *Drug and alcohol dependence, 180*, 272–278.
- Volkow, N. D., & Baler, R. D. (2014). Addiction science: Uncovering neurobiological complexity. *Neuropharmacology, 76*, 235–249.
- White, J., Walton, D., & Walker, N. (2015). Exploring comorbid use of marijuana, tobacco, and alcohol among 14 to 15-year-olds: Findings from a national survey on adolescent substance use. *BMC public health, 15*(1), 1–9.
- Yang, Y., Butt, Z. A., Leatherdale, S. T., Morita, P. P., Wong, A., Rosella, L., & Chen, H. H. (2022). Exploring the dynamic transitions of polysubstance use patterns among Canadian youth using latent Markov models on COMPASS data. *The Lancet Regional Health-Americas, 16*, 100389.
- Zapert, K., Snow, D. L., & Tebes, J. K. (2002). Patterns of substance use in early through late adolescence. *American Journal of Community Psychology, 30*(6), 835–852.
- Zuckermann, A. M., Williams, G., Battista, K., de Groh, M., Jiang, Y., & Leatherdale, S. T. (2019). Trends of poly-substance use among Canadian youth. *Addictive behaviors reports, 10*, 100189.
- Zuckermann, A. M., Williams, G. C., Battista, K., Jiang, Y., de Groh, M., & Leatherdale, S. T. (2020). Prevalence and correlates of youth poly-substance use in the COMPASS study. *Addictive behaviors, 107*, 106400.
- Zuckermann, A. M., Battista, K. V., Bélanger, R. E., Haddad, S., Butler, A., Costello, M. J., & Leatherdale, S. T. (2021a). Trends in youth cannabis use across cannabis legalization: Data from the COMPASS prospective cohort study. *Preventive medicine reports, 22*, 101351.
- Zuckermann, A. M., Gohari, M. R., Romano, I., & Leatherdale, S. T. (2021b). Changes in cannabis use modes among Canadian youth across recreational cannabis legalization: Data from the COMPASS prospective cohort study. *Addictive behaviors, 122*, 107025.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Yang Yang¹  · Zahid A. Butt¹ · Scott T. Leatherdale¹ · Helen H. Chen¹

✉ Yang Yang
y24yang@uwaterloo.ca

✉ Helen H. Chen
helen.chen@uwaterloo.ca

Zahid A. Butt
zahid.butt@uwaterloo.ca

Scott T. Leatherdale
sleatherdale@uwaterloo.ca

¹ School of Public Health Sciences, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada