# IDENTIFYING PREDICTORS OF MARIJUANA VAPING AND INVESTIGATING ITS POLYDRUG USE PATTERNS, USING MACHINE LEARNING AND LATENT CLASS ANALYSIS.

Siyoung Choe

Submitted to the faculty of the Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the School of Public Health
Indiana University
February 2024

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

Dong-Chul Seo, Ph.D.

Jon Agley, Ph.D.

Kit Elam, Ph.D.

Aurelian Bidulescu, Ph.D.

December 14, 2023

ACKNOWLEDGEMENTS

My dissertation would not have been possible without patience and guidance from all of my committee members – Drs. Agley, Bidulescu, and Elam. I would like to thank them sincerely for their patience and support. My sincerest gratitude goes to my research committee chair and advisor, Dr. Dong-Chul Seo. He has been the most incredible mentor anyone could have asked for, and I owe him everything I learned over the past decade. I also want to thank Dr. Hsien-Chang Lin for helping me begin my studies during Dr. Seo's sabbatical, and Drs. Feeney and Kitley for helping me manage my health. Without their help, my journey in academia would not have been possible.

Siyoung Choe

IDENTIFYING PREDICTORS OF MARIJUANA VAPING AND INVESTIGATING ITS POLYDRUG USE PATTERNS, USING MACHINE LEARNING AND LATENT CLASS ANALYSIS.

Approximately 36% of individuals who currently use marijuana use vaporized form of marijuana, and young adults are considered most vulnerable to marijuana vaping. However, the number of studies on marijuana vaping is limited, and predictors of marijuana vaping among U.S. young adults remain unclear. Existing studies on marijuana vaping have known limitations, as they (1) used regression-based methods over machine-learning methods, (2) focused on examining marijuana vaping initiation but not on its use over multiple years, (3) focused on mono-use of marijuana vaping, despite how poly-substance use is likely among individuals who vape marijuana, and (4) fail to account for states' non-medical marijuana laws. This dissertation research filled gaps in the literature, by identifying predictors of multi-year marijuana vaping among U.S. young adults using machine learning analysis (sub-study 1), and analyzing patterns of marijuana vaping and concurrent use of other drugs among U.S. young adults using latent class analysis (sub-study 2). Restricted data from the Population Assessment of Tobacco and Health Study was used for analyses. Sub-study 1 utilized regression with Least Absolute Shrinkage and Selection Operator (LASSO) and Classification and Regression Tree (CART) to create a five-terminal-node prediction model for states that legalized non-medical marijuana (split by marijuana use, cigarette use, bullying behavior, and ethnicity) and another five-terminal-node prediction model for states that have not legalized non-medical marijuana (split by marijuana use, heroin use, e-cigarette use, and hookah use). Sub-study 2 identified models with four classes (abstainers, drinkers, smoker-drinkers, and multi-substance users) to explain varying co-occurrence of 10 substance groups (marijuana vaping, joint marijuana smoking, blunt marijuana smoking, alcohol, cigarette, e-cigarette, cigar, other tobacco product, prescription drugs, and illegal drugs). Results suggest that characteristics predicting multi-year marijuana vaping may differ from those of marijuana vaping initiation. Results also highlight the importance of accounting for non-medical marijuana laws.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

## Chapter 1

## INTRODUCTION

*Marijuana smoking*

Marijuana refers to a dried hemp plant of the *Cannabis* genus (NIDA, 2019). Marijuana contains a variety of psychoactive and addictive chemicals, including its primary active component, delta-9-tetrahydrocannabinol (THC) (NIDA, 2019). The 2018 Farm Bill differentiates marijuana from similar expressions like *hemp* and *cannabis* based on THC concentration, so it specifically refers to cannabis that include more than 0.3% of THC by dry weight (Agricultural Improvement Act of 2018, Pub. L. No. 115-334), though terms *marijuana* and *cannabis* are sometimes used synonymously. Smoking is the most common method of marijuana use, though other method of use exists as well (e.g., ingestion of marijuana edibles or THC-rich resins) (NIDA, 2019). Marijuana smokers inhale combusted forms of marijuana in hand-rolled products (i.e., joints or blunts) or in pipes or water pipes (i.e., bongs), though inhalation of a vaporized form of THC extract has become more popular in recent years (NIDA, 2019).

Marijuana smoking is associated with short- and long-term health effects. One of the most notable short-term effects of marijuana smoking is stimulation and overactivation of the brain's cannabinoid receptors, which gives the user a sense of "high" (NIDA, 2021). Other short-term effects include altered senses, mood change, impaired body, impaired thoughts, impaired memory, hallucinations, delusions, and psychosis (NIDA, 2019). When used over a prolonged period, marijuana smoking may adversely affect the users' cognition capacity, heart rate, and breathing (Meier et al., 2012). Regular and prolonged marijuana smoking may even result in

Cannabinoid Hyperemesis Syndrome (Galli, Sawaya, & Friedenberg, 2011), Cannabis Use Disorder (Hasin et al., 2015; Lopez-Quintero et al., 2011), psychosis (Di Forti et al., 2015), and schizophrenia (Di Forti et al., 2009; Di Forti et al., 2014).

In addition to the aforementioned health effects, marijuana smoking may come with consequences not intended by the user. For example, studies have found that frequent marijuana smokers are likely to report lower life satisfaction, more relationship problems, high absenteeism, and higher risk of injury compared to non-users (McCaffrey et al., 2010; Zwerling, Ryan, & Orav, 1990). Marijuana smoking also impairs individuals' ability to drive motor vehicles, as studies found a significant negative correlation between blood THC concentration and individuals' motor coordination, judgment, and reaction time (Lenne et al., 2010; Hartman & Huestis, 2013; Hartman et al., 2015). Studies have suggested that marijuana use and subsequent rise in blood THC levels increase individuals' risk of vehicle crashes, such that the risk of a fatal car crash is twice as high among drivers with THC in their blood compared to drivers who did not use any drugs or alcohol (Brady & Li, 2014; Biecheler et al., 2008; Elvik, 2013). The risk of a fatal crash is greater when marijuana is used in combination with alcohol (Hartman & Huestis, 2013; Elvik, 2013), and such risk is known to be synergistic (i.e., the combined risk is greater than individual risks added together).

*Marijuana regulations*

Due to adverse effects associated with marijuana, many countries have banned the use of marijuana. Law and policy in the U.S., however, are complicated by differing rules at state and federal levels. At the federal level, marijuana is currently classified as a Schedule 1 substance, along with drugs like heroin and peyote (Shukla & Doshi, 2021; Volkow et al., 2014). As a

Schedule 1 drug, marijuana is considered a controlled substance with no accepted medical purpose while having a high potential for abuse by the Drug Enforcement Administration (DEA) and the Food and Drug Administration (FDA) (Shukla & Doshi, 2021). However, health professionals often disagree with marijuana's Schedule 1 classification and compare marijuana to Schedule 2 substances like opiates and stimulants, which have a high risk of abuse but are medically useful (Volkow et al., 2014). As such, many states have opted to go against established federal rules in recent years. Some states have merely decriminalized marijuana, while many states have partially or fully legalized the substance (NCSL, 2023). Currently, the majority of states allow medical use of marijuana and cannabis products, while twenty-four states, the District of Columbia, and two territories even allow non-medical ("recreational") use for adults (NCSL, 2023). The state-specific legal status of non-medical marijuana is summarized in **Table 1**.

**Table 1.** The legal status of non-medical adult use of marijuana and marijuana products in the U.S., as of December 13, 2023

| State | Allows non-medical adult use | Year in effect |
|---|---|---|
| Alabama | No | - |
| Alaska | Yes | 2014 |
| Arizona | Yes | 2020 |
| Arkansas | No | - |
| California | Yes | 2016 |
| Colorado | Yes | 2012 |
| Connecticut | Yes | 2021 |

| State | Allows non-medical adult use | Year in effect |
|---|---|---|
| Delaware | Yes | 2023- |
| Florida | No | - |
| Georgia | No | - |
| Hawaii | No | - |
| Idaho | No | - |
| Illinois | Yes | 2020 |
| Indiana | No | - |
| Iowa | No | - |
| Kansas | No | - |
| Kentucky | No | - |
| Louisiana | No | - |
| Maine | Yes | 2018 |
| Maryland | Yes | 2022 |
| Massachusetts | Yes | 2016 |
| Michigan | Yes | 2018 |
| Minnesota | Yes | 2023 |
| Mississippi | No | - |
| Missouri | Yes | 2022 |
| Montana | Yes | 2020 |
| Nebraska | No | - |
| Nevada | Yes | 2016 |
| New Hampshire | No | - |

| State | Allows non-medical adult use | Year in effect |
|---|---|---|
| New Jersey | Yes | 2020 |
| New Mexico | Yes | 2021 |
| New York | Yes | 2021 |
| North Carolina | No | - |
| North Dakota | No | - |
| Ohio | Yes | 2023 |
| Oklahoma | No | - |
| Oregon | Yes | 2014 |
| Pennsylvania | No | - |
| Rhode Island | Yes | 2022 |
| South Carolina | No | - |
| South Dakota | No | - |
| Tennessee | No | - |
| Texas | No | - |
| Utah | No | - |
| Vermont | Yes | 2018 |
| Virginia | Yes | 2021 |
| Washington | Yes | 2012 |
| West Virginia | No | - |
| Wisconsin | No | - |
| Wyoming | No | - |
| District of Columbia | Yes | 2014 |

| State | Allows non-medical adult use | Year in effect |
|---|---|---|
| Guam | Yes | 2019 |
| Northern Mariana Islands | Yes | 2018 |
| Puerto Rico | No | - |
| Virgin Islands | No | - |

*Note.* Data from NCSL, 2023

*Marijuana smoking in the U.S.*

Marijuana is one of the most commonly used federally illegal substances in the U.S., even in states that have not legalized its use (Bostwick, 2012; Center for Behavioral Health Statistics and Quality, 2015). It is estimated that approximately 18% of the U.S. population has used marijuana at least once in 2019 (SAMHSA, 2020), which is higher than the global estimate of 4% (Peacock et al., 2018).

Most common method of marijuana administration in the U.S. is joint smoking (i.e., using marijuana in rolling papers that do not contain tobacco) (Hindocha et al., 2016), though blunt smoking (i.e., rolling marijuana in hollowed-out cigarette, cigar, or cigarillo shells) may be as common in certain demographics (Fairman et al., 2015; Schauer et al., 2017). Blunt smoking has become increasingly popular in recent years among tobacco-marijuana co-users, young adults, and racial/ethnic minorities (Fairman et al., 2015; Schauer et al., 2017).

State legalization of marijuana has had a significant influence on the overall number of marijuana users in the U.S., as the number of marijuana users and marijuana poisoning have increased post-legalization of marijuana (Wang et al., 2014). However, it has been suggested that not all race/ethnicities are equally influenced by the legalization, as a 2021 study found that medical marijuana law increased marijuana use among Hispanic and non-Hispanic white individuals but not among non-Hispanic black individuals (Martins et al., 2021).

As the number of adult users increased, the number of non-adults intentionally and unintentionally exposed to marijuana also increased (Wang, 2014). In 2019, 37% of U.S. high school students reported using marijuana at least once in their lifetime, and 22% reported use in the past 30 days (Jones, 2020). Many of these adolescents use vaporized forms of marijuana, such that 6.5% of 8th graders, 13.1% of 10th graders, and 19.6% of 12th graders have reported past-year marijuana vaping (Miech et al., 2023).

*Marijuana vaping in the U.S.*

Marijuana vaping is a relatively novel form of THC delivery (NIDA, 2020). Individuals who vape marijuana use electronic vaporizers to inhale THC, similar to how electronic cigarettes (also known as *e-cigarettes* or *electronic nicotine delivery systems, or ENDS*) are used for inhaling vaporized nicotine (Daniulaityte et al., 2017; NIDA 2020). In the U.S., it is estimated that about 36% of individuals who use marijuana use a vaporized form of marijuana, the vast majority of which use multiple forms of marijuana delivery (i.e., there are individuals who concurrently vape and smoke marijuana) (Steigerwald et al., 2018). Marijuana vapers are likely to be younger, male, more frequent marijuana users, and co-users of e-cigarettes (Etter et al., 2015; Lee et al., 2016).

Marijuana vaping has gained attention in recent years, as many of the *e-cigarette or vaping product associated lung injury* (EVALI) cases in the U.S. have been linked to vitamin E-acetate-laced THC used by individuals who vape marijuana (Perrine et al., 2019). Dangers of marijuana vaping are not limited to EVALI, and studies have linked marijuana vaping to multiple health detriments, including accidental poisoning, explosion injury, and cardiovascular disease (Chadi, Minato,& Stanwick, 2020; Shahandeh, Chowdhary, & Middlekauff, 2021). Such detriments may be irreversible once they occur, and early intervention and prevention are warranted.

*Marijuana vaping among young adults*

Marijuana vaping is of special concern for young adults (Navon et al., 2019) due to its recent growth in prevalence (NIDA, 2020), increased access with state legalization of marijuana (NIDA, 2020; Whitehill et al., 2019), low perceived risk of use (Chadi, Minato,& Stanwick, 2020; Shahandeh, Chowdhary, & Middlekauff, 2021), and high convenience of use (Kenne et al., 2017). Indeed, the rate of past-month marijuana vaping has sharply increased for both collegiate and non-collegiate young adults (respectively, 4.9 and 6.7% to 11.2% and 18.9%) during 2017-2022 (Patrick et al., 2023), while the numbers have remained relatively unchanged for the older adult population during the similar period (NIDA, 2020). Given how young adults are particularly vulnerable to marijuana vaping (Navo et al., 2019) and given how health behaviors established during young adulthood tend to continue later in life (Lawrence et al., 2020), interventions must prevent initiation of marijuana vaping for young adults who are naïve to marijuana vaping, while preventing continuation for young adults who have already been exposed to marijuana vaping.

*Limitations of existing marijuana vaping studies*

Despite the danger it poses to young adults and the need to prevent its use, marijuana vaping has been relatively understudied. Past studies have largely focused on nicotine vaping or conventional forms of marijuana delivery (e.g., smoked or edible forms of marijuana) (Cherian et al., 2020; Perrine et al., 2019). Studies that did examine marijuana vaping were sometimes modeled after marijuana smoking studies (Cherian et al., 2020), despite how individuals who vape marijuana may have profiles that are unique compared to individuals who smoke marijuana (Han & Seo, 2022). Hence, there are potentially unidentified predictors of marijuana vaping that warrant exploration. A recent study by Han and Seo (2022) has been relatively comprehensive, and examined a total of 48 predictors of marijuana vaping. However, factors such as body mass index (BMI), housing, and health insurance status remain unexplored, despite how they may be significantly associated with individuals' substance use behavior (Mojtabai et al., 2020; Foltin, Fischman, & Byrne, 1988; Beulaygue & French, 2016; Gentzke et al., 2018; Douglas, 2010).

Past studies were mostly reliant on regression-based and linearity-based approaches to testing predictors of marijuana vaping (Boccio and Jackson, 2021; Cassidy et al., 2018; Kritikos et al., 2020; Lee et al., 2021; Pokhrel et al., 2020), though such approach may fail to examine complex interactive effects between variables beyond two- or three-way interactions (Yong et al., 2020). Our understanding of characteristics and behaviors related to marijuana vaping may be enhanced using novel analytic methods, such as machine learning, as demonstrated by Han and Seo (2022). Han and Seo (2022) used the least absolute shrinkage and selection operator (LASSO) and classification and regression tree (CART) to examine marijuana vaping initiation in states with and without recreational marijuana legalization. They found that bullying behavior may be an externalized endophenotypic symptom of marijuana vaping in states without recreational marijuana legalization. While Han and Seo's study was relatively rigorous, a follow-up study is warranted because not all of those individuals who initiate marijuana vaping are going to be long-term marijuana users. A study exploring continued marijuana vaping is needed to understand if the characteristics of prolonged marijuana vapers are different from those who quit after short-term experimentation.

Lastly, past studies have focused on the mono-use of marijuana vaping, and studies on polydrug use seldom included marijuana vaping as a substance of interest (Stanton et al., 2020; Miech et al., 2016; Buu et al.,2020). While concurrent or sequential use of multiple substances may increase the severity of withdrawal symptoms and hinder cessation efforts (Lemyre, Poliakova, & Bélanger, 2019), little is known about how marijuana vaping is used alongside other substances. Hence, a comprehensive investigation into patterns of polydrug use that include marijuana vaping is warranted (Agrawal, Budney, & Lynskey, 2012).

*Need for the research*

Literature on marijuana vaping is relatively sparse, and a lot remains unknown regarding marijuana vaping. As mentioned above, the literature on marijuana vaping is limited in four key ways: (1) existing studies have not rigorously examined BMI, housing, and health insurance status as potential predictors of marijuana vaping; (2) existing studies have examined lifetime/ever marijuana vaping and marijuana vaping initiation, but not marijuana vaping continuation; (3) existing studies relied on methods that may be less suitable for exploring emerging problem like marijuana vaping; and (4) exiting studies have focused on marijuana vaping as a mono-use substance, and how marijuana vaping is done concurrently with other substances (e.g., tobacco and alcohol) remain largely unknown.

Chapter 2 of this dissertation document includes a comprehensive summary of the literature to show how gaps in the literature needed to be filled, and illustrate how this dissertation research filled those gaps in the literature.

*Purpose of the research*

This dissertation research filled gaps in the literature via two sub-studies.

The purpose of sub-study 1 was to use machine learning analysis to identify risk profiles of marijuana vaping. Specifically, the study was designed to identify and examine emerging predictors of multi-year marijuana vaping among U.S. young adults while accounting for variables (e.g., housing, insurance status, BMI, and legalization of non-medical marijuana) that have not been rigorously explored in past studies. This was one of the first studies to focus on multi-year marijuana vaping, which is unique given how other measures of marijuana vaping (e.g., initiation, past-year use, and ever-use) often fail to differentiate long-term users from short-term experimenters. Moreover, machine learning is a technique not often used in public health research, and the methods described in this study may serve as a guide for additional machine learning research in public health.

The purpose of sub-study 2 was to comprehensively examine polydrug use patterns of multiple substances, including marijuana vaping, using latent class analysis. Specifically, the study (1) identified classes of substance use to examine which of the ten legal and illegal substances commonly abused by young adults are used concurrently, (2) examined predictors of each substance class, and (3) explored potential heterogeneity by user characteristics. In addition to identifying patterns of polydrug use, this study sought to examine predictors of underlying substance use patterns to identify individuals who are most vulnerable to polydrug use. This is especially important because it is currently unclear whether or not some predictors' influence on individual substances also extends to co-occurring substances. This study filled an important gap in the literature, as no studies to date have rigorously examined marijuana vaping and its co-occurring substance use among U.S. young adults, especially using nationally representative data.

*Hypotheses*

For sub-study 1, I hypothesized that variables identified as important by machine learning technique would differ from those identified by a non-machine learning technique. I hypothesized, based on the marijuana vaping initiation study by Han and Seo (2022), that the state's law on non-medical marijuana would have a significant influence on multi-year marijuana vaping. Given how housing, health insurance status, and BMI have shown significant association with substance use in previous studies, I also expected housing, health insurance status, and BMI to be significant predictors of multi-year marijuana vaping.

For sub-study 2, I hypothesized that marijuana vaping is done concurrently with e-cigarettes and marijuana because marijuana vaping has a chemical profile similar to conventional (i.e., smoked) marijuana while having a mode of delivery (i.e., vapor inhalation using an electronic device) similar to e-cigarettes. I hypothesized, based on previous studies (Lanza, Motlagh, & Orozco, 2020; Steigerwald et al, 2018; Sawdey, 2017; Buu et al., 2020), that (1) patterns of polydrug use would differ by state legalization of non-medical marijuana; and (2) individuals' race/ethnicity, sexual orientation, housing, employment, degree enrollment, health insurance status, and perceived quality of life significantly predict class membership.

*Delimitations*

The scope of this dissertation research has been delimited as follows:

Both sub-studies defined young adults as individuals aged 18-25. This definition was based on studies in the literature (Higley, 2019; Arnett, 2000), though studies were largely inconsistent on when young adulthood ends (Higley, 2019).

Sub-study 1 only included young adults (age 18-25) with non-missing outcomes at waves 4-6. Sub-study 2 included all young adults (age 18-25) with non-missing outcomes at wave 6. While some past studies have defined young adulthood as having a wider age range (age 18-34) (Wadhahi et al., 2021; Ng Fat & Shelton, 2012), this dissertation research chose only to include individuals aged 18-25, as brain development tends to differ before and after age 25 (Arain et al., 2013), and the way substance use behavior interacts with brain chemistry in individuals aged 18-25 may differ from those aged 26-34 (Winters & Arria, 2012; Squeglia et al., 2009).

*Limitations*

This dissertation research has known limitations: Both sub-study 1 and sub-study 2 were limited by the data. While the data used for analyses is widely used nationally representative data with a relatively comprehensive set of variables, this research is still considered secondary, and includes limitations inherent in non-primary data analyses.

Sub-study 1 was limited by the inclusion of individuals with three waves of matched data. This was a limitation because individuals sometimes dropped out in subsequent waves, which made their data unusable for this study and influenced the final sample size. Multiple imputation was used to prevent further sample reduction. Sub-study 1 was also limited by an indirect measurement of the outcome. Because none of the existing survey items directly asked about multi-year marijuana vaping, a proxy measure based on when they last vaped marijuana at each survey year was necessary.

Sub-study 2 was limited by its use of complete case analysis. While the sample had less than five percent of missing data, and complete case analysis is deemed the computational default by the syntax utilized for the study, the resulting loss of observations was a limitation that nevertheless influenced the size of the analytic sample. Sub-study 2 was also limited by its cross-sectional design. While latent class analysis (LCA) was an appropriate tool for the current polydrug use study, a future study using longitudinal LCA (LLCA) may show how polydrug use may change over time.

*Definitions of terms and abbreviations*

- AIC = Akaike Information Criteria. A criterion for model fit and selection, along with Bayesian information criterion (BIC). Models with lower AIC are typically considered better than those with higher AIC. Small-sample corrected AIC (also known as AICc) is generally preferred over a standard AIC, especially when *n* is small.
- AUROC = Area Under the Receiver Operating Characteristic curve. AUROC score summarizes the performance of a model, where a value of 0.5 represents a random guess, and a value of 1 represents a perfect score.
- BMI = Body mass index. A commonly used metric to estimate individual's adiposity, obtained by body weight in kilograms divided by the square of height in meters. Depending on BMI, individuals are typically classified into one of four categories: underweight (BMI<18.5), normal (BMI=18.5-24.9), overweight (BMI=25.0-29.9), and obesity (BMI>30). At times, the term "BMI" may refer to the BMI-based weight category rather than the raw BMI.

- CART = Classification and regression trees. A type of machine learning decision tree technique. Useful for both prediction and classification, similar to CHAID. Results are highly visual and typically easy to interpret. Typically uses binary split for growing trees, followed by pruning to prevent overgrowth. See: CHAID.
- CHAID = Chi-square automatic interaction detection. A type of machine learning decision tree technique. Useful for both prediction and classification, similar to CART. Results are highly visual and typically easy to interpret. Typically uses multi-way split and pre-pruning. See: CART.
- CV LASSO = Cross-validated LASSO. A specific form of LASSO, with K-fold validation to find the smallest MSPE. See: LASSO, MSPE
- E-cigarette = An electronic device that is intended to simulate tobacco smoking. Typically includes an atomizer, a power source, and a container for cartridge or liquid pods. Also known as electronic cigarette or electronic nicotine delivery systems (ENDS). Juul, Vuse, Pax Labs, and Altria Group are examples of brands that sell e-cigarettes, and those brand names are sometimes used synonymously with the term "e-cigarette."
- EVALI = E-cigarette or vaping product associated lung injury. An emerging form of acute lung injury that has been recently linked to the inhalation of vitamin E-acetate-laced THC commonly in vaporized marijuana. See: THC
- High risk = The term *high risk* is used to label one of the classes in sub-study 2, meaning that individuals within the class are likely to be concurrent users of almost all legal and illegal substances. The assignment of a *high risk* label to the group that engages in all co-occurring behavior was based on an existing study (Hautala, 2018)

- LASSO = Least absolute shrinkage and selection operator. A type of machine learning method useful for the analysis of high-dimension data, as it can simultaneously perform regularization and variable selection, thus improving prediction accuracy and interpretation.
- LCA = Latent class analysis. A subset of structural equational modeling (SEM), cross-sectional in typical nature, that can be used to find underlying groups or classes in multivariate categorical data. Longitudinal LCA (abbreviated as LLCA) was not used in this dissertation research.
- Marijuana = Conventionally, refers to the psychoactive dried resinous flower buds and leaves of the cannabis plant. May also refer to marijuana products, or any substance derived from the plant. Synonyms include, but are not limited to: pot, weed, cannabis, hemp, ganja, hashish, dope, grass, blow, and reefer.
- Marijuana vaping continuation = For the purpose of this research, an individual was considered to continually vape marijuana if they have initiated marijuana vaping at wave 4, and they have continued its use at wave 5. See: Table 2
- MSPE = Mean-squared prediction error. Measures the expected squared distance between predicted and true value for cross-validation of a prediction model. Models with lower values of MSPE are typically considered better.
- PATH = Population Assessment of Tobacco and Health study. Data source for this dissertation research. See: Chapter 3.
- Polydrug use = Using more than one drug together. Also known as poly-substance use. Concurrently and simultaneous use of different drugs, as well as alternating use, can be considered a type of polydrug use. Differentiating sub-types of polydrug use is often

difficult due to problems like measurement specification and survey recall. For the purpose of this research, individuals who reported the use of substances A and B during the same survey wave were considered concurrent users of substances A and B.

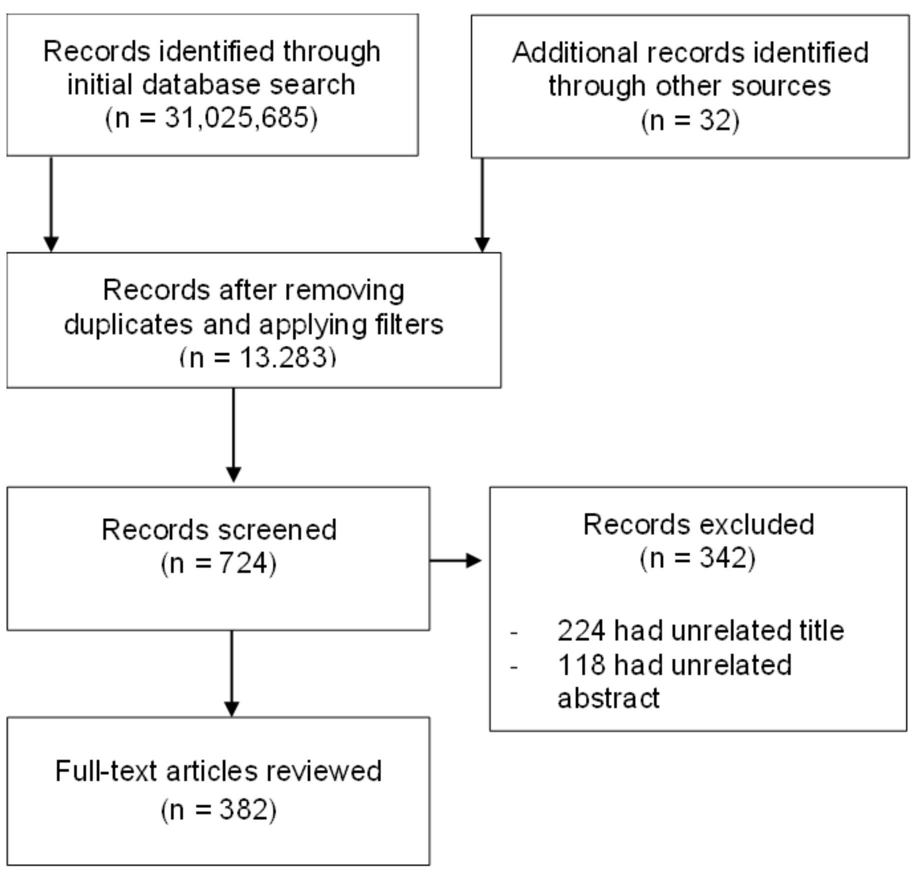
- RML = *recreational marijuana legalization* or *recreational marijuana law*. This is the term and abbreviation often found in the literature to indicate states that have legalized non-medical marijuana use for adults (e.g., "states with RML").
- SSC = Statistical Software Components from the Boston College. User-defined syntax developed by (and uploaded to) SSC allows STATA users to use new analytic methods that are not yet available in the official release of STATA.
- THC = Delta-9-tetrahydrocannabinol. The main psychoactive chemical ingredient of marijuana.
- Vaping = The act of inhaling and exhaling vapors, typically containing nicotine, flavoring, and/or liquid THC, produced by electronic devices such as e-cigarettes or *vape pen*s.
- Greek letters have been used in this dissertation. While these are commonly found in statistical analysis of studies in the literature, studies sometimes opt to write out these letters in long form, such that:
    - α = alpha
    - β = beta
    - γ = gamma
    - λ = lambda
    - ρ = rho
    - χ = chi

## Chapter 2

## LITERATURE REVIEW

A comprehensive literature review was conducted for each sub-study to examine information that is available in the literature and identify potential gaps in the literature. Online keyword search was used as the primary method of finding relevant studies via EBSCO Host for scholarly articles and Google for government documents and webpages. All databases within EBSCO Host were included. Only articles with full text in English published by November 2023 were reviewed.

**Figure 1**. Flow chart of literature search and review



Search keywords included a combination of the following: *marijuana, weed, pot, thc, *cannabinol, cannab*, hemp, dope, smok*, ends, cigar*, e-cig*, vap*, ml, machine*, learn*, lasso, cart, chaid, lca, sem, latent*, factor, predict*, covariate, preval*, onset, initi*, continu*, abus*, misus*, legal*, law*, and *rml.*

The process of literature search and review is summarized in Figure 1. Information gathered via literature review is summarized below.

*Sub-study 1 – Predictors of marijuana vaping*

Past studies (Han & Seo, 2022; Han et al., 2021; Loukas et al., 2019; Pericot-Valverde et al., 2019; Riehm et al., 2019; Spears et al., 2019) have identified clusters of variables that may be related to marijuana vaping, which include (1) socio-demographic characteristics (e.g., sex, race/ethnicity), (2) psychosocial factor (e.g. stress, bullying behavior), (3) digital media use (e.g., screen time, social media use), (4) substance use and misuse (e.g., cigarette, alcohol), (5) family and peer influence, (6) exposure to tobacco advertisement/promotion, (7) life quality and health, and (8) risk perception of substance use. However, many of these variables were adopted from studies that examined marijuana use and ENDS use, and only a limited number of studies have directly examined predictors of marijuana vaping. Existing studies have found that marijuana vaping among young adults is related to their other substance use and social network characteristics (Cassidy et al., 2018; Pokhrel et al., 2020; Lee et al., 2023). One study on U.S. adolescents found that marijuana vaping initiation is predicted by substance use history, peer influence, mental health, parent education, and Hispanic ethnicity (Lee et al., 2021). A longitudinal study found that past 30 days of marijuana vaping was influenced by cigarette and electronic cigarette use, as well as the number of cigarette and marijuana users with whom people spend the most time or talk to most often (Cassidy et al., 2018; Pokhrel et al., 2020). A different longitudinal study found that baseline e-cigarette use significantly predicts marijuana vaping and any form of marijuana use, while also suggesting that being non-Hispanic black, other tobacco use, alcohol use, perceiving e-cigarette use as having little or some harm, age, mental health problems, and state's legalization status of recreational marijuana are also related to marijuana vaping (Wang et al., 2022). The study also found that the association of e-cigarette use with marijuana vaping was not stronger than its association with other modes of marijuana use, and that additional studies are needed to explain the mechanisms linking e-cigarettes and marijuana (Wang et al., 2022). A later study on a nationally representative cohort of 9,828 youths affirmed those findings, and suggested that while e-cigarette use predicts future marijuana use, difference across modes of marijuana use is limited (Sun, Mendez, & Warner, 2022). These studies together provide a generous list of predictors. However, it should be noted that these studies were often limited by the non-representativeness of their sample and ad hoc covariate selection. Moreover, many of the existing studies have relied on regression-based approach to testing predictors (Boccio and Jackson, 2021; Cassidy et al., 2018; Kritikos et al., 2020; Lee et al., 2021; Pokhrel et al., 2020; Lee et al., 2023), which limit our understanding of characteristics and behaviors related to marijuana vaping. Regression-based approach cannot examine the complex and interactive effect between variables beyond two- or three-way interactions (Yong et al., 2020), while such is possible with machine learning methods.

A recent study by Han and Seo (2022) has provided much-needed evidence on marijuana vaping initiation among U.S. young adults. The study found that state legalization of recreational marijuana not only influenced people's risk perception and increased marijuana smoking, but it also significantly increased marijuana vaping among young adults (Han & Seo, 2022). Moreover, the study used machine learning to identify and examine variables not previously included in conventional models, though it had certain limitations. The first limitation of Han and Seo's study was that it lacked certain variables that are still potentially related to young adults' marijuana vaping. For example, sleep is a variable that may predict marijuana vaping, as it is a behavior that has been associated with electronic cigarette vaping and marijuana smoking (Jackson, Boccio,& Leal, 2020; Advani et al., 2022). As marijuana vaping has characteristics of electronic cigarette vaping and marijuana smoking, it makes sense to test whether or not sleep is related to marijuana vaping. However, the study only included sleep disorder as a variable, which is merely a proxy measure of individual's sleep.

Han and Seo (2022) also left housing, health insurance status, and BMI unexplored in their study. While housing, health insurance status, and BMI were never presented as strong predictors of marijuana vaping in previous studies, they have been presented as significant predictors of various substance-use behavior (Mojtabai at al., 2020; Foltin, Fischman, and Byrne, 1988; Beulaygue & French, 2016; Gentzke et al., 2018; Douglas, 2010). Housing, for example, can have a significant influence on individual's health, especially among collegiate young adults. This is because where and with whom college students reside can influence their peer relations and their chance of attending parties, which in turn influence their chance of substance experimentation (Anastasiou et al., 2020; DiBello et al., 2018; Lyke, 2018; Gentzke et al., 2018). Health insurance status coverage may be an important predictor of marijuana vaping in a similar regard. As health insurance coverage significantly influence whether or not individuals seek treatment for substance use disorders (Mojtabai at al., 2020), it may influence individual's decision to use substances that may lead to such illness. Lastly, BMI as a predictor of marijuana product use have been mixed. Older studies have found a positive association between marijuana and BMI, presumably because marijuana use and subsequent appetite stimulation lead to weight gain (Foltin, Fischman, and Byrne, 1988). However, more recent studies have found that marijuana product use is negatively associated with an individual's BMI (Beulaygue & French, 2016; Sansone & Sansone, 2014)). The relationship between BMI and marijuana use must be further explored to resolve the discrepancy in the literature

It should be noted that, in addition to the aforementioned limitations, existing marijuana vaping studies have varied outcome measures. The majority of existing studies have examined lifetime or ever-use of marijuana vaping, while others examined future initiation of marijuana vaping. Only one study has examined extended marijuana vaping, using a non-representative sample of Hawaiians (Pokhrel et al., 2020). Although predicting substance ever-use and initiation are important, more studies examining extended substance use are needed. This is because not everyone who initiate substance use continue its use (Vu et al., 2018; Bailey, Flewelling, & Rachal, 1992; Bray et al., 2021; Van den Bree & Pickworth, 2005). A proportion of individuals who reported marijuana vaping initiation may have been mere experimenters, which means that they are less likely to experience risks associated with prolonged exposure. Given how individuals with extended substance use are likely to experience most harm (Hasin et al, 2015; Winters & Lee, 2007), a nationally representative study examining extended marijuana vaping (rather than initiation) is warranted.

*Sub-study 1 – Initiating substance use vs. continued substance use*

Substance use studies in the literature have often focused on their initiation, because cessation is often difficult for addictive substances, which makes prevention of initiation the ideal strategy. However, a limited number of studies have examined continued substance use, given how there are individuals who are considered mere experimenters (i.e., individuals who initiate substance use, but only use it for a short period of time and do not return to its use). For example, Ansari-Moghaddam et al (2012) examined predictors of initiation, continuation and transition of drug use in south-eastern Iran, and found that the age of the first drug use and marital status were significant predictors of continued drug use, while predictors differed for drug initiation, transition, and cessation. Tsering and Pal (2009) studies high school students in India, and found that drug initiation was largely predicted by family and peers influence, while drug continuation was predicted by easy availability in the neighborhood.

Recent studies have examined predictors of continued e-cigarettes and marijuana use in the U.S. For example, Mantey et al (2022) conducted a longitudinal study of Texan adolescents to compare never-users, ever-users, and current-users of e-cigarettes, and found that perceived stress predicted greater risk of e-cigarette continuation, but not initiation. A systematic review by Notley et al (2021) examined patterns of use of flavored e-cigarettes and their associations with continued vaping, tobacco smoking uptake, or cessation, and found that flavoring is considered important for both initiation and continuation, while nicotine dependence was important only for continuation. Choi, DiNitto, and Marti (2018) used waves 1-2 of the PATH data to examine marijuana use initiation, return to use, and continuation. The study found substantial movement between marijuana nonuse and use during the 12-month follow-up. The study also found that other substance use, initiation of marijuana use before the age of 18, and frequent marijuana use at baseline predicts continued marijuana use. However, it is unclear whether such study on conventional marijuana use is generalizable to marijuana vaping as well. To the best of my knowledge, no studies to date have rigorously examined continuation of marijuana vaping among U.S. young adults.

*Sub-study 1 – Past use of machine learning*

Machine learning is a novel analytic tool in public health. It is particularly useful for uncovering new and emerging predictors to build a robust prediction model, out of many potential predictors in a high-dimension data (Ahn et al, 2016). It is also useful for analyzing imbalanced (i.e., skewed) or highly-correlated data that may lead to inflated Type II errors (Ahn et al, 2016; Volkow, Koob, & Baler, 2015). As such, machine learning models often perform better at prediction than conventional statistical models with imbalanced data (Gradus et al., 2019; Han, Lee, & Seo, 2020). Using conventional regression methods to analyze imbalanced data is known to have low prediction capacity (Wiles, 2006), but the method has been widely used, and the problem has been largely ignored. While predictive power can decrease with machine learning methods when data are highly imbalanced (Krawczyk, 2016), it still shows better predictive performance than conventional models (Alghamdi et al., 2017; Sakr, 2018; Chiew, 2019). This is an advantage that may be particularly useful in substance use research because individuals reporting substance use often represent a small proportion of the data and make the data imbalanced (Mak, Lee, & Park, 2019; Loughran, Larroulet, & Thornberry, 2018).

In addition to being robust to data imbalance, machine learning techniques have been shown to mitigate problems arising from multicollinearity and overfitting, at least more so than conventional regression approach to data analysis (Altman & Krzywinski, 2018; Yarkoni & Westfall, 2017). As hinted earlier, this makes machine learning especially useful for analyzing high-dimension data. One of the ways machine learning mitigates such problems is by imposing penalties for having too many variables and shrinking regression coefficients of the less explanatory variables toward zero. This is known as regularization, and is found in penalized regression models such as ridge regression and Least Absolute Shrinkage and Selection Operator (LASSO) regression. Ridge regression performs best when the outcome is a function of many predictors, all with coefficients of roughly equal size, because ridge regression tends to shrink coefficients towards zero, but it does not set any of them to exactly zero (Gareth et al., 2013). On the other hand, LASSO regression can impose penalty that force coefficients with minor contribution to the model to be exactly zero, which produces simpler and more interpretable models (Gareth et al., 2013). The optimal model selection can be done using cross-validation. Regardless of which specific method you use, machine learning with penalized regression can substantially reduce the variance of the model without a sizable increase in bias (Yarkoni & Westfall, 2017).

Given its usefulness, the National Institute on Drug Abuse's guideline suggested that machine learning is one of the most promising approaches in classifying at-risk populations with high-dimensional data as well as in identifying predictive factors for psychiatric and substance use disorders (Volkow, Koob, & Baler, 2015). Mak et al. (2019) suggested, in their systematic review, that machine learning would play a crucial role in identifying and profiling behavioral groups on substance addiction. Despite such recommendations, machine learning has been under-utilized in examining marijuana vaping.

Past studies have used machine learning to examine conventional tobacco use behaviors. For example, Coughlin et al. (2020) used classification and regression trees (CART) to identify clinical and psychosocial factors that predict the success of cognitive-behavioral therapy for cigarette smoking cessation. Dumortier et al. (2016) also used classification tree to analyze 248 adult smokers seeking to quit smoking and established up to 86% of classification accuracy. However, many of existing studies were conducted in clinical settings, and the samples were too small or non-representative to generalize its findings to a larger population.

Past studies have used machine learning to examine e-cigarette use and marijuana use behaviors as well. For example, Han et al. (2021) used penalized regression to identify emerging predictors of e-cigarette use among a representative sample of U.S. adolescents. Parekh and Fahim (2021) used machine learning to predict daily marijuana with high efficiency, and found current use of e-cigarette and combustible cigarette use, male gender, unmarried, poor mental health, depression, cognitive decline, abnormal sleep pattern, and high-risk behavior as most important predictors of daily marijuana use. A recent study by Han and Seo (2022) examined marijuana vaping using LASSO and CART, but studies that use machine learning to examine marijuana vaping are still very sparse.

*Sub-study 2 – Marijuana and polydrug use*

Marijuana vaping is a threat to public health that warrant further investigation (Daniulaityte et al., 2017; NIDA 2020; Steigerwald et al., 2018). Studies have examined predictors of marijuana vaping, though those studies often had known limitations. Many of the marijuana vaping studies to date have examined marijuana vaping as a single-use substance, though individuals who vape marijuana tend to not vape marijuana alone (Steigerwald et al., 2018; Chadi, Minato, & Stanwick, 2020). It has been suggested that the majority of individuals who vape marijuana tend to use at least one other substance concurrently or sequentially, though the pattern and the extent of concurrent/sequential use is unclear (Steigerwald et al., 2018; Chadi, Minato, & Stanwick, 2020).

A cross-sectional study of California adolescents found that actual and perceived socio-economic status are related to single-, dual-, and multiple substance use, and parental socio-economic status predicts nicotine and marijuana products (Bello et al., 2019). However, this study examined polydrug use by having the number of substances used as its regression outcome, and it does not clearly answer which products are likely used together. A different study found that 55% of all marijuana users only smoke marijuana, 13% use a combination of smoking, vaping, and edibles, while only 3% use vaporized form only (Steigerwald et al., 2018). However, findings of this study was limited, as it sampled relatively small number of general adult population, and the study only considers co-occurrence of marijuana products (i.e., it does not consider polydrug use with alcohol or nicotine products). Some studies have used latent class analysis to examine patterns of polydrug use among young adults (Lanza, Motlagh, & Orozco, 2020; Linden-Carmichael et al., 2022; Park & Kim, 2018; Mattingly, Elliott, & Fleischer, 2023), but most of them do not sufficiently differentiate electronic cigarette vaping (i.e., nicotine vaping) from marijuana vaping.

*Sub-study 2 – Predictors of polydrug use*

Jessor's Problem Behavior Theory (PBT) is a conceptual framework that may be useful in understanding substance use, abuse, and dependence as a form of risk-taking behavior (Knight, 2009). PBT includes personality, behavioral, and environmental constructs that predict future substance use, particularly among adolescents (Knight, 2009). PBT may serve as a starting point for variable selection, as individuals' characteristics that fit any one of these constructs are potential predictors of substance use.

Numerous studies have examined predictors of polydrug use. For example, Collins, Ellickson, and Bell (1998) found that the best predictors of concurrent use of alcohol and marijuana were pro-drug environment, pro-drug beliefs, social deviance, and family disruption. Vaughn et al (2009) utilized the National Longitudinal Study of Adolescent Health (Add Health) data to identify gene-environment interplay and self-control as important predictors of polydrug use. Kelly and Parsons (2009) found male gender as a significant predictor of polydrug cocaine use, but race and sexual orientation as insignificant predictors in club subcultures. A 12-year longitudinal study by Galaif and Newcomb (1999) found that teenage polydrug use was a significant predictor of adult polydrug use among white, black, and Latino individuals, but not among Asians, suggesting that prevention strategies should consider population diversity and factors that uniquely affect each racial/ethnic groups. Mattingly, Elliott, and Fleischer (2023) conducted latent class analysis and multinomial regression of U.S. youth and young adults and found that compared to non-Hispanic white young adults, non-Hispanic black young adults have higher odds of *cigarettes + cigar use*, and racial/ethnic minorities generally have lower odds of other poly-substance use.

*Sub-study 2 – Past use of latent class analysis*

Latent class analysis is a statistical procedure that can identify subgroups with shared underlying characteristics (Weller, Bowen, & Faubert, 2020). Studies have used latent class analysis to examine patterns of polydrug use among young adults (Lanza, Motlagh, & Orozco, 2020; Linden-Carmichael et al., 2022; Park & Kim, 2018). However, almost none of them sufficiently differentiate electronic cigarette vaping (i.e., nicotine vaping) from marijuana vaping. For example, a study by Lanza et al. (2020) focused explicitly on co-use and correlates of nicotine vaping. A study by Linden-Carmichael et al. (2022) included various substance use in its latent class analysis but measured non-specific *cannabis use* instead of marijuana vaping. One study (Mattingly, Elliott, & Fleischer, 2023) does sufficiently differentiate nicotine vaping from marijuana vaping, but the study was limited in three other ways: (1) it only examined combination of nicotine and marijuana products, despite how alcohol and illegal substances may also be used concurrently; (2) it used data from wave 4 of the PATH study instead of the most recent (wave 6) data; and (3) it failed to consider state legalization of non-medical marijuana and its potential impact on their multinomial model.

Existing latent class analysis studies have often only tested a small number of characteristics as predictors of class membership, while a more comprehensive list of predictors may help identify individuals at the highest risk. For example, housing is an ecological variable that has not been rigorously examined in polydrug studies, though it has the potential to influence young adult's peer relations, which in turn influence their chance of multiple substance use (Anastasiou et al., 2020; DiBello et al., 2018; Lyke, 2018).

## Chapter 3.

## METHODS

The main analyses of this dissertation research utilized machine learning techniques and structural equation modeling. Specifically, sub-study 1 utilized various Least Absolute Shrinkage and Selection Operator (LASSO) models for variable selection and Classification and Regression Trees (CART) for predictor classification, while sub-study 2 utilized Latent Class Analyses models to assess patterns of polydrug use among U.S. young adults.

All analyses were conducted using STATA 17 MP (Stata Press, College Station, TX), and relevant user-defined ADO packages, available for download at the Boston College Statistical Software Components (SSC) archive and the Latent Class Analysis Knowledge Base (https://www.latentclassanalysis.com), installed for the Virtual Data Enclave (VDE) by the Inter-University Consortium For Political And Social Research (ICPSR). This study was verified as a not human subjects research by the Indiana University's Institutional Review Board (IRB protocol #18942; See **Appendix A**).

Detailed methods of the sub-studies are presented below. Please note that methods presented below may be similar those found in the literature, as sub-study 1 was modelled after Han and Seo's (2022) study, and sub-study 2 followed the LCA Plugin Users' Guide by Lanza et al (2018).

### *Sub-study 1 – Study design and participants*

This study utilized data from the most recent three waves (Waves 4-6, December 2016-November 2021) of the Population Assessment of Tobacco and Health (PATH) Study (US

DHHS, 2023). PATH Study is a nationally representative longitudinal cohort study designed to collect data on patterns of nicotine product use and related health outcomes among non-institutionalized U.S. residents aged 12 and above. Information in the PATH data is not limited to mere nicotine product use, and information such as respondents' demographic characteristics and other substance use are also available. Only three waves of the PATH data (i.e., waves 4-6) were included for analysis, because wave 4 cohort was the first complete sample to provide information related to adult marijuana vaping. Additional details about the PATH data can be found elsewhere (Hyland et al., 2017).

This study utilized the restricted access files of the PATH data (US DHHS, 2023), given how state identifiers are only available in the restricted access files of the PATH data. Data with state identifiers was accessed via the Inter-university Consortium for Political and Social Research (ICPSR), as done by a previous study (Han & Seo, 2022). State identifiers were used to determine and stratify by legal status of non-medical marijuana. There were eight states that had legalized non-medical marijuana before the study period and ten states that underwent legalization during the study period. To prevent significant sample reduction and account for likely anticipatory effect of legalization (Sheehan et al., 2021), states that have completed legalization and states that underwent legalization were both considered as legal states (AZ, CA, CO, CT, DC, IL, MA, ME, MI, MT, NJ, NM, NY, OR, VA, VT, WA), while others were considered as illegal states (AL, AR, FL, GA, HI, IA, ID, IN, KS, KY, LA, MD, MN, MO, MS, NC, NE, NH, OH, OK, PA, RI, SC, SD, TN, TX, UT, WI, WV, WY). Sensitivity analyses using models with different cutoffs for legal-illegal differentiation produced similar results or failed to converge due to small cell size. The study sample included all young adults with non-missing data on marijuana vaping at waves 4-6.

*Sub-study 2 – Study design and participants*

This study used data from the most recent wave (Wave 6, March 2021 - November 2021) of the Population Assessment of Tobacco and Health (PATH) Study (US DHHS, 2023). As mentioned earlier, PATH Study is a nationally representative longitudinal cohort study designed to collect data on patterns of nicotine product use and related health outcomes among non-institutionalized U.S. residents aged 12 or more years. Information available in the PATH data is not limited to mere nicotine product use, and information such as respondents' demographic characteristics and other substance use are also available. Only the most recent wave (wave 6) of the data was used, as this study is cross-sectional in nature. Additional details about the PATH data can be found in a previous study (Hyland et al., 2017).

This study utilized the restricted access files of the PATH data (US DHHS, 2023), given how state identifiers are only available in the restricted access files of the PATH data. Data with state identifiers was accessed via the Inter-university Consortium for Political and Social Research (ICPSR), as done by a previous study (Han & Seo, 2022). State identifiers were used to determine and stratify by legal status of non-medical marijuana. There were 14 states that had legalized non-medical marijuana before the study period and four states that underwent legalization during the study period. To prevent significant sample reduction and account for likely anticipatory effect of legalization (Sheehan et al., 2021), states that have completed legalization and states that underwent legalization were both considered as legal states (AK, AZ, CA, CO, CT, DC, IL, MA, ME, MI, MT, NJ, NM, NV, NY, OR, VA, VT, WA), while others were considered as illegal states (AL, AR, FL, GA, HI, IA, ID, IN, KS, KY, LA, MD, MN, MO, MS, NC, ND, NE, NH, OH, OK, PA, RI, SC, SD, TN, TX, UT, WI, WV, WY). Sensitivity analyses using models with different cutoffs for legal-illegal differentiation produced similar results or failed to converge due to small cell size. The study sample included all young adults with non-missing data on substance use at wave 6.

*Sub-study 1 – Outcome and predictor measures*

Outcome variable was multi-year marijuana vaping, a binary measurement of whether respondent vaped marijuana for at least two years. Multi-year marijuana vaping was generated by first creating dummy variables for current marijuana vaping at each survey wave (based on the survey item: *when did you last use marijuana, marijuana concentrates, marijuana waxes, THC, or hash oils in an electronic product such as an e-cigarette, vape, mod, personal vaporizer, e-hookah, or hookah pen?*), and then adding the dummy variables to identify individuals who vaped marijuana at more than two waves of data. Other measures of marijuana vaping have been considered during preliminary analyses, but were dismissed for having too many missing values or having inconsistent survey item across waves.

Predictor variables were selected based on what was done by a previous study (Han & Seo, 2022), followed by addition of three unique variables (i.e., housing, insurance status, and BMI). A total of 54 predictor variables were included from Wave 4, which were categorized under socio-demographics, psychosocial factors, digital media use, substance use, family/peer influence, tobacco advertisement/promotion, life quality and health, and risk perception. A full list of predictor variables, along with outcome and weight variables, is provided in **Table 2**. BMI was originally a continuous variable, recoded as a 4-category variable (1=underweight, 2=normal, 3=overweight, and 4=obese) based on the Center of Disease Control and Prevention's adult weight classification (Garrow &Webster, 1985; CDC, 2022). All other variables were binary and categorical in the original PATH data, and were analyzed without recoding. All variables were manually examined for numeric missing values (e.g., missing responses coded as "-99"), and replaced by blanks ("." in STATA) when considered appropriate by the data's documentation.

**Table 2.** List of variables included in sub-study 1

| | Category | Variable | Variable name in data |
|---|---|---|---|
| 1 | Socio-demographics | Sex | R04R_A_SEX_IMP |
| 2 | | Sexual orientation | R04R_A_SEXORIENT2 |
| 3 | | Race | R04R_A_RACECAT3 |
| 4 | | Hispanic ethnicity | R04R_A_HISP_IMP |
| 5 | | Household income | R04R_A_AM0030 |
| 6 | | Employment status | R04_AM0014 |
| 7 | | Health insurance[1] | R04R_A_AM0026_V2 |
| 8 | | Current housing[1] | R04R_A_AM0042 |
| 9 | | Wave 4 state identifier[2] | R04_A_STATE |
| 10 | | Degree program enrollment status | R04_AM0019 |
| 11 | Psychosocial factors | Feeling very trapped, lonely, sad, blue, depressed or hopeless about the future | R04_AX0161 |
| 12 | | Sleep trouble - such as bad dreams, sleeping restlessly or falling asleep during the day | R04_AX0162 |

| | Category | Variable | Variable name in data |
|---|---|---|---|
| 13 | | Feeling very anxious, nervous, tense, scared, panicked or like something bad was going to happen | R04_AX0163 |
| 14 | | Becoming very distressed and upset when something reminded you of the past | R04_AX0164 |
| 15 | | Lied or conned to get things you wanted or to avoid having to do something | R04_AX0165 |
| 16 | | Had a hard time paying attention at school, work or home | R04_AX0166 |
| 17 | | Had a hard time listening to instructions at school, work or home | R04_AX0167 |
| 18 | | Were a bully or threatened other people | R04_AX0168 |
| 19 | | Started physical fights with other people | R04_AX0169 |
| 20 | | Felt restless or the need to run around or climb on things | R04_AX0250 |
| 21 | Digital media use | Amount of time spent watching TV or videos that include commercials on an average weekday | R04_AX0494 |
| 22 | | Amount of time spent listening to the radio or online radio stations that include commercials on an average weekday | R04_AX0060 |
| 23 | | How often visit social media accounts | R04_AX0317 |

| | Category | Variable | Variable name in data |
|---|---|---|---|
| 24 | Substance use | Current established cigarette smoker | R04R_A_CUR_ESTD_CIGS |
| 25 | | Current established electronic nicotine product user | R04R_A_CUR_ESTD_EPRODS |
| 26 | | Current established cigar smoker | R04R_A_CUR_ESTD_CIGAR |
| 27 | | Current established pipe smoker | R04R_A_CUR_ESTD_PIPE |
| 28 | | Current established hookah smoker | R04R_A_CUR_ESTD_HOOK |
| 29 | | Current established snus pouch user | R04R_A_CUR_ESTD_SNUS |
| 30 | | Current established smokeless tobacco user | R04R_A_CUR_ESTD_SMKLS |
| 31 | | Past-year alcohol use | R04_AX0673_12M |
| 32 | | Past-year marijuana use | R04_AX0085_12M |
| 33 | | Past-year use of prescription drugs not prescribed to you: Ritalin or Adderall | R04_AX0089_12M_01 |
| 34 | | Past-year use of prescription drugs not prescribed to you: Painkillers, sedatives or tranquilizers | R04_AX0089_12M_02 |
| 35 | | Past-year use of drugs: Cocaine or crack | R04_AX0220_12M_01 |

| | Category | Variable | Variable name in data |
|---|---|---|---|
| 36 | | Past-year use of drugs: Stimulants like methamphetamine or speed | R04_AX0220_12M_02 |
| 37 | | Past-year use of drugs: Any other drugs like heroin, inhalants, solvents or hallucinogens | R04_AX0220_12M_03 |
| 38 | Family/peer influence | Anyone who lives with you now use any tobacco products | R04R_A_AX0066 |
| 39 | | Lax rules about smoking a combustible tobacco product inside your home | R04_AR1045 |
| 40 | | Lax rules about using non-combustible tobacco products inside your home | R04_AR1050 |
| 41 | | Lax rules about using e-cigarettes and other electronic nicotine products inside your home | R04_AR1051 |
| 42 | Tobacco advertisement /promotion | In the past 12 months, received discounts or coupons for any of the following products: Cigarettes | R04_AX0708_01 |
| 43 | | In the past 12 months, received discounts or coupons for any of the following products: E-cigarettes or other electronic nicotine products (including e-liquid) | R04_AX0708_02 |
| 44 | | In the past 12 months, received discounts or coupons for any of the following products: Cigars | R04_AX0708_03 |

| | Category | Variable | Variable name in data |
|---|---|---|---|
| 45 | | In the past 12 months, received discounts or coupons for any of the following products: Shisha or hookah tobacco | R04_AX0708_04 |
| 46 | | In the past 12 months, received discounts or coupons for any of the following products: Snus | R04_AX0708_05 |
| 47 | | In the past 12 months, received discounts or coupons for any of the following products: Other types of smokeless tobacco | R04_AX0708_06 |
| 48 | Life quality and health | Self-perception of place on social ladder relative to other people in the United States | R04_AM0040 |
| 49 | | Self-perception of quality of life | R04_AX0185 |
| 50 | | Level of satisfaction with social activities and relationships | R04_AX0092 |
| 51 | | How many days in typical week you do any physical activity or exercise of at least moderate intensity | R04_AX0241 |
| 52 | | How many days in a typical week you do physical activities specifically designed to strengthen your muscles | R04_AX0243 |
| 53 | | Weight classification based on Body Mass Index (BMI) [1] | R04R_A_BMI |

| | Category | Variable | Variable name in data |
|---|---|---|---|
| 54 | Risk perception | General perception: Harmfulness of using e-cigarettes or other electronic nicotine products to health | R04_AE9050 |
| | Sample | Age category | R04R_A_AGECAT7 |
| | Outcome | Multi-year marijuana vaping, derived from when last vaped marijuana | R04-R06_AM0100 |
| | Weights | Wave 6 Adult All-waves Longitudinal Weight for the Wave 4 Cohort | R06_A_A04WGT |
| | | Wave 6 Adult All-waves Longitudinal Replicates for the Wave 4 Cohort | R06_A_A04WGT1 - R06_A_A04WGT100 |

*Note.* Data from Population Assessment of Tobacco and Health (PATH) study

[1]Variables not examined by Han & Seo (2022)

[2]Variable obtained from the restricted PATH data

*Sub-study 2 – Outcome and predictor measures*

This study examined the past-month use of a total of 16 legal and illegal substances as outcome indicators: alcohol, cigarette, e-cigarette, cigar, pipe, hookah, snus, smokeless tobacco, joint marijuana, blunt marijuana, marijuana vaping, Adderall, painkiller, cocaine, methamphetamine, and heroin, all available in the PATH dataset with binary response options (yes/no). Substances with low number of users (pipe, hookah, snus, smokeless tobacco, Adderall, painkillers, cocaine, methamphetamine, and heroin) prevented proper convergence of the preliminary analytic model, and were combined with other similar substances for the final analysis. The final model included past-month marijuana vaping, past-month joint marijuana smoking, past-month blunt marijuana smoking, past-month alcohol use, past-month cigarette smoking, past-month e-cigarette use, past-month cigar smoking, past-month other tobacco product use, past-month misuse of prescription drugs, and past-month use of illegal drugs, as shown in **Table 3.**

**Table 3.** List of substance variables included in sub-study 2

| Variable | Variable name in data |
|---|---|
| Past-month marijuana vaping | R06_AX0290_03 |
| Past-month joint marijuana smoking | R06_AX0290_01 |
| Past-month blunt marijuana smoking | R06_AX0290_02 |
| Past-month alcohol use | R06_AX0673 |
| Past-month cigarette smoking | R06R_A_P30D_CIGS |
| Past-month electronic nicotine product (e-cigarette) use | R06R_A_P30D_EPRODS |
| Past-month cigar smoking | R06R_A_P30D_CIGAR |
| Past-month other tobacco product use (pipe; hookah; snus; smokeless tobacco) | R06R_A_P30D_PIPE; R06R_A_P30D_HOOKAH; R06R_A_P30D_SNUS; R06R_A_P30D_SMKLS |
| Past-month use of prescription drugs not prescribed to you (Ritalin, Adderall, painkillers, sedatives or tranquilizers) | R06_AX0676_01-03 |

| Variable | Variable name in data |
| --- | --- |
| Past-month use of illegal drugs (Cocaine or crack; stimulants like methamphetamine or speed; any other drugs like heroin, inhalants, solvents or hallucinogens) | R06_AX0676_04-06 |

*Note.* Data from Population Assessment of Tobacco and Health (PATH) study

Predictor selection was based on Jessor's Problem Behavior Theory (PBT) constructs (i.e., personality, behavioral, and environmental constructs predicting future substance use) (Knight, 2009). The final model included sex, race/ethnicity, sexual orientation, housing, employment status, degree enrollment, health insurance status, and perceived quality of life as part of personality, behavioral, and environmental PBT constructs that predict and reinforce young adults' substance use behavior. BMI, past-year use of substances, and perceived harm of substances were considered but removed from the final model to optimize model convergence and prevent severe multicollinearity. Most of the variables were used as presented in the original PATH dataset, though housing, employment, and quality of life had categories collapsed due to small cell size. Race/ethnicity variable was created using separate race and ethnicity variables in the PATH dataset, such that individuals were categorized as one of four racial/ethnic groups: non-Hispanic white, non-Hispanic black, Hispanic, or other.

*Sub-study 1 – Statistical analysis*

Statistical analyses of this study were modelled after Han and Seo's (2022) study on marijuana vaping initiation. For descriptive statistics, baseline characteristics for selected variables were reported with unweighted frequency and weighted percentages. Pearson's chi-squared test was used to test the association between variables and the legal status of non-medical marijuana, and variables with significant p-values were reported.

Two-stage machine learning approach were used for the main analysis, as suggested by previous studies (Suchting et al., 2018; Walss-Bass et al., 2018; Walters et al., 2021). These two stages were necessary, because machine learning identifies new predictors and builds a robust prediction model out of many potential predictors of interests or with imbalanced or highly-correlated data, which may inflate type II error and require steps of reduction for proper identification. These two stages were chosen over other supervised machine learning methods, because they required minimal supervision and relatively low computing power while producing accurate and interpretable results. The first stage of the analysis involved fitting least absolute shrinkage and selection operator (LASSO) models to identify subset of variables most related to extended marijuana vaping, and reduce biases and imbalances (Altman and Krzywinski, 2018; Mak et al., 2019). To illustrate how LASSO models performs better than other models of variable selection, I also conducted stepwise regression with backward elimination (threshold=0.20) and ridge regression, and reported their results. Model diagnostic statistics were also be reported, to illustrate in detail how model and variable selections were made. For each LASSO model, the top 10 performing predictors were selected based on the scaled importance, unless less than 10 predictors are selected by the model. Scaled importance was obtained by dividing each coefficient by the highest-magnitude coefficient.

The second stage of the analysis involved the classification and regression tree (CART) to predict multi-year marijuana vaping (Breiman et al., 1984; Tuffery, 2011; Yong et al.,2020), using variables selected from the first stage. 10-fold cross-validation was used to prioritize out-of-sample performance, and guard against overfitting that hinders out-of-sample estimation

(McBride & Nichols, 2018). Gini impurity measure was used to compute optimal split and store option in the prediction matrix. The process is looped until no nodes remain eligible for splitting, or maximum tree depth is reached. Nodes are considered ineligible for splitting if all observations at a node are in one class, have zero variance, or have too few observations (Hastie, Tibshirani, & Friedman, 2009). Stratified CART was performed to evaluate whether the decision-making process differs by legal status of non-medical marijuana.

All analyses were conducted using STATA version 17 MP (Stata Press, College Station, TX), including user-defined *lassopack, plogit, crtrees,* and *cart* packages from the Statistical Software Components (SSC). These STATA packages have been modelled after *glmnet* package in R. To account for different probabilities of sample selection, all analyses have been weighted based on PATH data documentation. Missing data in the predictor variables (25.7%%, n = 1,908) were imputed using the multiple imputation with chained equations (MICE), similar to what was done in a previous study (Han & Seo, 2022). Specifically, the procedure used STATA's -*mi*- syntax to generate 5 sets of imputed data with 40 iterations per set, without the use of an auxiliary variable. MICE was used over other imputation method for not requiring a certain pattern of missingness (Jakobsen et al, 2017).

*Sub-study 2 – Statistical analysis*

Statistical analysis and reporting for this study followed guidelines set by Weller, Bowen, and Faubert (2020). For descriptive statistics, characteristics for all variables were reported with unweighted frequency and weighted percentages, stratified by legal status of non-medical marijuana. Pearson's chi-squared test was used to test the association between variables and the legal status of non-medical marijuana, and variables with significant p-values were reported. For the main analysis, latent class analysis (LCA) was used to identify substance use classes and examine how some substances are used with other substances. LCA has been used by previous polydrug use studies, as it uses underlying patterns of observations to group individuals with similar responses, behaviors, or characteristics into latent classes. LCA began by specifying the unconditional model (i.e., model with one class), which was then compared to subsequent models with increasing number of classes. Full information maximum likelihood estimation was used for iteration process. As suggested in the literature (Ramaswamy et al, 1993), model with optimal number of classes was determined by comparing fit statistics, including AIC, BIC, bootstrapped likelihood-ratio test (BLRT). Because BLRT is not available in STATA's standard syntax for LCA, -*gsem*-, I used a modified version of an external plug-in -*LCAbootstrap*- (LCABootstrap SAS Macro (Version 4.0) [Software]. (2016). University Park: The Methodology Center, Penn State).

Latent class posterior probabilities were used to create a latent class variable, which was then followed by multinomial logistic regression to find predictors of class membership. The aforementioned LCA and regression protocols were repeated with stratification to examine heterogeneity by legal status of non-medical marijuana. Unless otherwise noted, level of significance was set at $p < 0.05$ for all statistical tests. Item response probability graph was generated using Microsoft Excel, Office 2019. All other analyses, including the user-defined syntax -*doLCA*- and the external plug-in -*LCAbootstrap*-, were conducted using STATA version 17 (Stata Press, College Station, TX).

## Chapter 4

## ANALYSIS OF DATA

*Sub-study 1 – Descriptive statistics*

**Table 4** summarizes the characteristics of the analytic sample (age, mean [SD] = 21.5 [2.29] years) by legal status of non-medical marijuana. Of the 7,417 young adults, 4,123 lived in states where non-medical marijuana use was considered legal, and 3,294 lived in states where non-medical marijuana use was considered illegal. Four percent of those who lived in legal states reported multi-year marijuana vaping (n = 127), while two percent of those who lived in illegal states reported multi-year marijuana vaping. (n = 76). Pearson's $\chi$ test revealed significant differences in multi-year marijuana vaping, as well as race, Hispanic ethnicity, household income, degree enrollment, health insurance coverage, housing, past-year alcohol use, current cigarette use, e-cigarette use, cigar use, pipe use, hookah use, smokeless tobacco use, past-year marijuana use, cocaine use, methamphetamine use, coupons for Hookah, BMI weight classification, and starting physical fights (*p*'s <0.05) by legal status of non-medical marijuana.

**Table 4.** Weighted descriptive statistics for selected variables in sub-study 1

| Variable | Non-medical marijuana legal states (N = 4,123) | Non-medical marijuana illegal states (N = 3,294) |
|---|---|---|
| Multi-year marijuana vaping* | 127 (3.87%) | 76 (1.97%) |

| Variable | Non-medical marijuana legal states (N = 4,123) | Non-medical marijuana illegal states (N = 3,294) |
|---|---|---|
| ***Sociodemographics*** | | |
| Male | 1,600 (50.7%) | 1,937 (50.8%) |
| Race* | | |
| White | 2,231 (66.8%) | 2,772 (72.9%) |
| Black | 419 (11.9%) | 883 (17.9%) |
| Asian | 186 (9.98%) | 84 (3.26%) |
| Others | 458 (11.3%) | 384 (5.92%) |
| Hispanic* | 1,116 (26.6%) | 880 (16.9%) |
| Income* | | |
| <$10,000 | 583 (18.4%) | 916 (22.2%) |
| $10,000-24,999 | 646 (21.4%) | 893 (24.4%) |

| Variable | Non-medical marijuana legal states (N = 4,123) | Non-medical marijuana illegal states (N = 3,294) |
|---|---|---|
| $25,000-49,000 | 602 | 808 |
| | (19.8%) | (21.8%) |
| $50,000-99,999 | 604 | 687 |
| | (21.4%) | (18.7%) |
| $100,000+ | 561 | 455 |
| | (19.0%) | (13.0%) |
| Employment status | | |
| Work full-time (35+ hours per week) | 1,059 | 1,437 |
| | (35.6%) | (39.2%) |
| Work full-time (15-34 hours per week | 767 | 1,000 |
| | (22.8%) | (22.7%) |
| Work part-time (less than 15 hours per week | 327 | 356 |
| | (9.31%) | (8.43%) |
| Do not work for pay | 1,118 | 1,309 |
| | (32.3%) | (29.7%) |
| Enrolled in a degree program* | 1,372 | 1,613 |
| | (42.6%) | (38.2%) |
| Health insurance coverage* | 2,694 | 3,114 |

| Variable | Non-medical marijuana legal states (N = 4,123) | Non-medical marijuana illegal states (N = 3,294) |
|---|---|---|
| | (85.8%) | (79.0%) |
| Housing* | | |
| Apartment, condo, house | 2,504 | 2,975 |
| | (84.9%) | (81.3%) |
| Dorm, campus housing, fraternity/sorority | 222 | 285 |
| | (6.38%) | (7.06%) |
| Someplace else | 277 | 446 |
| | (8.71%) | (11.7%) |
| ***Substance use*** | | |
| Past-year alcohol use* | 2,219 | 2,456 |
| | (74.9%) | (66.9%) |
| Current cigarette use* | 420 | 638 |
| | (12.7%) | (17.1%) |
| Current e-cigarette use* | 218 | 263 |
| | (6.83%) | (6.86%) |
| Current cigar use* | 109 | 203 |
| | (3.20%) | (4.62%) |
| Current pipe use* | 14 | 19 |

| Variable | Non-medical marijuana legal states (N = 4,123) | Non-medical marijuana illegal states (N = 3,294) |
|---|---|---|
| | (0.50%) | (0.54%) |
| Current hookah use* | 114 | 75 |
| | (3.83%) | (1.83%) |
| Current snus use | 20 | 30 |
| | (0.56%) | (0.94%) |
| Current smokeless tobacco use* | 58 | 113 |
| | (1.95%) | (3.43%) |
| Past-year marijuana use* | 1322 | 1,304 |
| | (39.1%) | (30.2%) |
| Past-year painkiller use | 189 | 257 |
| | (5.91%) | (6.23%) |
| Past-year misuse of prescription drugs like Adderall | 147 | 174 |
| | (4.59%) | (4.01%) |
| Past-year Cocaine use* | 171 | 117 |
| | (5.73%) | (3.42%) |
| Past-year methamphetamine use* | 33 | 62 |
| | (1.05%) | (1.69%) |
| Past-year heroin use | 135 | 133 |

| Variable | Non-medical marijuana legal states (N = 4,123) | Non-medical marijuana illegal states (N = 3,294) |
| --- | --- | --- |
| | (4.19%) | (3.41%) |
| ***Perception, psychosocial factors, advertisement, and physical health*** | | |
| Received discounts and coupons for Hookah* | 24 | 24 |
| | (0.97%) | (0.46%) |
| Received discounts and coupons for cigars | 36 | 60 |
| | (0.96%) | (1.46%) |
| Perceived harmfulness of e-cigarettes | | |
| Not at all harmful | 77 | 130 |
| | (2.26%) | (3.06%) |
| Slightly harmful | 404 | 530 |
| | (12.0%) | (13.2%) |
| Somewhat harmful | 987 | 1,218 |
| | (30.2%) | (29.6%) |
| Very harmful | 982 | 1,140 |
| | (30.9%) | (28.6%) |
| Extremely harmful | 833 | 1,079 |

| Variable | Non-medical marijuana legal states (N = 4,123) | Non-medical marijuana illegal states (N = 3,294) |
|---|---|---|
| | (24.6%) | (25.5%) |
| BMI weight classification* | | |
| Underweight | 142 | 155 |
| | (3.89%) | (3.83%) |
| Normal | 1,558 | 1,805 |
| | (49.4%) | (44.6%) |
| Overweight | 806 | 963 |
| | (24.8%) | (24.7%) |
| Obesity | 706 | 1,084 |
| | (21.9%) | (26.9%) |
| Last time started physical fights* | | |
| Past month | 46 | 104 |
| | (1.20%) | (2.42%) |
| 2 -12 months ago | 108 | 127 |
| | (3.62%) | (3.25%) |
| Over a year ago | 338 | 425 |
| | (10.7%) | (10.2%) |
| Never | 2,797 | 3,461 |
| | (84.5%) | (84.1%) |

| Variable | Non-medical marijuana legal states (N = 4,123) | Non-medical marijuana illegal states (N = 3,294) |
|---|---|---|
| Last time bullied or threatened others | | |
| Past month | 75 (2.11%) | 122 (2.98%) |
| 2 -12 months ago | 106 (3.73%) | 141 (3.39%) |
| Over a year ago | 394 (12.1%) | 516 (13.0%) |
| Never | 2,712 (82.1%) | 3,335 (80.6%) |

*Note.* Numbers are unweighted frequencies with weighted column percentages in parentheses. Difference between respondents living in states where non-medical marijuana use is legal (AZ, CA, CO, CT, DC, IL, MA, ME, MI, MT, NJ, NM, NY, OR, VA, VT, WA) and those in states where non-medical marijuana use illegal (AL, AR, FL, GA, HI, IA, ID, IN, KS, KY, LA, MD, MN, MO, MS, NC, NE, NH, OH, OK, PA, RI, SC, SD, TN, TX, UT, WI, WV, WY) were tested using weighted Pearson's χ tests.

*p<0.05

*Sub-study 1 – LASSO regression*

**Tables 5** demonstrates variable selection using stepwise and ridge regressions. Stepwise regression used backward elimination at threshold = 0.20 to generate a model with 18 predictors. Ridge regression retained most of the original variables, though more than half of those variables had coefficients penalized to near-zero. **Table 6** shows variables selected via various unstratified least absolute shrinkage and selection operator (LASSO) models. Standard, square-root, CV, and adaptive LASSO all selected up to 10 predictors, while rigorous LASSO selected only two predictors. Standard LASSO and square-root LASSO selected identical variables with identical order of coefficients. CV LASSO and adaptive LASSO selected variables that were similar, but not identical to, standard LASSO. Specifically, CV LASSO included *starting fights* variable instead of *smokeless tobacco use*, while adaptive LASSO included *snus use* variable instead of *smokeless tobacco use* variable found in standard and square-root LASSO models.

**Table 7** summarizes diagnostic information for comparing and selecting unstratified LASSO models. CV LASSO was selected as the best performing model, because it had reasonable number of variables, lowest RMSE (0.1438) and highest $R^2$ (0.0592) in the validation sample, and highest AUROC (0.8655).

**Table 5.** Comparison of stepwise and ridge regression models and resulting variable set

| | Stepwise regression | | Ridge regression | |
|---|---|---|---|---|
| | Variable | $\|\beta\|$ | Variable | Penalized $\|\beta\|$ |
| 1 | Marijuana | 2.8565 | Coupons for hookah | 0.0588 |
| 2 | Coupons for hookah | 1.4907 | Heroin | 0.0470 |

| | Stepwise regression | | Ridge regression | |
|---|---|---|---|---|
| | Variable | \|β\| | Variable | Penalized \|β\| |
| 3 | Heroin | 0.7875 | Marijuana | 0.0390 |
| 4 | E-cigarette | 0.7608 | E-cigarette | 0.0377 |
| 5 | Hispanic | 0.7352 | Coupons for cigar | 0.0345 |
| 6 | Smokeless tobacco | 0.7178 | Cocaine | 0.0337 |
| 7 | Cocaine | 0.4996 | Snus | 0.0301 |
| 8 | Cigarette | 0.3853 | Smokeless tobacco | 0.0234 |
| 9 | Lax rules about non-combustible tobacco use at home | 0.2627 | Pipe | 0.0231 |
| 10 | Started fights | 0.2332 | Coupons for snus | 0.0163 |
| 11 | Bullied others | 0.1985 | Adderall | 0.0158 |
| 12 | Sexual orientation | 0.1865 | Hispanic | 0.0151 |
| 13 | Hard time paying attention | 0.1519 | Painkiller | 0.0141 |
| 14 | Employment | 0.1435 | Cigarette | 0.0125 |
| 15 | Income | 0.1160 | Started fights | 0.0086 |
| 16 | Perception of e-cigarette | 0.1120 | Cigar | 0.0084 |
| 17 | Exercise days per week | 0.0726 | Hookah | 0.0078 |
| 18 | Muscle strengthening days per week | 0.0677 | Methamphetamine | 0.0071 |
| 19 | - | | Bullied others | 0.0066 |

| | Stepwise regression | | Ridge regression | |
|---|---|---|---|---|
| | Variable | \|β\| | Variable | Penalized \|β\| |
| 20 | - | | Lax rules about combustible tobacco use at home | 0.0054 |
| 21 | - | | Lax rules about non-combustible tobacco use at home | 0.0048 |
| 22 | - | | Sexual orientation | 0.0048 |
| 23 | - | | Coupons for smokeless tobacco | 0.0040 |
| 24 | - | | Insurance | 0.0034 |
| 25 | - | | Hard time paying attention | 0.0028 |
| 26 | - | | Lives with tobacco user | 0.0024 |
| 27 | - | | Income | 0.0024 |
| 28 | - | | Employment | 0.0023 |
| 29 | - | | Enrolled in degree program | 0.0021 |
| 30 | - | | Housing | 0.0019 |
| 31 | - | | Sex | 0.0018 |
| 32 | - | | Perception of e-cigarette | 0.0014 |
| 33 | - | | Lied to get things | 0.0014 |
| 34 | - | | Quality of life | 0.0012 |
| 35 | - | | Exercise days per week | 0.0012 |
| 36 | - | | Sleep troubles | 0.0012 |

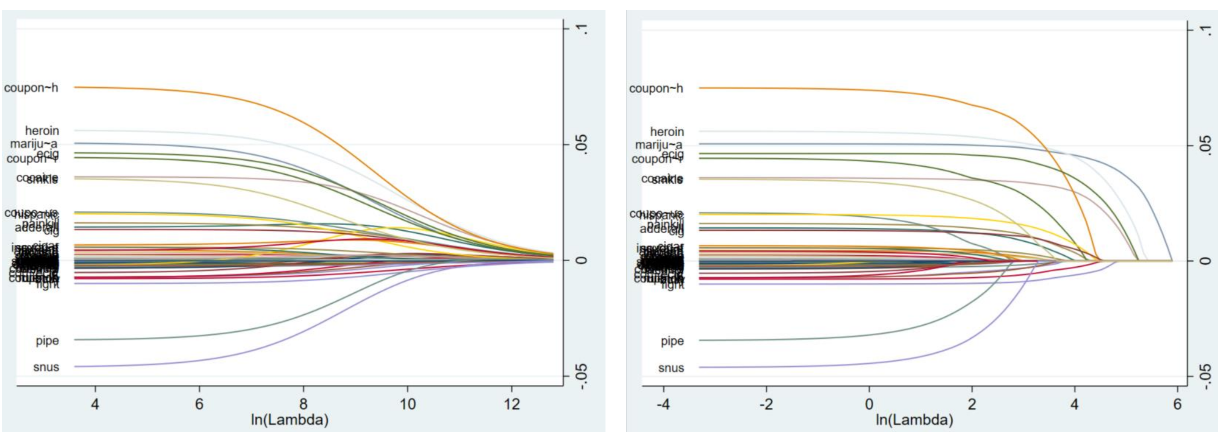
| | Stepwise regression | | Ridge regression | |
|---|---|---|---|---|
| | Variable | \|β\| | Variable | Penalized \|β\| |
| 37 | - | | Muscle strengthening days per week | 0.0010 |
| 38 | - | | Coupons for cigarette | 0.0010 |
| 39 | - | | Social media | 0.0009 |
| 40 | - | | Lax rules about e-cigarette use at home | 0.0009 |
| 41 | - | | Social satisfaction | 0.0008 |
| 42 | - | | Hard time following instructions | 0.0008 |
| 43 | - | | Coupons for e-cigarette | 0.0007 |
| 44 | - | | Felt restless | 0.0005 |
| 45 | - | | Radio time | 0.0004 |
| 46 | - | | Felt anxious | 0.0004 |
| 47 | - | | Distressed | 0.0004 |
| 48 | - | | Felt trapped | 0.0003 |
| 49 | - | | Alcohol | 0.0002 |
| 50 | - | | BMI | 0.0002 |
| 51 | - | | TV time | 0.0001 |
| 52 | - | | Race | 0.0001 |

*Note.* BMI = body mass index.

Stepwise regression used backward elimination at threshold = 0.20. Ridge regression used default settings of Stata’s lasso2 package, and selected model that minimized small-sample corrected Akaike Information Criteria values (AICc).

Additional diagnostic tests are visually represented in **Figures 2-3**. Figure 2 displays coefficient paths of standardized variables in ridge regression and LASSO, representing variables’ behavior during model convergence. Figure 3 is the AUROC graph of the best-performing LASSO model, with AUROC=0.8655.

**Figure 2.** Comparison of coefficient paths of variables – ridge regression vs. LASSO




*Note.* LASSO = least absolute shrinkage and selection operator.

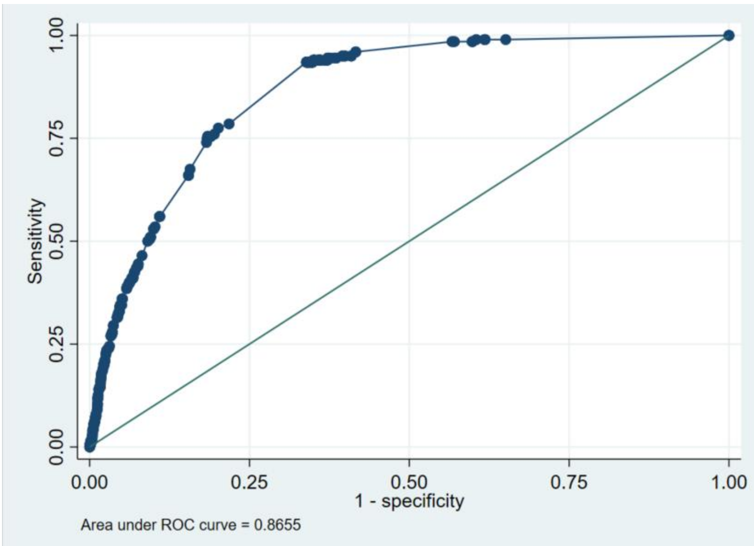
Graph on the left represents coefficient path plot of ridge regression; Graph on the right represents coefficient path plot of LASSO

**Table 7.** Model performance diagnostics

| | Standard LASSO | Square-Root LASSO | CV LASSO | Adaptive LASSO | Rigorous LASSO |
|---|---|---|---|---|---|
| Variable count | 10 | 10 | 10 | 10 | 2 |
| Training sample: | | | | | |
| RMSE | 0.1618 | 0.1618 | 0.1619 | 0.1618 | 0.1635 |
| $R^2$ | 0.0663 | 0.0663 | 0.0658 | 0.0663 | 0.0512 |
| Validation sample: | | | | | |
| RMSE | 0.1459 | 0.1459 | 0.1438 | 0.1461 | 0.1464 |
| $R^2$ | 0.0529 | 0.0529 | 0.0592 | 0.0498 | 0.0368 |
| AUROC | 0.8629 | 0.8629 | 0.8655 | 0.8570 | 0.8203 |

*Note.* RMSE = root-mean-square error. LASSO = least absolute shrinkage and selection operator; CV = cross-validated. AUROC = area under the receiver operating characteristic curve. Ratio of training sample to validation sample was 3:1.

**Figure 3.** Performance of unstratified LASSO model based on AUROC



*Note.* AUROC = area under the receiver operating characteristic curve; LASSO = Least Absolute Shrinkage and Selection Operator.

AUROC = 0.8655

**Table 6.** Comparison of LASSO models and resulting variable set

| | Standard LASSO | | Square-Root LASSO | | CV LASSO | | Adaptive LASSO | | Rigorous LASSO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Variable | Penalized $\lvert\beta\rvert$ | Variable | Penalized $\lvert\beta\rvert$ | Variable | Penalized $\lvert\beta\rvert$ | Variable | Penalized $\lvert\beta\rvert$ | Variable | Penalized $\lvert\beta\rvert$ |
| 1 | Coupon for Hookah | 0.0581 | Coupon for Hookah | 0.0575 | Heroin | 0.0468 | Coupon for Hookah | 0.0685 | Marijuana | 0.0448 |
| 2 | Heroin | 0.0510 | Heroin | 0.0509 | Marijuana | 0.0465 | Heroin | 0.0557 | E-cigarette | 0.0010 |
| 3 | Marijuana | 0.0490 | Marijuana | 0.0489 | E-cigarette | 0.0381 | Marijuana | 0.0518 | - | - |
| 4 | E-cigarette | 0.0437 | E-cigarette | 0.0435 | Coupon for Hookah | 0.0369 | E-cigarette | 0.0482 | - | - |
| 5 | Cocaine | 0.0340 | Cocaine | 0.0339 | Cocaine | 0.0313 | Cocaine | 0.0364 | - | - |
| 6 | Coupon for cigar | 0.0286 | Coupon for cigar | 0.0281 | Coupon for cigar | 0.0131 | Coupon for cigar | 0.0341 | - | - |
| 7 | Hispanic | 0.0155 | Hispanic | 0.0154 | Hispanic | 0.0096 | Smokeless tobacco | 0.0230 | - | - |
| 8 | Smokeless tobacco | 0.0134 | Smokeless tobacco | 0.0128 | Started fights | 0.0067 | Snus | 0.0226 | - | - |

| | Standard LASSO | | Square-Root LASSO | | CV LASSO | | Adaptive LASSO | | Rigorous LASSO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Variable | Penalized $\lvert\beta\rvert$ | Variable | Penalized $\lvert\beta\rvert$ | Variable | Penalized $\lvert\beta\rvert$ | Variable | Penalized $\lvert\beta\rvert$ | Variable | Penalized $\lvert\beta\rvert$ |
| 9 | Painkiller | 0.0114 | Painkiller | 0.0112 | Cigarette | 0.0063 | Hispanic | 0.0189 | - | - |
| 10 | Cigarette | 0.0107 | Cigarette | 0.0106 | Painkiller | 0.0054 | Painkiller | 0.0126 | - | - |

*Note.* LASSO = Least Absolute Shrinkage and Selection Operator; CV = cross-validated

Standard, square-root, and adaptive LASSO used λ parameter that minimize small-sample corrected Akaike Information Criteria values (AICc). CV LASSO used 10-fold validation to find the smallest mean-squared prediction error (MSPE), and then estimated model with lambda that minimized MSPE.

*Sub-study 1 – Stratified LASSO*

**Table 8** shows penalized coefficients and scaled importance of predictors estimated from CV LASSO, stratified by respondents' states' legalization of non-medical marijuana. While stratified LASSO underwent same model selection and iteration procedure as unstratified LASSO demonstrated earlier, the procedure produced different selection and rank order of predictors.

**Table 8.** LASSO regression of marijuana vaping with penalized coefficients and scaled importance, by legal status of non-medical marijuana

| | Variable | Penalized β | Penalized \|β\| | Scaled importance |
|---|---|---|---|---|
| **Legal** | | | | |
| 1 | Marijuana | 0.0564 | 0.0564 | 1 |
| 2 | E-cigarette | 0.0491 | 0.0491 | 0.8706 |
| 3 | Heroin | 0.0463 | 0.0463 | 0.8209 |
| 4 | Painkiller | 0.0411 | 0.0411 | 0.7209 |
| 5 | Cigarette | 0.0276 | 0.0276 | 0.4895 |
| 6 | Started fights | -0.0154 | 0.0154 | 0.2735 |
| 7 | Coupon for snus | 0.0108 | 0.0108 | 0.1923 |
| 8 | Hispanic | 0.0101 | 0.0101 | 0.1782 |
| 9 | Bullied others | -0.0084 | 0.0084 | 0.1488 |
| 10 | Cocaine | 0.0066 | 0.0066 | 0.1178 |
| **Not legal** | | | | |
| 1 | Heroin | 0.0377 | 0.0377 | 1 |

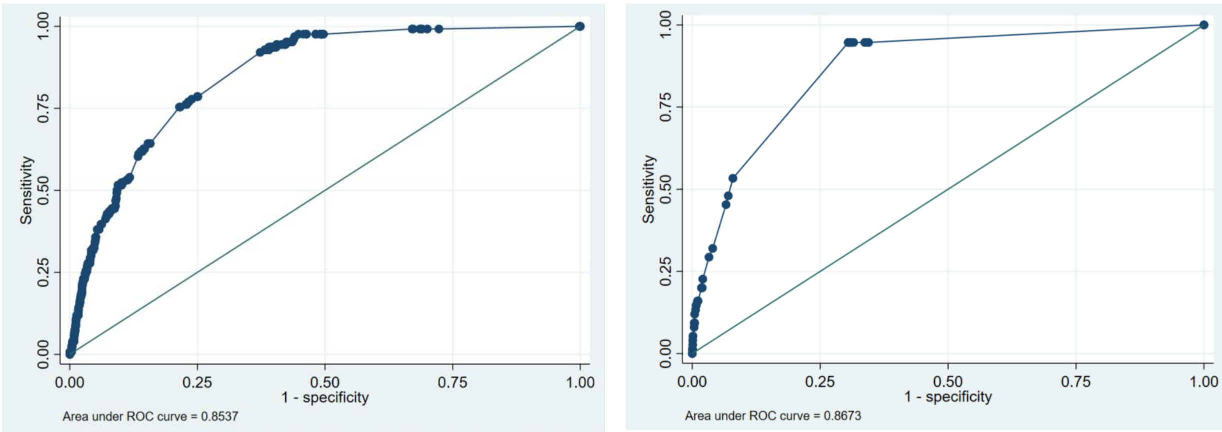
| | Variable | Penalized β | Penalized \|β\| | Scaled importance |
|---|---|---|---|---|
| 2 | Coupon for hookah | 0.0354 | 0.0354 | 0.9375 |
| 3 | Coupon for cigar | 0.0343 | 0.0343 | 0.9086 |
| 4 | Cocaine | 0.0303 | 0.0303 | 0.8035 |
| 5 | Marijuana | 0.0284 | 0.0284 | 0.7527 |
| 6 | Hookah | 0.0221 | 0.0221 | 0.5866 |
| 7 | E-cigarette | 0.0144 | 0.0144 | 0.3821 |

*Note.* Rank order was determined by relative contribution (i.e., scaled importance) of variables in prediction, calculated by dividing the penalized |β| of each predictor by the highest-magnitude |β|. If more than ten variables were identified by LASSO, only variables ranked 1-10 in importance are shown here.

Ten predictors (marijuana use, current e-cigarette use, past-year heroin use, past-year painkiller abuse, current cigarette use, starting fights, coupons for snus, Hispanic ethnicity, bullying others, and past-year cocaine use, in decreasing scaled importance) were selected for respondents residing in non-medical marijuana legal states, while seven predictors (past-year heroin use, coupon for hookah, coupon for cigar, past-year cocaine use, past-year marijuana use, current hookah use, and current e-cigarette use, in decreasing scaled importance) were selected for respondents residing in non-medical marijuana illegal states.

**Figure 4** represents the AUROC curve of the best performing LASSO (CV LASSO) model, when stratified by legal status of non-medical marijuana. The model appears to have reasonable discriminatory performance, with AUROC value of 0.854 for respondents living in non-medical marijuana legal states and 0.867 for respondents living in illegal states.

**Figure 4.** AUROC comparison of models for individuals by legal status of non-medical marijuana




*Note.* AUROC = area under the receiver operating characteristic curve; LASSO = Least Absolute Shrinkage and Selection Operator.

Graph on the left represents the model in states where non-medical marijuana is legal (AUROC = 0.8537); Graph on the right represents the model in states where non-medical marijuana is not legal (AUROC = 0.8673)

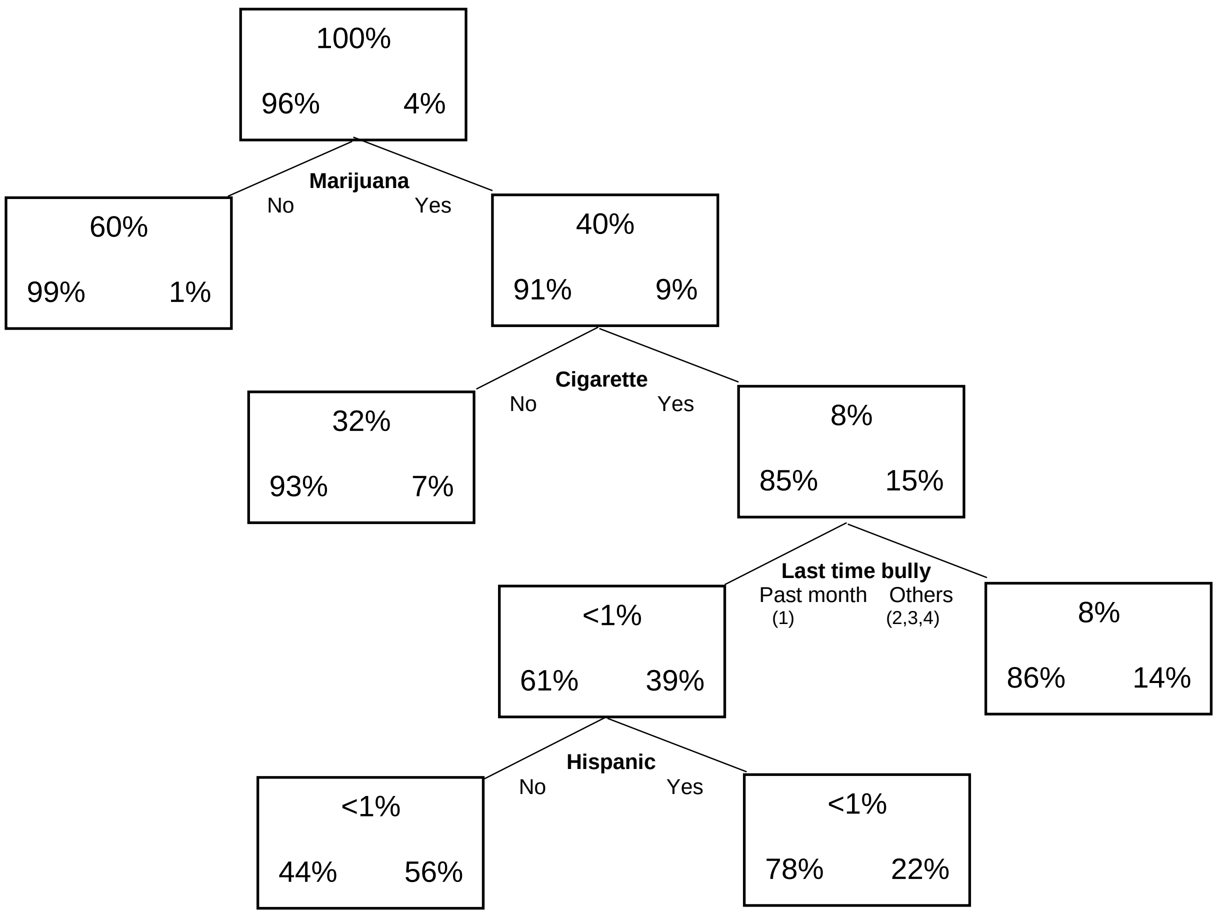
*Sub-study 1 – Stratified CART*

**Figures 5 and 6** displays results of stratified classification and regression tree (CART) model for multi-year marijuana vaping among U.S. young adults living in non-medical-marijuana-legal and illegal states. Stratified CART generated a five-terminal-node prediction model for states that legalized non-medical marijuana (split by past-year marijuana use, current cigarette use, last time respondent bullied others, and Hispanic ethnicity) and another five-terminal-node prediction model for states that have not legalized non-medical marijuana (split by past-year marijuana use, past-year heroin use, current e-cigarette use, and current hookah use).

Figure 5 represents CART model for states that have legalized non-medical adult use of marijuana. Here, the proportion of respondents reporting multi-year marijuana vaping at wave 6 was 1% for individuals who reported not smoking marijuana at wave 4, and 9% for individuals who reported smoking marijuana at wave 4. No additional splitting was observed for individuals who reported not smoking marijuana, while multiple splitting was observed for individuals who reported smoking marijuana. Among individuals who reported smoking marijuana, the next split occurred by current cigarette smoking, such that 7% of individuals identified as non-smokers reported multi-year marijuana vaping, compared to 15% of those identified as smokers reporting multi-year marijuana vaping. Cigarette smokers were further split by bullying behavior, such that multi-year marijuana vaping was 14% among those who never bullied others or did so more than a month ago, while multi-year marijuana vaping was 39% among those who bullied others less than a month ago. The final split occurred among individuals who recently bullied others, such that 22% of those with Hispanic ethnicity reported multi-year marijuana vaping, while 56% of non-Hispanics reported multi-year marijuana vaping.

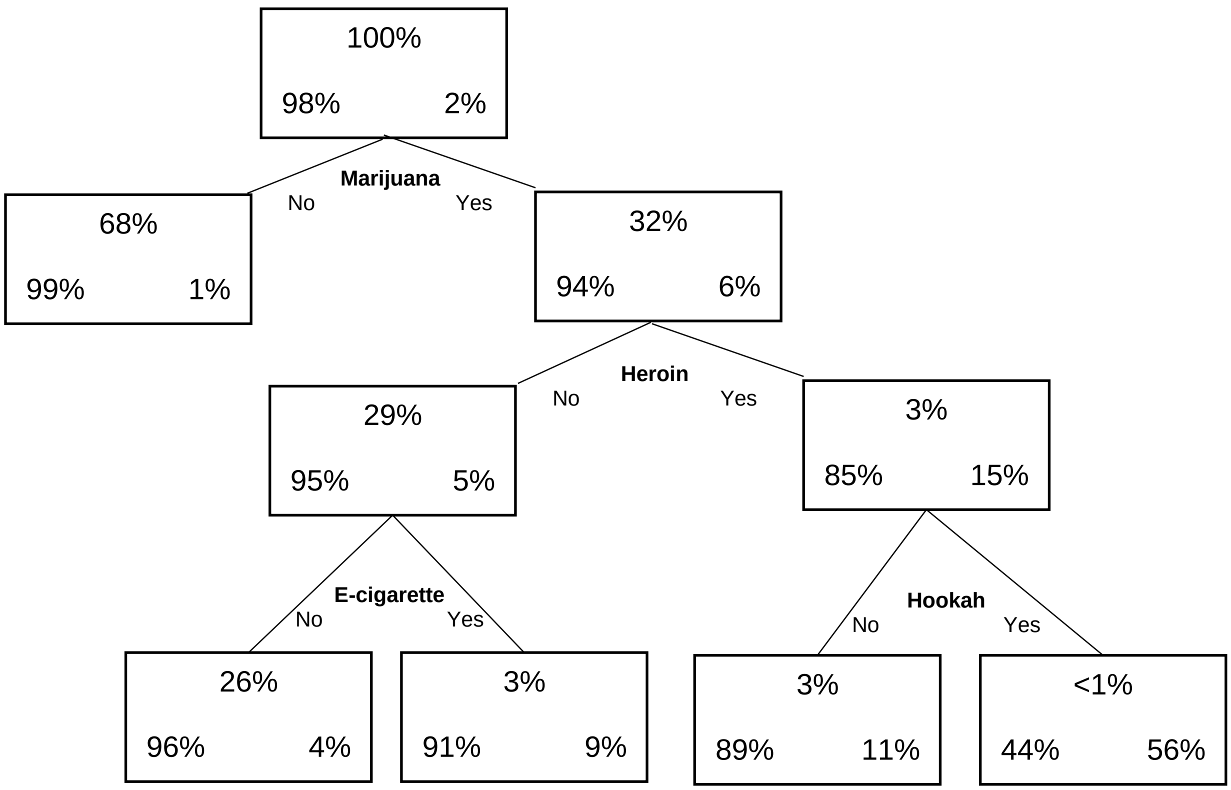
Figure 6 represents CART model for states that have not legalized non-medical adult use of marijuana. The first split occurred by marijuana smoking, such that 6% of those who smoked marijuana at wave 4 eventually reported multi-year marijuana vaping, compared to 1% of those who did not smoke marijuana at wave 4 reporting multi-year marijuana vaping. The next split occurred among marijuana smokers, where 15% of heroin users reported multi-year marijuana vaping, and 5% of heroin non-users reported multi-year marijuana vaping. Among heroin users, 56% of hookah users and 11% of hookah non-users reported multi-year marijuana vaping. Among heroin non-users, 9% of e-cigarette users and 4% of e-cigarette non-users reported multi-year marijuana vaping.

**Figure 5.** Results of CART model of multi-year marijuana vaping in states that have legalized non-medical adult use of marijuana




*Note.* The percentage on the top row of each rectangle indicates the proportion of the node from the entire sample. The percentage on the bottom left of each rectangle indicates the proportion of the respondents within that node who did not vape marijuana, while the percentage on the bottom right indicates the proportion of the respondents within that node who reported multi-year marijuana vaping by wave 6

**Figure 6.** Results of CART model of multi-year marijuana vaping in states that have not legalized non-medical adult use of marijuana




*Note.* The percentage on the top row of each rectangle indicates the proportion of the node from the entire sample. The percentage on the bottom left of each rectangle indicates the proportion of the respondents within that node who did not vape marijuana, while the percentage on the bottom right indicates the proportion of the respondents within that node who reported multi-year marijuana vaping by wave 6.

*Sub-study 2 – Descriptive statistics*

**Table 9** summarizes the characteristics of the analytic sample (age, mean [SD] = 21.5 [2.29] years) by legal status of non-medical marijuana. Overall, alcohol was the most comm used among all included substances, followed e-cigarette use, joint marijuana smoking, and blunt marijuana smoking. Only a few respondents reported misusing prescription drugs and illegal drugs, even after collapsing many of the sub-categories to increase the cell size.

Of the 11,449 young adults, 5,042 lived in states where non-medical marijuana use was considered legal, and 6,407 lived in states where non-medical marijuana use was considered illegal. For individuals residing in non-medical marijuana legal states, past-month use of marijuana vaping was 11.8% (n = 602), joint marijuana was 21.9% (n = 1,029), blunt marijuana was 9.29% (n = 479), alcohol was 51.6% (n = 2,386), cigarette was 10.2% (n = 448), e-cigarette was 20.2% (n = 923), cigar was 5.10% (n = 216), other tobacco products was 5.60% (n = 250), prescription drugs was 1.46% (n = 73), and Illegal drugs was 1.15% (n = 48). For individuals residing in states where non-medical marijuana use is considered not legal, past-month use of marijuana vaping was 8.17%,(n = 501), joint marijuana was 14.0% (n = 849), blunt marijuana was 9.73% (n = 640), alcohol was 46.0% (n = 2,734), cigarette was 12.2% (n = 689), e-cigarette was 19.9% (n = 1,169), cigar was 6.68% (n = 413), other tobacco products was 5.34% (n = 306), prescription drugs was 1.44% (n = 89), and Illegal drugs was 0.69% (n = 38).

Pearson's $\chi$ test revealed significant differences in Past-month marijuana vaping, joint marijuana smoking, alcohol use, cigarette smoking, cigar smoking, illegal drug use, race/ethnicity, sexual minority, housing, health insurance coverage, and perceived quality of life ($p$'s $<0.05$) by legal status of non-medical marijuana.

**Table 9.** Weighted descriptive statistics for all variables in sub-study 2

| Variable | Non-medical marijuana legal states (N = 5,042) | Non-medical marijuana illegal states (N = 6,407) |
|---|---|---|
| Marijuana vaping* | 602 (11.8%) | 501 (8.17%) |
| Joint marijuana smoking* | 1,027 (21.9%) | 849 (14.0%) |
| Blunt marijuana smoking | 479 (9.29%) | 640 (9.73%) |
| Alcohol* | 2,386 (51.6%) | 2,734 (46.0%) |
| Cigarette* | 448 (10.2%) | 689 (12.2%) |
| E-cigarette | 923 (20.2%) | 1,169 (19.9%) |
| Cigar* | 216 (5.10%) | 413 (6.68%) |
| Other tobacco products | 250 (5.60%) | 306 (5.34%) |
| Prescription drugs | 73 | 89 |

| Variable | Non-medical marijuana legal states (N = 5,042) | Non-medical marijuana illegal states (N = 6,407) |
|---|---|---|
| | (1.46%) | (1.44%) |
| Illegal drugs* | 48 | 38 |
| | (1.15%) | (0.69%) |
| Male | 2,484 | 3,098 |
| | (49.7%) | (51.0%) |
| Race/ethnicity* | | |
| Non-Hispanic white | 2,108 | 3,029 |
| | (48.0%) | (56.8%) |
| Non-Hispanic black | 427 | 1,181 |
| | (9.33%) | (16.0%) |
| Hispanic | 1,858 | 1,529 |
| | (27.9%) | (18.1%) |
| Other | 649 | 668 |
| | (14.8%) | (8.98%) |
| Sexual minority* | 940 | 1,071 |
| | (20.9%) | (18.1%) |
| Housing* | | |
| Apartment | 4,219 | 5,253 |
| | (91.0%) | (88.2%) |
| Dorm or others | 403 | 653 |

| Variable | Non-medical marijuana legal states (N = 5,042) | Non-medical marijuana illegal states (N = 6,407) |
|---|---|---|
| | (9.01%) | (11.8%) |
| Employment | | |
| Work full-time | 1,676 | 2,320 |
| | (39.6%) | (40.6%) |
| Work part-time | 1,422 | 1,832 |
| | (29.1%) | (30.4%) |
| Do not work for pay | 1,524 | 1,754 |
| | (31.2%) | (29.0%) |
| Enrolled in degree program | 1,940 | 2,354 |
| | (37.7%) | (36.0%) |
| Insurance* | 4,200 | 5,035 |
| | (87.0%) | (80.4%) |
| Perceived quality of life* | | |
| Excellent | 1,413 | 1,977 |
| | (28.4%) | (30.9%) |
| Very good | 2,057 | 2,538 |
| | (41.4%) | (39.5%) |
| Good | 1,227 | 1,513 |

| Variable | Non-medical marijuana legal states (N = 5,042) | Non-medical marijuana illegal states (N = 6,407) |
|---|---|---|
| | (23.4%) | (23.6%) |
| Fair | 305 | 337 |
| | (6.41%) | (5.25%) |
| Poor | 35 | 36 |
| | (0.63%) | (0.71%) |

*Note.* Numbers are unweighted frequencies with weighted column percentages in parentheses. Difference between respondents living in states where non-medical marijuana use is legal (AK, AZ, CA, CO, CT, DC, IL, MA, ME, MI, MT, NJ, NM, NV, NY, OR, VA, VT, WA) and those in states where non-medical marijuana use illegal (AL, AR, FL, GA, HI, IA, ID, IN, KS, KY, LA, MD, MN, MO, MS, NC, ND, NE, NH, OH, OK, PA, RI, SC, SD, TN, TX, UT, WI, WV, WY) were tested using weighted Pearson's $\chi$ tests.

*$p<0.05$

*Sub-study 2 – LCA*

Latent class analyses (LCA) were conducted to identify latent classes of co-occurring substance groups. **Tables 10-12** provide statistical fit indices for 1-5 classes. Fit indices showed, based on combination of BIC and bootstrap likelihood ratio test (BLRT), that models with 4-classes were optimal for both unstratified and stratified LCA.

**Table 10.** Latent class model fit indices, unstratified model

| Classes | Log likelihood | df | AIC | BIC | BLRT |
|---|---|---|---|---|---|
| 1 | -35437.87 | 10 | 70895.75 | 70969.20 | - |
| 2 | -31783.11 | 20 | 63606.22 | 63753.14 | <0.01 |
| 3 | -31426.2 | 31 | 62914.41 | 63142.12 | <0.01 |
| **4** | **-31246.79** | **42** | **62577.58** | **62886.09** | **<0.01** |
| 5 | -31208.15 | 53 | 62522.30 | 62911.62 | 1.00 |

*Note.* df = degrees of freedom; AIC = Akaike information criterion; BIC = Bayesian information criterion; BLRT = bootstrap likelihood ratio test.

BLRT shows p-values for relative adequacy of current model (with K class) compared to previous model (with K-1 class)

**Table 11.** Latent class model fit indices, in states that have legalized non-medical adult use of marijuana

| Classes | Log likelihood | df | AIC | BIC | BLRT |
|---|---|---|---|---|---|
| 1 | -14370.42 | 10 | 28760.85 | 28825.24 | - |
| 2 | -12772.31 | 21 | 25586.61 | 25721.82 | <0.01 |
| 3 | -12639.88 | 31 | 25341.76 | 25541.36 | <0.01 |
| **4** | **-12557.98** | **42** | **25199.95** | **25470.37** | **0.01** |
| 5 | -12549.92 | 54 | 25205.83 | 25547.08 | 0.69 |

*Note.* df = degrees of freedom; AIC = Akaike information criterion; BIC = Bayesian information criterion; BLRT = bootstrap likelihood ratio test.

BLRT shows p-values for relative adequacy of current model (with K class) compared to previous model (with K-1 class)

**Table 12.** Latent class model fit indices, in states that have not legalized non-medical adult use of marijuana

| Classes | Log likelihood | df | AIC | BIC | BLRT |
|---|---|---|---|---|---|
| 1 | -17648.25 | 10 | 35316.51 | 35383.35 | - |
| 2 | -15832.29 | 21 | 31706.58 | 31846.93 | <0.01 |
| 3 | -15641.36 | 31 | 31344.72 | 31551.91 | <0.01 |
| **4** | **-15542.07** | **43** | **31170.14** | **31457.54** | **<0.01** |
| 5 | -15506.45 | 53 | 31118.91 | 31473.14 | 1.00 |

*Note.* df = degrees of freedom; AIC = Akaike information criterion; BIC = Bayesian information criterion; BLRT = bootstrap likelihood ratio test.

BLRT shows p-values for relative adequacy of current model (with K class) compared to previous model (with K-1 class)

Item-response probabilities diagram (**Tables 13-14 and Figures 7-8**) were used to confirm that individuals in each latent class had similar response patterns to the observed indicators, and that classes were well-separated. Classes were labelled for identification purposes, based on clearly co-occurring substances: Class 1 = no substance use; Class 2 = alcohol only; Class 3 = alcohol and tobacco; and Class 4 = high risk. High risk class represents individuals who concurrently use almost all included substances.

**Table 13.** Item-response probabilities of substance use classes in states that have legalized non-medical adult use of marijuana

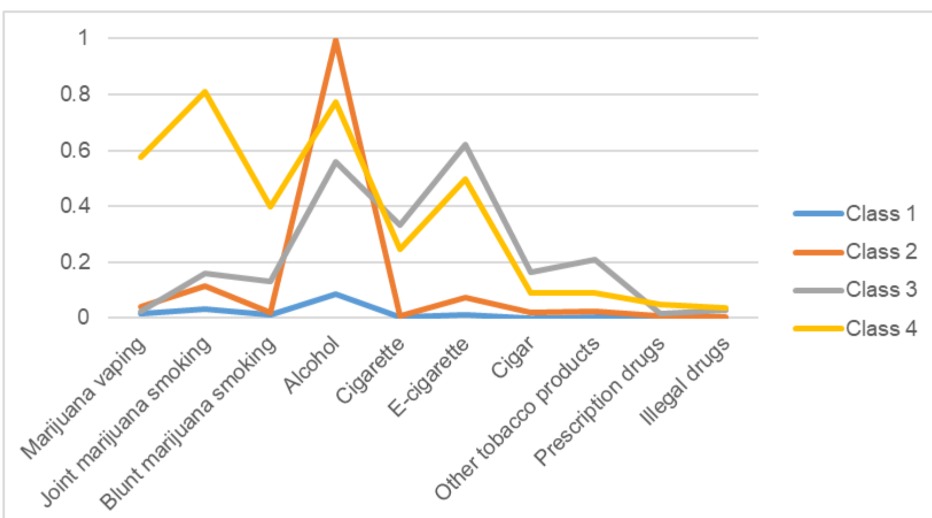
| Substance | Class 1 (γ=0.4803) | Class 2 (γ=0.2266) | Class 3 (γ=0.1222) | Class 4 (γ=0.1709) |
|---|---|---|---|---|
| Marijuana vaping | 0.01495 | 0.04081 | 0.02279 | 0.57613 |
| Joint marijuana smoking | 0.03219 | 0.11207 | 0.15904 | 0.81062 |
| Blunt marijuana smoking | 0.01087 | 0.01763 | 0.13001 | 0.39875 |
| Alcohol | 0.08417 | 0.99578 | 0.55732 | 0.77253 |
| Cigarette | 0.00122 | 0.00593 | 0.33201 | 0.24579 |
| E-cigarette | 0.00977 | 0.07454 | 0.62139 | 0.49587 |
| Cigar | 0.00044 | 0.01906 | 0.16464 | 0.08794 |
| Other tobacco products | 0.00416 | 0.02239 | 0.20909 | 0.08980 |
| Prescription drugs | 0.00660 | 0.00652 | 0.01688 | 0.04861 |
| Illegal drugs | 0.00000 | 0.00086 | 0.02705 | 0.03723 |

**Table 14.** Item-response probabilities of substance use classes in states that have not legalized non-medical adult use of marijuana

| Substance | Class 1 (γ=0.5193) | Class 2 (γ=0.2042) | Class 3 (γ=0.1727) | Class 4 (γ=0.1037) |
|---|---|---|---|---|
| Marijuana vaping | 0.00256 | 0.06094 | 0.04927 | 0.54269 |
| Joint marijuana smoking | 0.01303 | 0.09084 | 0.10348 | 0.85334 |
| Blunt marijuana smoking | 0.00990 | 0.02964 | 0.14385 | 0.60402 |

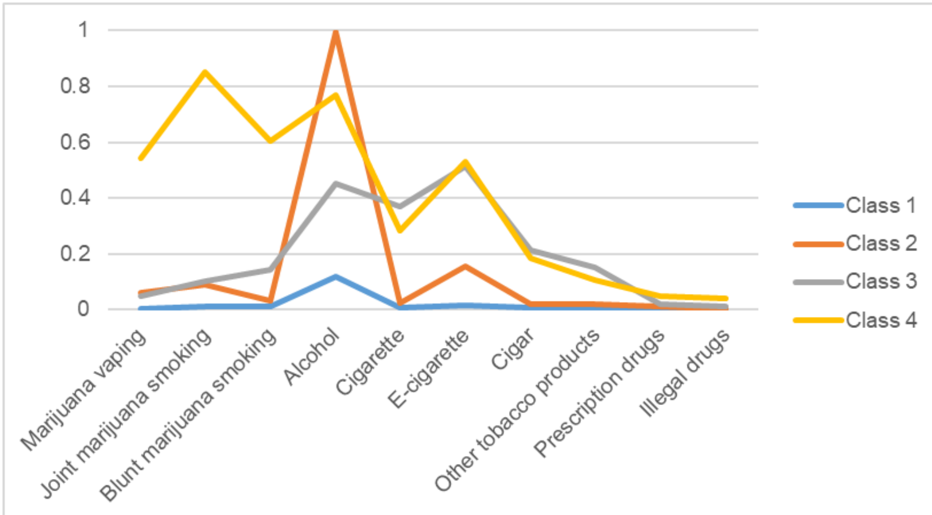
| Substance | Class 1 (γ=0.5193) | Class 2 (γ=0.2042) | Class 3 (γ=0.1727) | Class 4 (γ=0.1037) |
|---|---|---|---|---|
| Alcohol | 0.11737 | 0.99454 | 0.44974 | 0.76859 |
| Cigarette | 0.00634 | 0.02488 | 0.36933 | 0.28413 |
| E-cigarette | 0.01434 | 0.15558 | 0.51377 | 0.53108 |
| Cigar | 0.00526 | 0.01858 | 0.21272 | 0.18278 |
| Other tobacco products | 0.00724 | 0.01946 | 0.14994 | 0.10737 |
| Prescription drugs | 0.00702 | 0.00989 | 0.02007 | 0.04693 |
| Illegal drugs | 0.00000 | 0.00088 | 0.00919 | 0.03962 |

**Figure 7.** Item-response probabilities of substance use classes in states that have legalized non-medical adult use of marijuana




*Note*. y-axis represents probability. Labels assigned to each class are as follows: Class 1 = no substance use; Class 2 = alcohol only; Class 3 = alcohol and tobacco; Class 4 = high risk

**Figure 8.** Item-response probabilities of substance use classes in states that have not legalized non-medical adult use of marijuana




*Note*. y-axis represents probability. Labels assigned to each class are as follows: Class 1 = no substance use; Class 2 = alcohol only; Class 3 = alcohol and tobacco; Class 4 = high risk

It should be noted that, while the number of classes and patterns of co-occurring substances are similar in marijuana-legal and illegal states, relative proportion of class members differ by marijuana law. For example, 17% of respondents from marijuana-legal states were members of class 4, while only 10% of respondents from marijuana illegal states were members of its class 4. As another example, class 1 (no substance use) represent about 48% of respondents in marijuana legal states, while class 1 (no substance use) represents relatively higher proportion (52%) of respondents.in marijuana illegal states.

*Sub-study 2 – Multinomial logistic regression*

**Tables 15-16** provides results of the multinomial logistic regression, generated after identifying the best-fitting model of substance use classes. Multinomial logistic regression was used to identify which of the predictors (sex, race/ethnicity, sexual minority, housing, employment, degree enrollment, health insurance, and quality of life) were significant predictors of class membership, in in states that have legalized non-medical adult use of marijuana (Table 15) and states that have not legalized non-medical adult use of marijuana (Table 16).

**Table 15.** Estimated odds ratio (OR) of substance use class membership based on covariates in states that have legalized non-medical adult use of marijuana

| Reference class: No substance use | Alcohol only<br>OR<br>(95% CI) | Alcohol and tobacco<br>OR<br>(95% CI) | High risk<br>OR<br>(95% CI) |
|---|---|---|---|
| Male | 0.64***<br>(0.50-0.82) | 1.30<br>(0.96-1.75) | 1.13<br>(0.91-1.40) |
| Race/ethnicity | | | |
| Non-Hispanic white | Reference | Reference | Reference |
| Non-Hispanic black | 0.28**<br>(0.13-0.58) | 1.26<br>(0.81-1.98) | 0.21***<br>(0.12-0.38) |
| Hispanic | 0.38***<br>(0.27-0.55) | 0.42***<br>(0.29-0.60) | 0.48***<br>(0.38-0.61) |
| Other | 0.39***<br>(0.26-0.59) | 0.49*<br>(0.28-0.84) | 0.58**<br>(0.42-0.79) |
| Sexual minority | 1.14<br>(0.83-1.58) | 1.24<br>(0.85-1.82) | 2.30***<br>(1.80-2.92) |

| Reference class: No substance use | Alcohol only OR (95% CI) | Alcohol and tobacco OR (95% CI) | High risk OR (95% CI) |
|---|---|---|---|
| Housing | | | |
| Apartment | Reference | Reference | Reference |
| Dormitory and others | 0.91 | 1.44 | 1.20 |
| | (0.59-1.41) | (0.90-2.31) | (0.85-1.70) |
| Employment | | | |
| Works full-time | Reference | Reference | Reference |
| Works part-time | 0.36*** | 0.49*** | 0.65** |
| | (0.26-0.51) | (0.33-0.71) | (0.49-0.86) |
| No work | 0.22*** | 0.48*** | 0.49*** |
| | (0.14-0.34) | (0.34-0.69) | (0.37-0.66) |
| Degree enrollment | 1.81*** | 0.32*** | 0.99 |
| | (1.37-2.38) | (0.20-0.51) | (0.78-1.24) |
| Health insurance | 2.67** | 0.74 | 1.22 |
| | (1.48-4.80) | (0.53-1.04) | (0.90-1.64) |
| Quality of life | | | |
| Excellent | Reference | Reference | Reference |
| Very good | 1.18 | 0.86 | 1.62** |

| Reference class: No substance use | Alcohol only OR (95% CI) | Alcohol and tobacco OR (95% CI) | High risk OR (95% CI) |
|---|---|---|---|
| | (0.90-1.54) | (0.58-1.27) | (1.23-2.14) |
| Good | 0.81 | 1.95** | 2.15*** |
| | (0.56-1.15) | (1.33-2.84) | (1.59-2.91) |
| Fair/Poor | 0.60 | 2.69*** | 3.47*** |
| | (0.28-1.30) | (1.60-4.51) | (2.30-5.23) |

*Note.* OR = odds ratio, **p<.05, **p<.01, ***p<.001*

When the "no substance use" class was treated as the reference class, individuals in marijuana-legal states had sex, race/ethnicity, employment, degree enrollment, and health insurance as significant predictors of "alcohol only" class membership. Being men (OR=0.64), being one of the racial/ethnic minorities (OR=0.28-0.39), and not working full-time (OR=0.22-0.36) all decreased the likelihood of "alcohol only" class membership, while being enrolled in a degree program (OR=1.81) and having health insurance (OR=2.67) increased the likelihood of "alcohol only" class membership. In marijuana-legal states, membership to "alcohol and tobacco" class was predicted by race/ethnicity, employment, degree enrollment, and quality of life, such that being Hispanic (OR=0.42), not working full-time (OR=0.48-0.49), and being enrolled in a degree program (OR=0.32) decreased the likelihood of class membership, while having relatively poorer quality of life (OR=1.95-2.69) increased the likelihood of belonging to the "alcohol and tobacco" class. Membership to "high risk" class was predicted by race/ethnicity, sexual orientation, employment, and quality of life, such that being one of the racial/ethnic minorities (OR=0.21-0.58) and not working full-time (OR=0.49-0.65) decreased likelihood of membership, while identifying as sexual minority (OR=2.30) and having relatively poorer quality of life (OR=1.62-3.47) increased likelihood of membership.

**Table 16.** Estimated odds ratio (OR) of substance use class membership based on covariates in states that have not legalized non-medical adult use of marijuana

| Reference class: No substance use | Alcohol only OR (95% CI) | Alcohol and tobacco OR (95% CI) | High risk OR (95% CI) |
|---|---|---|---|
| Male | 0.72** | 1.33* | 1.26* |
| | (0.57-0.89) | (1.05-1.69) | (1.02-1.58) |
| Race/ethnicity | | | |
| Non-Hispanic white | Reference | Reference | Reference |
| Non-Hispanic black | 0.12*** | 0.62** | 0.50*** |
| | (0.06-0.25) | (0.43-0.89) | (0.36-0.70) |
| Hispanic | 0.56*** | 0.48*** | 0.49*** |
| | (0.42-0.74) | (0.35-0.65) | (0.36-0.66) |
| Other | 0.45*** | 0.66 | 1.18 |
| | (0.31-0.65) | (0.43-1.01) | (0.87-1.61) |
| Sexual minority | 1.98*** | 1.68** | 3.74*** |
| | (1.47-2.66) | (1.22-2.33) | (2.92-4.79) |
| Housing | Reference | Reference | Reference |

| Reference class: No substance use | Alcohol only OR (95% CI) | Alcohol and tobacco OR (95% CI) | High risk OR (95% CI) |
|---|---|---|---|
| Apartment | | | |
| Dormitory and others | 0.84 | 1.52* | 0.76 |
| | (0.59-1.20) | (1.09-2.11) | (0.53-1.09) |
| Employment | | | |
| Works full-time | Reference | Reference | Reference |
| Works part-time | 0.42*** | 0.40*** | 0.83 |
| | (0.32-0.56) | (0.29-0.59) | (0.64-1.08) |
| No work | 0.29*** | 0.57*** | 0.53*** |
| | (0.22-0.38) | (0.43-0.74) | (0.40-0.70) |
| Degree enrollment | 2.02*** | 0.30*** | 0.95 |
| | (1.61-2.54) | (0.19-0.46) | (0.74-1.21) |
| Health insurance | 2.57*** | 0.53*** | 1.11 |
| | (1.71-3.87) | (0.41-0.69) | (0.84-1.47) |
| Quality of life | | | |
| Excellent | Reference | Reference | Reference |
| Very good | 1.54*** | 1.78*** | 1.98*** |
| | (1.21-1.97) | (1.31-2.42) | (1.49-2.62) |

| Reference class: No substance use | Alcohol only OR (95% CI) | Alcohol and tobacco OR (95% CI) | High risk OR (95% CI) |
|---|---|---|---|
| Good | 1.13 | 2.46*** | 2.91*** |
| | (0.83-1.54) | (1.79-3.40) | (2.16-3.91) |
| Fair/Poor | 1.07 | 3.49*** | 3.78*** |
| | (0.56-2.05) | (2.23-5.47) | (2.44-5.85) |

*Note.* OR = odds ratio, $*p<.05$, $**p<.01$, $***p<.001$

Similar but different pattern was observed for individuals living in states where marijuana has not been legalized. When the "no substance use" class was treated as the reference class, individuals in marijuana-illegal states had sex, race/ethnicity, sexual orientation, employment, degree enrollment, and health insurance as significant predictors of "alcohol only" class membership. That is, being men (OR=0.72), being one of the racial/ethnic minorities (OR=0.12-0.56), and not working full-time (OR=0.29-0.42) all decreased the likelihood of "alcohol only" class membership, while identifying as sexual minority (OR=1.98), being enrolled in a degree program (OR=2.02), and having health insurance (OR=2.57) increased the likelihood of "alcohol only" class membership. Membership to "alcohol and tobacco" class was predicted by all predictor variables in the model (i.e., sex, race/ethnicity, sexual orientation, housing, employment, degree enrollment, health insurance, and quality of life), such that being black or Hispanic (OR=0.48-62), not working full-time (OR=0.40-0.57), being enrolled in a degree program (OR=0.30), and having health insurance (OR=0.53) all decreased the likelihood of class membership. Being men (OR=1.33), living in a dormitory or any other non-apartment residence

(OR=1.52), and having poorer quality of life (OR=1.78-3.49) all increased the likelihood of belonging to the “alcohol and tobacco” class. Membership to “high risk” class in marijuana-illegal states was predicted by sex, race/ethnicity, sexual orientation, employment, and quality of life, such that being black or Hispanic (OR=0.49-0.50) and not working for pay (OR=0.53) decreased likelihood of membership, while being men (OR=1.26), identifying as sexual minority (OR=3.74), and having relatively poorer quality of life (OR=1.98-3.78) all increased likelihood of belonging to the “high risk” class.

Chapter 5

# DISCUSSION AND CONCLUSION

*Sub-study 1 – Purpose of study*

The purpose of sub-study 1 was to use machine learning to identify risk profiles of marijuana vaping. Specifically, this study was designed to identify and examine emerging predictors of marijuana vaping continuation among U.S. young adults while including variables (e.g., housing, insurance status, and BMI) unexplored by existing studies. This study was one of the few studies to examine marijuana vaping using the machine learning method and one of the first to focus on multi-year vaping to differentiate prolonged users from temporary experimenters. This study identified and classified risk factors of marijuana vaping using the most up-to-date nationally representative longitudinal sample of U.S. young adults.

At the beginning of the research, I hypothesized that variables identified as important by machine learning techniques would differ from those identified by non-machine learning techniques. After data analysis, I verified that variables identified as important by machine learning technique differed from those identified by non-machine learning techniques, as LASSO models were more parsimonious than stepwise and ridge regression models. I also hypothesized that the state's law on non-medical marijuana would have a significant influence on the outcome, and that housing, health insurance status, and BMI would emerge as significant predictors of marijuana vaping. This hypothesis was only partially correct, as models were notably different when stratified by legal status of non-medical marijuana, but none of the models identified housing, health insurance status, and BMI as important predictors of multi-year marijuana vaping.

*Sub-study 2 – Purpose of study*

The purpose of sub-study 2 was to comprehensively examine polydrug use patterns of 10 substance groups, including marijuana vaping, using latent class analysis. Specifically, the study aimed to (1) identify classes of substance use to examine which of the 10 substance groups commonly abused and misused by young adults are used concurrently, (2) examine predictors of each substance class, and (3) explore potential heterogeneity by user characteristics. This study filled an important gap in the literature, as no studies to date have rigorously examined marijuana vaping and its polydrug use among U.S. young adults, especially using the most recent data from a nationally representative longitudinal study.

For sub-study 2, I hypothesized that marijuana vaping is done concurrently with e-cigarettes and marijuana. This hypothesis was somewhat correct, as one of the four identified classes (i.e., "high risk" class) had individuals who concurrently used a majority of substances, including marijuana vaping, marijuana smoking, and e-cigarette vaping. This is in line with many of the existing studies, which suggest that nicotine and cannabis addiction share a common genetic risk factor (Lemyre, Poliakova, & Bélanger, 2019), that a significant number of individuals use nicotine and marijuana products concurrently (Jacobs et al., 2022; Agrawal, Budney, & Lynskey, 2012), and that e-cigarettes can serve as a gateway drug to marijuana use, including marijuana vaping (Wong et al., 2020; Sun, Mendez, & Warner, 2022). However, this does not necessarily mean that e-cigarette vapers concurrently smoke or vape marijuana, as e-cigarette vapers may belong to a different class (i.e., “alcohol and tobacco” class) and limit their substance use to alcohol and tobacco products.

I also hypothesized that (1) patterns of polydrug use would differ by state legalization of non-medical marijuana, and (2) individuals' race/ethnicity, sexual orientation, housing, employment, degree enrollment, health insurance status, and perceived quality of life significantly predict class membership. Results of sub-study 2 indicated that patterns of polydrug use are similar in non-medical marijuana legal and illegal states, and class membership is predicted by individuals' race/ethnicity, sexual orientation, housing, employment, degree enrollment, health insurance status, and perceived quality of life.

### *Sub-study 1 – Discussion of results*

Past studies have reported mixed findings on whether or not the legalization of non-medical marijuana increases the likelihood of drug abuse and misuse (Coley et al., 2021; Chu, 2015; Buttorff et al., 2023; Mennis, McKeon, & Stahler, 2023; Stormshak et al., 2019). Those mixed findings suggest that the way substance use interacts with marijuana law is complex, and legalization may sometimes decrease substance use by providing safer alternatives, while it may act as a gateway to increase the use of another substance. While an examination of substance use over time was outside the scope of this study, descriptive statistics showed that multi-year marijuana vaping was more common in states that legalized non-medical marijuana than states that have not legalized non-medical marijuana.

CART diagrams (Figures 5-6) showed that substance use generally increases the likelihood of multi-year marijuana vaping. This agrees not only with existing polydrug studies that examine substance use as co-occurring with other substances (Lanza, Motlagh, & Orozco, 2020; Linden-Carmichael et al., 2022; Park & Kim, 2018; Mattingly, Elliott, & Fleischer, 2023), but it may also align with studies that view marijuana products as potential substitute for substances (Reiman, 2009; Charoenporn, Charoenporn, & Mackie, 2023)

A closer examination of stratified CART models showed that, in states that have legalized non-medical marijuana, split occurred by past-year marijuana use, current cigarette smoking, bullying others, and Hispanic ethnicity. This is notably different from the CART model in Han and Seo's (2022) study on marijuana vaping initiation, which observed splitting by past-year marijuana use, risk perception of e-cigarettes, and having tobacco use home rule. While past-year marijuana use predicted marijuana vaping in both, this shows how differences in outcome measurement (initiation vs multi-year use) create drastically different results. This also shows how rules and risk perception may influence individuals' decision to try a substance, but they may have no influence on individuals' decision to use what they have been using for a prolonged period of time. With the legalization of marijuana, chronic marijuana vapers may have little incentive to re-evaluate their substance use behavior, and the effect of rule and perception variables are masked by other variables. This explanation is largely consistent with the explanation provided in the literature by Ducci and Goldman (2012). According to Ducci and Goldman (2012), whether or not occasional substance use progresses toward a pathological pattern to become an addiction is often determined by intrinsic and epigenetic factors (i.e., genotype, preexisting addictive disorder, and mental illness). Given how this study's outcome captured the most serious of marijuana vapers, it is unsurprising that its predictors are close approximations of intrinsic and epigenetic determinants of addiction (i.e., ethnicity, past substance use, and bullying behavior) (Ducci and Goldman, 2012).

The explanation for marijuana-illegal states can generally be similar to the above, as all of the splits were based on past substance use. It should be noted that one of the substance use variables to split the CART model and predict multi-year marijuana vaping was heroin use. While the association between heroin and marijuana vaping was unsurprising, the fact that it was the second splitting variable and occurred only in marijuana-illegal states was somewhat surprising. A possible explanation for this observation is an underlying criminality and risk-taking tendency (Dahlback, 1990; Junger & Dekovid, 2017). In other words, individuals who vape marijuana over multiple years in states that consider it illegal may be risk takers who are less inclined to follow the law, which consequently makes them more likely to break other laws and use heroin.

It should be noted that the results of stratified CART were similar to, but not identical to, that of stratified LASSO. For example, LASSO selected *e-cigarette*s, *heroin*, and *painkiller*s as second to fourth most important variables in legal states (Table 8), but those variables were not identified as splitting variables in CART (Figure 5). Moreover, LASSO and CART appear to disagree on Hispanic ethnicity as a predictor. That is, one of the final splits of CART (Figure 5) showed multi-year marijuana vaping as being less common among Hispanic individuals than non-Hispanic counterparts, while LASSO showed Hispanic individuals as being more likely to report multi-year marijuana (Table 8). Such discrepancy may indicate the presence of a non-linear interaction between predictors. While LASSO is widely used for reducing dimensionality (Shi et al., 2018), CART is an analytical technique known to reveal potential interactions between predictors (Han and Seo, 2022; Breiman et al., 2017; Tufféry, 2011). This insight into complex interactions between predictors may not have been captured if the study only relied on a conventional linearity-based approach (Yong et al., 2020). Given how many of the past studies that reported Hispanic individuals as being more likely to vape marijuana than non-Hispanic counterparts (Trivers et al., 2018; Taleb et al., 2020; Kritikos, Johnson, & Hodgkin, 2021) used models with linearity-assumption and no interaction terms, the inclusion of non-linear interaction term may be necessary for future studies seeking to understand marijuana vaping among Hispanic individuals.

It should also be noted that this study's outcome – multi-year marijuana vaping – was rarer than other measures of marijuana vaping (e.g., ever-use, initiation). Identifying risk profiles of multi-year marijuana vaping would have been difficult without the use of machine learning techniques, as analysis of such imbalanced data with rare observation is prone to bias when using conventional analytic models (Gradus et al., 2019; Wiles, 2006). Future substance use research may wish to utilize machine learning as demonstrated by this study, especially when examining multi-year substance use to tease out temporary experimenters from prolonged users,

*Sub-study 2 – Discussion of results*

Descriptive statistics of sub-study 2 showed that past-month marijuana vaping, past-month joint smoking, past-month alcohol use, and past-month illegal drug use were higher in states that have legalized non-medical marijuana, while past-month cigarette smoking and past-month cigar smoking were relatively less common in those states. This is consistent with past studies that suggested that the way substance use interacts with marijuana law is complex, and legalization may sometimes decrease substance use by providing safer alternatives, while it may act as a gateway to increase the use of another substance (Coley et al., 2021; Chu, 2015; Buttorff et al., 2023; Mennis, McKeon, & Stahler, 2023; Stormshak et al., 2019). It should be noted that, according to descriptive statistics and its chi-squared tests of sub-study 2, blunt marijuana smoking did not significantly differ by marijuana law while joint marijuana smoking did show a significant difference. While this is not the main focus of this study, it highlights how past studies that examined joint and blunt smoking together as one marijuana smoking may have been incomplete. Indeed, a limited number of studies have compared blunts and joints, and found blunt users were likely to be unemployed, black, men, with a history of legal and social troubles (Cohn et al., 2016; Montgomery et al., 2019). These studies suggest that blunt smoking poses challenges that are distinct from joint smoking (Cohn et al., 2016; Montgomery et al., 2019) and merit individual attention. Future marijuana use studies may benefit by examining joint and blunt use separately.

The influence state legalization has on substance use is hinted in latent classes as well. Upon examining $\gamma$ for classes in Tables 13 and 14, it is evident that the proportion of individuals that belong to classes differs for legal and illegal states. That is, states that legalized non-medical marijuana showed a smaller portion of individuals belonging to the “no substance use” class, while it showed a relatively larger proportion of individuals belonging to the "high risk” class. This is in line with findings by Collins, Ellickson, and Bell (1998), which suggested that the best predictors of concurrent use of drugs include "pro-drug environment" and "pro-drug beliefs." States that legalized non-medical marijuana may have significant overlap with such pro-drug environments and pro-drug beliefs and push individuals in marijuana-legal states toward polydrug use (Stormshak et al., 2019; Fleming et al., 2016).

Since stratified LCA produced near-identical 4-class plots for legal- and illegal states, legalization of non-medical marijuana appears to not alter patterns of polydrug use. Unstratified analysis also produced a 4-class model, which was notably different from an earlier study by Mattingly, Elliott, & Fleischer (2023). Mattingly’s study utilized LCA to examine marijuana-tobacco co-use using wave 4 of the PATH data. Mattingly’s study examined fewer substances to produce a model with more classes, and many of the individual substances appeared as separate classes. The difference between this study and Mattingly’s study is likely attributed to substances not included in Mattingly’s study (i.e., alcohol use, prescription drugs, and illegal drugs). According to Wurpts and Geiser’s 2014 study on LCA methodology, LCA model convergence may be influenced by both the number and quality (i.e., conditional response probability) of indicator variables. The inclusion of alcohol use, prescription drugs, and illegal drugs in my LCA model may have suppressed the separation of individual substance classes to produce a simpler 4-class model overall.

Results of multinomial regression showed that predictors of class membership were similar, but different, in marijuana-legal and illegal states. For example, housing appeared to have had no significance in marijuana-legal states, but it was significantly associated with "alcohol and tobacco" class membership in marijuana-illegal states. Sex only showed significance in predicting the "alcohol only" class in marijuana-legal states, but it showed significance in predicting “alcohol only,” “alcohol and tobacco,” and “high risk” classes in marijuana-illegal states. Sexual orientation only predicted the "high risk" class membership in marijuana legal states but predicted "alcohol only," "alcohol and tobacco," and "high risk" in illegal states. Additional studies are needed to test whether these observed differences translate into statistically significant difference between marijuana-legal and illegal states. Taken together, these results are consistent with the existing notion that polydrug use is common among men, non-Hispanic whites, sexual minorities, and unemployed individuals (Mattingly, 2023; Laudet, 2012). Men may be more likely to partake in polydrug use because they are more likely than women to be risk-takers (Byrnes, Miller, & Schafer, 1999; Charness & Gneezy, 2012; Bornovalova et al., 2005; Leland & Paulus, 2005).

Degree enrollment and health insurance status as predictors deserve special attention, as they have often been ignored by past polydrug use studies. Individuals enrolled in degree programs (i.e., college students) were likely to use just alcohol, but were less likely to concurrently use alcohol and tobacco. This can be explained by how college students tend to drink more than their non-collegiate counterparts (Carter, Brandon, & Goldman, 2010), but they also have negative attitudes toward smoking (Al Omari et al., 2021) that hinder membership to the "alcohol and tobacco" class. A similar pattern was observed for individuals with health insurance, as insurance status was positively and significantly associated with the "alcohol only" class regardless of state law on non-medical marijuana. This similarity between degree enrollment and health insurance status may be partly attributed to shared health awareness and temperament (Morrell, Cohen, & Dempsey, 2008), but it may also be attributed to how many states and colleges require students to have health insurance (Caulfield, 2002). Studies examining interaction between degree enrollment, health insurance, and substance use may further our understanding of this matter.

*Sub-study 1 - Limitations*

The first limitation of sub-study 1 was the data. While the data used for analyses was obtained from a widely used nationally representative study and had a relatively comprehensive set of variables, this research was still secondary in nature and included limitations inherent in non-primary data analyses. For example, there may be local and federal laws that interact with state laws on marijuana (Andrews et al., 2019; Pacula et al., 2015), and those local laws may influence the location of high-access dispensaries, which may have an impact that supersedes and suppresses other variables. The study was also limited by the inclusion of individuals with three waves of matched data. This was a limitation because individuals sometimes dropped out in subsequent waves, which made their data unusable for this study. An indirect measurement of the outcome was necessary but was somewhat of a limitation. Because none of the existing survey items directly asked about multi-year marijuana vaping, a proxy measure based on when they last vaped marijuana at each survey year was necessary. Finally, it should be noted that there are other machine learning techniques (e.g., deep learning, elastic net, and CHAID) that have better computational performance than LASSO and CART. LASSO and CART were chosen over better methods for their interpretability, but their use may still be considered by some as unresolved limitations

*Sub-study 2 - Limitations*

The first limitation of sub-study 2 was its cross-sectional design. While a cross-sectional latent class analysis (LCA) is capable of contributing to the literature, longitudinal LCA (LLCA) could have provided more information by showing how polydrug use may change over time. The second limitation of sub-study 2 comes from the data. Like sub-study 1, this study is still considered secondary in nature, and includes limitations inherent in non-primary data analyses. The third limitation of sub-study 2 was choosing complete case analyses over multiple imputations. While the study had very few missing observations (<5%) and complete case analysis is unlikely to impact its findings in a meaningful way, the resulting loss of observations was still a limitation that influenced the size of the analytic sample. The fourth limitation of sub-study 2 was the scope of the study. While the study was useful in identifying patterns of polydrug use and testing predictors of those polydrug use, this study could not fully explain the mechanism and etiology behind polydrug use. The final limitation of sub-study 2 was having rare observations that limited cell size. This caused the model to collapse categories, use dichotomous indicators instead of multi-category indicators, and use two categories to define the legal status of marijuana (i.e., legal/illegal instead of legal/pending/illegal).

*Conclusion*

Despite these limitations, this dissertation research made contributions to the literature in several ways: Our knowledge about risk factors of marijuana vaping was limited, especially in the context of how it is influenced by non-medical marijuana laws and how focusing on prolonged substance use differs from substance initiation. Sub-study 1 addressed this gap in the literature by identifying important predictors of multi-year marijuana vaping and showing how risk profiles differed by state legalization of non-medical marijuana. Sub-study 1 also generated classification trees that may be used to guide intervention efforts that target the most serious marijuana vapers. In doing so, sub-study 1 verified the usefulness of machine learning techniques in predicting substance use behavior.

Sub-study 2 was one of the first studies to combine LCA and marijuana vaping for studying polydrug use. Sub-study 2 highlights the importance of examining patterns of co-occurring substance use that include marijuana vaping, as it created four-class models (classes: no substance use, alcohol only, alcohol and tobacco, and high risk) as best fit to examine which of the 10 substances (marijuana vaping, joint marijuana smoking, blunt marijuana smoking, alcohol, cigarette, e-cigarette, cigar, other tobacco product, prescription drugs, and illegal drugs) or subset of those substances are used together. Interventions should also be mindful that individuals who vape marijuana may be "high-risk" individuals with polydrug-use problems and prepare for adequate help.

BIBLIOGRAPHY

Advani, I., Gunge, D., Boddu, S., Mehta, S., Park, K., Perera, S., ... & Crotty Alexander, L. E. (2022). Dual use of e-cigarettes with conventional tobacco is associated with increased sleep latency in cross-sectional Study. Scientific Reports, 12(1), 1-10.

Agrawal, A., Budney, A. J., & Lynskey, M. T. (2012). The co-occurring use and misuse of cannabis and tobacco: a review. Addiction, 107(7), 1221-1233.

Agricultural Improvement Act of 2018, Pub. L. No. 115-334

Ahn, W. Y., Ramesh, D., Moeller, F. G., & Vassileva, J. (2016). Utility of machine-learning approaches to identify behavioral markers for substance use disorders: impulsivity dimensions as predictors of current cocaine dependence. Frontiers in Psychiatry, 7, 34.

Al Omari, O., Abu Sharour, L., Heslop, K., Wynaden, D., Alkhawaldeh, A., Al Qadire, M., & Khalaf, A. (2021). Knowledge, attitudes, prevalence and associated factors of cigarette smoking among university students: a cross sectional study. Journal of community health, 46, 450-456.

Alghamdi, M., Al-Mallah, M., Keteyian, S., Brawner, C., Ehrman, J., & Sakr, S. (2017). Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford ExercIse Testing (FIT) project. PloS one, 12(7), e0179805.

Ali, M. M., Teich, J. L., & Mutter, R. (2017). Reasons for not seeking substance use disorder treatment: variations by health insurance coverage. The journal of behavioral health services & research, 44(1), 63-74.

Altman, N., & Krzywinski, M. (2018). The curse (s) of dimensionality. Nat Methods, 15(6), 399-400.

Anastasiou, E., Chennareddy, S., Wyka, K., Shelley, D., & Thorpe, L. E. (2020). Self-reported secondhand marijuana smoke (SHMS) exposure in two New York City (NYC) subsidized housing settings, 2018: NYC Housing Authority and lower-income private sector buildings. Journal of Community Health, 45(3), 635-639.

Andrews, C. M., Pollack, H. A., Abraham, A. J., Grogan, C. M., Bersamira, C. S., D'aunno, T., & Friedmann, P. D. (2019). Medicaid coverage in substance use disorder treatment after the Affordable Care Act. Journal of substance abuse treatment, 102, 1-7.

Ansari-Moghaddam, A., Habybabady, R. H., Shakiba, M., Mirzaei, R., Shahriyari, F., & Aghaei, S. (2012). Predictors of initiation, continuation and transition of drug use in south-eastern Iran. J Pak Med Assoc, 62(7), 698-703.

Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. American psychologist, 55(5), 469.

Audrain-McGovern, J., Rodriguez, D., Testa, S., Alexander, E., & Pianin, S. (2021). Adolescent e-cigarette onset and escalation: associations with internalizing and externalizing symptoms. Journal of Adolescent Health, 68(4), 801-807.

Bailey, S. L., Flewelling, R. L., & Rachal, J. V. (1992). Predicting continued use of marijuana among adolescents: The relative influence of drug-specific and social context factors. Journal of Health and Social Behavior, 51-65.

Bello, M. S., Khoddam, R., Stone, M. D., Cho, J., Yoon, Y., Lee, J. O., & Leventhal, A. M. (2019). Poly-product drug use disparities in adolescents of lower socioeconomic status: Emerging trends in nicotine products, marijuana products, and prescription drugs. Behaviour Research and Therapy, 115, 103-110.

Beulaygue, I. C., & French, M. T. (2016). Got Munchies? Estimating the Relationship between Marijuana Use and Body Mass Index. The journal of mental health policy and economics, 19(3), 123-140.

Biecheler, M. B., Peytavin, J. F., Sam Group, Facy, F., & Martineau, H. (2008). SAM survey on “drugs and fatal accidents”: search of substances consumed and comparison between drivers involved under the influence of alcohol or cannabis. Traffic injury prevention, 9(1), 11-21.

Boccio, C. M., & Jackson, D. B. (2021). Adolescent nicotine and marijuana vaping activity and the use of other illicit substances. Drug and Alcohol Dependence, 219, 108469.

Böke, B. N., Mills, D. J., Mettler, J., & Heath, N. L. (2019). Stress and coping patterns of university students. Journal of college student development, 60(1), 85-103.

Bornovalova, M. A., Daughters, S. B., Hernandez, G. D., Richards, J. B., & Lejuez, C. W. (2005). Differences in impulsivity and risk-taking propensity between primary users of crack cocaine and primary users of heroin in a residential substance-use program. Experimental and clinical psychopharmacology, 13(4), 311.

Bostwick J. M. (2012). Blurred boundaries: the therapeutics and politics of medical marijuana. Mayo Clinic proceedings, 87(2), 172–186

Brady, J. E., & Li, G. (2014). Trends in alcohol and other drugs detected in fatally injured drivers in the United States, 1999–2010. American journal of epidemiology, 179(6), 692-699.

Bray, B. C., Berglund, P. A., Evans-Polce, R. J., & Patrick, M. E. (2021). A latent transition analysis of self-reported reasons for marijuana use during young adulthood. Evaluation & the Health Professions, 44(1), 9-24.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). Classification and regression trees. Routledge.

Brieman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression tree analysis.

Brooks-Russell, A., Ma, M., Levinson, A. H., Kattari, L., Kirchner, T., Anderson Goodell, E. M., & Johnson, R. M. (2019). Adolescent marijuana use, marijuana-related perceptions, and use of other substances before and after initiation of retail marijuana sales in Colorado (2013–2015). Prevention science, 20, 185-193.

Buttorff, C., Wang, G. S., Wilks, A., Tung, G., Kress, A., Schwam, D., & Pacula, R. L. (2023). Impact of recreational cannabis legalization on opioid prescribing and opioid-related hospital visits in Colorado: an observational study. Journal of general internal medicine, 1-8.

Buu, A., Hu, Y. H., Wong, S. W., & Lin, H. C. (2020). Comparing American college and noncollege young adults on e-cigarette use patterns including polysubstance use and reasons for using e-cigarettes. Journal of American College Health, 68(6), 610-616.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. Psychological bulletin, 125(3), 367.

Carter, A. C., Brandon, K. O., & Goldman, M. S. (2010). The college and noncollege experience: A review of the factors that influence drinking behavior in young adulthood. Journal of studies on alcohol and drugs, 71(5), 742-750.

Cassidy, R. N., Meisel, M. K., DiGuiseppi, G., Balestrieri, S., & Barnett, N. P. (2018). Initiation of vaporizing cannabis: individual and social network predictors in a longitudinal study of young adults. Drug and Alcohol Dependence, 188, 334-340.

Caulfield, S. C. (2002). Student Health Insurance. The History and Practice of College Health, 328.

CDC (2022). Defining Adult Overweight & Obesity: Adult Body Mass Index. Retrieved from https://www.cdc.gov/obesity/basics/adult-defining.html on 2022, November 20.

Center for Behavioral Health Statistics and Quality. (2015). Behavioral health trends in the United States: Results from the 2014 National Survey on Drug Use and Health (HHS Publication No. SMA 15–4927, NSDUH Series H–50). Rockville, MD: Substance Abuse and Mental Health Services Administration

Chadi, N., & Hadland, S. E. (2018). Adolescents and perceived riskiness of marijuana: Why care?. Journal of Adolescent Health, 63(4), 377-378.

Chadi, N., Minato, C., & Stanwick, R. (2020). Cannabis vaping: Understanding the health risks of a rapidly emerging trend. Paediatrics & Child Health, 25(Supplement_1), S16-S20.

Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. Journal of Economic Behavior & Organization, 83(1), 50-58.

Charoenporn, V., Charernboon, T., & Mackie, C. J. (2023). Medical cannabis as a substitute for prescription agents: a systematic review and meta-analysis. Journal of Substance Use, 28(4), 522-534.

Cherian, S. V., Kumar, A., & Estrada-Y-Martin, R. M. (2020). E-cigarette or vaping product-associated lung injury: a review. The American Journal of Medicine, 133(6), 657-663.

Chiew, C. J., Liu, N., Wong, T. H., Sim, Y. E., & Abdullah, H. R. (2020). Utilizing machine learning methods for preoperative prediction of postsurgical mortality and intensive care unit admission. Annals of surgery, 272(6), 1133.

Choi, K., Hennrikus, D. J., Forster, J. L., & Moilanen, M. (2013). Receipt and redemption of cigarette coupons, perceptions of cigarette companies and smoking cessation. Tobacco control, 22(6), 418-422.

Choi, N. G., DiNitto, D. M., & Marti, C. N. (2018). Marijuana use among adults: Initiation, return to use, and continued use versus quitting over a one-year follow-up period. Drug and alcohol dependence, 182, 19-26.

Chu, Y. W. L. (2015). Do medical marijuana laws increase hard-drug use?. The Journal of Law and Economics, 58(2), 481-517.

Cohn, A., Johnson, A., Ehlke, S., & Villanti, A. C. (2016). Characterizing substance use and mental health profiles of cigar, blunt, and non-blunt marijuana users from the National Survey of Drug Use and Health. Drug and alcohol dependence, 160, 105-111.

Cole, M. B., Trivedi, A. N., Wright, B., & Carey, K. (2018). Health insurance coverage and access to care for community health center patients: evidence following the Affordable Care Act. Journal of General Internal Medicine, 33(9), 1444-1446.

Coley, R. L., Kruzik, C., Ghiani, M., Carey, N., Hawkins, S. S., & Baum, C. F. (2021). Recreational marijuana legalization and adolescent use of marijuana, tobacco, and alcohol. Journal of Adolescent Health, 69(1), 41-49.

Collins, R. L., Ellickson, P. L., & Bell, R. M. (1998). Simultaneous polydrug use among teens: prevalence and predictors. Journal of substance abuse, 10(3), 233-253.

Coughlin, L. N., Tegge, A. N., Sheffer, C. E., & Bickel, W. K. (2020). A machine-learning approach to predicting smoking cessation treatment outcomes. Nicotine and Tobacco Research, 22(3), 415-422.

Dahlbäck, O. (1990). Criminality and risk-taking. Personality and individual differences, 11(3), 265-272.

Daniulaityte, R., Lamy, F. R., Barratt, M., Nahhas, R. W., Martins, S. S., Boyer, E. W., ... & Carlson, R. G. (2017). Characterizing marijuana concentrate users: A web-based survey. Drug and Alcohol Dependence, 178, 399-407.

Di Forti, M., Marconi, A., Carra, E., Fraietta, S., Trotta, A., Bonomo, M., ... & Murray, R. M. (2015). Proportion of patients in south London with first-episode psychosis attributable to use of high potency cannabis: a case-control study. The Lancet Psychiatry, 2(3), 233-238.

Di Forti, M., Morgan, C., Dazzan, P., Pariante, C., Mondelli, V., Marques, T. R., ... & Murray, R. M. (2009). High-potency cannabis and the risk of psychosis. The British Journal of Psychiatry, 195(6), 488-491.

Di Forti, M., Sallis, H., Allegri, F., Trotta, A., Ferraro, L., Stilo, S. A., ... & Murray, R. M. (2014). Daily use, especially of high-potency cannabis, drives the earlier onset of psychosis in cannabis users. Schizophrenia bulletin, 40(6), 1509-1517.

DiBello, A. M., Benz, M. B., Miller, M. B., Merrill, J. E., & Carey, K. B. (2018). Examining residence status as a risk factor for health risk behaviors among college students. Journal of American College Health, 66(3), 187-193.

Douglas C. E. (2010). Smoke-free Environments Law Project, Restricting the Use of Medical Marijuana in Multi-Unit Residential Settings: Legal and Practical Considerations. Retrieved from http://www.mismokefreeapartment.org/MManalysis.pdf on 2022, November 20.

Ducci, F., & Goldman, D. (2012). The genetic basis of addictive disorders. Psychiatric Clinics, 35(2), 495-519.

Dumortier, A., Beckjord, E., Shiffman, S., & Sejdić, E. (2016). Classifying smoking urges via machine learning. Computer methods and programs in biomedicine, 137, 203-213.

Elvik, R. (2013). Corrigendum to:"Publication bias and time-trend bias in meta-analysis of bicycle helmet efficacy: A re-analysis of Attewell, Glase and McFadden, 2001"[Accid. Anal. Prev. 43 (2011) 1245–1251]. Accident Analysis & Prevention, 60, 245-253.

Etter, J. F. (2015). Electronic cigarettes and cannabis: an exploratory study. European addiction research, 21(3), 124-130.

Fairman, B. J. (2015). Cannabis problem experiences among users of the tobacco–cannabis combination known as blunts. Drug and alcohol dependence, 150, 77-84.

Fleming, C. B., Guttmannova, K., Cambron, C., Rhew, I. C., & Oesterle, S. (2016). Examination of the divergence in trends for adolescent marijuana use and marijuana-specific risk factors in Washington State. Journal of Adolescent Health, 59(3), 269-275.

Foltin, R. W., Fischman, M. W., & Byrne, M. F. (1988). Effects of smoked marijuana on food intake and body weight of humans living in a residential laboratory. Appetite, 11(1), 1-14.

Galaif, E. R., & Newcomb, M. D. (1999). Predictors of polydrug use among four ethnic groups: A 12-year longitudinal study. Addictive behaviors, 24(5), 607-631.

Galli, JA., Sawaya, RA., & Friedenberg, FK. (2011). Cannabinoid hyperemesis syndrome. Current drug abuse reviews, 4(4), 241-249.

Gareth, J., Daniela, W., Trevor, H., & Robert, T. (2013). An introduction to statistical learning: with applications in R. Spinger

Garrow, J. S., & Webster, J. (1985). Quetelet's index (W/H2) as a measure of fatness. International journal of obesity, 9(2), 147-153.

Geiser, C., & Wurpts, I. C. (2014). Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte Carlo study. Frontiers in Psychology: Quantitative Psychology and Measurement, 5.

Gentzke, A. S., Hyland, A., Kiviniemi, M., & Travers, M. J. (2018). Attitudes and experiences with secondhand smoke and smoke-free policies among subsidised and market-rate multiunit housing residents living in six diverse communities in the USA. Tobacco control, 27(2), 194-202.

Gold, M. (2020). Medicinal marijuana, stress, anxiety, and depression: primum non nocere. Missouri Medicine, 117(5), 406.

Gradus, J. L., Rosellini, A. J., Horváth-Puhó, E., Street, A. E., Galatzer-Levy, I., Jiang, T., ... & Sørensen, H. T. (2020). Prediction of sex-specific suicide risk using machine learning and single-payer health care registry data from Denmark. JAMA psychiatry, 77(1), 25-34.

Han, D. H., & Seo, D. C. (2022). Identifying risk profiles for marijuana vaping among US young adults by non-medical marijuana legalization status: A machine learning approach. Drug and Alcohol Dependence, 232, 109330.

Han, D. H., Lee, S. H., Lee, S., & Seo, D. C. (2021). Identifying emerging predictors for adolescent electronic nicotine delivery systems use: A machine learning analysis of the Population Assessment of Tobacco and Health Study. Preventive Medicine, 145, 106418.

Han, D. H., Lee, S., & Seo, D. C. (2020). Using machine learning to predict opioid misuse among US adolescents. Preventive medicine, 130, 105886.

Haque, M., Johal, L., Mathur, P., Nel, W., Rais, A., & Sandhu, R. (2013). Maturation of the adolescent brain Arain M. Neuropsychiatr Dis Treat [Internet], 9, 449-61.

Hartman, R. L., & Huestis, M. A. (2013). Cannabis effects on driving skills. Clinical chemistry, 59(3), 478-492.

Hartman, R. L., Brown, T. L., Milavetz, G., Spurgin, A., Pierce, R. S., Gorelick, D. A., ... & Huestis, M. A. (2015). Cannabis effects on driving lateral control with and without alcohol. Drug and alcohol dependence, 154, 25-37.

Hasin, D. S., Saha, T. D., Kerridge, B. T., Goldstein, R. B., Chou, S. P., Zhang, H., ... & Grant, B. F. (2015). Prevalence of marijuana use disorders in the United States between 2001-2002 and 2012-2013. JAMA psychiatry, 72(12), 1235-1242.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer Series in Statistics. Springer New York.

Hautala, D., Abadie, R., Thrash, C., Reyes, J. C., & Dombrowski, K. (2018). Latent risk subtypes based on injection and sexual behavior among people who inject drugs in rural Puerto Rico. The Journal of Rural Health, 34(3), 236-245.

Hawks, L., Himmelstein, D. U., Woolhandler, S., Bor, D. H., Gaffney, A., & McCormick, D. (2020). Trends in unmet need for physician and preventive services in the United States, 1998-2017. JAMA internal medicine, 180(3), 439-448.

Hefner, K. R., Sollazzo, A., Mullaney, S., Coker, K. L., & Sofuoglu, M. (2019). E-cigarettes, alcohol use, and mental health: Use and perceptions of e-cigarettes among college students, by alcohol use and mental health status. Addictive behaviors, 91, 12-20.

Higley, E., (2019). Defining Young Adulthood. DNP Qualifying Manuscripts. 17.

Hindocha, C., Freeman, T. P., Ferris, J. A., Lynskey, M. T., & Winstock, A. R. (2016). No smoke without tobacco: a global overview of cannabis and tobacco routes of administration and their association with intention to quit. Frontiers in psychiatry, 7, 104.

Hyland, A., Ambrose, B. K., Conway, K. P., Borek, N., Lambert, E., Carusi, C., ... & Compton, W. M. (2017). Design and methods of the Population Assessment of Tobacco and Health (PATH) Study. Tobacco Control, 26(4), 371-378.

Jackson, D. B., Boccio, C. M., & Leal, W. E. (2020). Do youth who vape exhibit risky health lifestyles? Monitoring the future, 2017. Preventive Medicine, 136, 106101.

Jacobs, W., Idoko, E., Montgomery, L., Smith, M. L., & Merianos, A. L. (2021). Concurrent E-cigarette and marijuana use and health-risk behaviors among US high school students. Preventive Medicine, 145, 106429.

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. BMC medical research methodology, 17(1), 1-10.

Jones, C. M., Clayton, H. B., Deputy, N. P., Roehler, D. R., Ko, J. Y., Esser, M. B., ... & Hertz, M. F. (2020). Prescription opioid misuse and use of alcohol and other substances among high school students—Youth Risk Behavior Survey, United States, 2019. MMWR supplements, 69(1), 38.

Junger, M., & Dekovid, M. (2017). Crime as risk-taking: Co-occurrence of delinquent behavior, health-endangering behaviors, and problem behaviors. In Control theories of crime and delinquency (pp. 213-248). Routledge.

Kelly, B. C., & Parsons, J. T. (2008). Predictors and comparisons of polydrug and non-polydrug cocaine use in club subcultures. The American journal of drug and alcohol abuse, 34(6), 774-781.

Kenne, D. R., Fischbein, R. L., Tan, A. S., & Banks, M. (2017). The Use of Substances Other Than Nicotine in Electronic Cigarettes Among College Students. Substance Abuse: Research and Treatment, 11.

Knight, J. R. (2009). Substance use, abuse, and dependence and other risk-taking behaviors. In Developmental-Behavioral Pediatrics (pp. 437-451). WB Saunders.

Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. Progress in Artificial Intelligence, 5(4), 221-232.

Kritikos, A. F., Johnson, J. K., & Hodgkin, D. (2021). Past 30–day marijuana vaping: prevalence and predictors of use in a nationally representative study of US youth. American Journal of Preventive Medicine, 60(2), 258-266.

Kritikos, A. F., Johnson, J. K., & Hodgkin, D. (2021). Past 30–day marijuana vaping: prevalence and predictors of use in a nationally representative study of US youth. American journal of preventive medicine, 60(2), 258-266.

Ksinan, A. J., Spindle, T. R., Thomas, N. S., Eissenberg, T., Dick, D. M., & Spit for Science Working group. (2020). E-cigarette use is prospectively associated with initiation of cannabis among college students. Addictive Behaviors, 106, 106312.

Lanza, H. I., & Teeter, H. (2018). Electronic nicotine delivery systems (e-cigarette/vape) use and co-occurring health-risk behaviors among an ethnically diverse sample of young adults. Substance use & misuse, 53(1), 154-161.

Lanza, H. I., Motlagh, G., & Orozco, M. (2020). E-cigarette use among young adults: A latent class analysis examining co-use and correlates of nicotine vaping. Addictive Behaviors, 110, 106528.

Lanza, S. T., Dziak, J. J., Huang, L., Wagner, A. T., & Collins, L. M. (2015). LCA Stata plugin users' guide (Version 1.2). University park: the methodology center, Penn State.

Laudet, A. B. (2012). Rate and predictors of employment among formerly polysubstance dependent urban individuals in recovery. Journal of Addictive Diseases, 31(3), 288-302.

Lawrence, E., Mollborn, S., Goode, J., & Pampel, F. (2020). Health lifestyles and the transition to adulthood. Socius, 6, 2378023120942070.

Lee, D. C., Crosier, B. S., Borodovsky, J. T., Sargent, J. D., & Budney, A. J. (2016). Online survey characterizing vaporizer use among cannabis users. Drug and alcohol dependence, 159, 227-233.

Lee, J., Kong, G., Kassas, B., & Salloum, R. G. (2021). Predictors of vaping marijuana initiation among US adolescents: Results from the Population Assessment of Tobacco and Health (PATH) study Wave 3 (2015–2016) and Wave 4 (2016–2018). Drug and Alcohol Dependence, 226, 108905.

Lee, J., Krishnan-Sarin, S., & Kong, G. (2023). Social media use and cannabis vaping initiation among US youth. Drug and Alcohol Dependence, 109949.

Leland, D. S., & Paulus, M. P. (2005). Increased risk-taking decision-making but not altered response to punishment in stimulant-using young adults. Drug and alcohol dependence, 78(1), 83-90.

Lemyre, A., Poliakova, N., & Bélanger, R. E. (2019). The relationship between tobacco and cannabis use: a review. Substance Use & Misuse, 54(1), 130-145.

Lemyre, A., Poliakova, N., & Bélanger, R. E. (2019). The relationship between tobacco and cannabis use: a review. Substance Use & Misuse, 54(1), 130-145.

Lenné, M. G., Dietze, P. M., Triggs, T. J., Walmsley, S., Murphy, B., & Redman, J. R. (2010). The effects of cannabis and alcohol on simulated arterial driving: influences of driving experience and task demand. Accident Analysis & Prevention, 42(3), 859-866.

Linden-Carmichael, A. N., Van Doren, N., Bray, B. C., Jackson, K. M., & Lanza, S. T. (2022). Stress and affect as daily risk factors for substance use patterns: an application of latent class analysis for daily diary data. Prevention science, 23(4), 598-607.

Littlefield, A. K., Gottlieb, J. C., Cohen, L. M., & Trotter, D. R. (2015). Electronic cigarette use among college students: Links to gender, race/ethnicity, smoking, and heavy drinking. Journal of American College Health, 63(8), 523-529.

Lopez-Quintero, C., de los Cobos, J. P., Hasin, D. S., Okuda, M., Wang, S., Grant, B. F., & Blanco, C. (2011). Probability and predictors of transition from first use to dependence on nicotine, alcohol, cannabis, and cocaine: Results of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). Drug and alcohol dependence, 115(1-2), 120-130.

Loughran, T. A., Larroulet, P., & Thornberry, T. P. (2018). Definitional elasticity in the measurement of intergenerational continuity in substance use. Child development, 89(5), 1625-1641.

Loukas, A., Marti, C. N., Cooper, M., Pasch, K. E., & Perry, C. L. (2018). Exclusive e-cigarette use predicts cigarette initiation among college students. Addictive behaviors, 76, 343-347.

Loukas, A., Paddock, E. M., Li, X., Harrell, M. B., Pasch, K. E., & Perry, C. L. (2019). Electronic nicotine delivery systems marketing and initiation among youth and young adults. Pediatrics, 144(3).

Lyke, A. (2018). Marijuana Laws: Relationships between State Policy and Drug and Alcohol Violations in On-Campus Housing. Journal of College and University Student Housing, 44(3), 96-111.

Mak, K. K., Lee, K., & Park, C. (2019). Applications of machine learning in addiction studies: A systematic review. Psychiatry research, 275, 53-60.

Mantey, D. S., Clendennen, S. I., Sumbe, A., Wilkinson, A. V., & Harrell, M. B. (2022). Perceived stress and E-cigarette use during emerging adulthood: A longitudinal examination of initiation, progression, and continuation. Preventive Medicine, 160, 107080.

Martins, S. S., Segura, L. E., Levy, N. S., Mauro, P. M., Mauro, C. M., Philbin, M. M., & Hasin, D. S. (2021). Racial and ethnic differences in cannabis use following legalization in US states with medical cannabis laws. JAMA network open, 4(9), e2127002-e2127002

Mattingly, D. T., Elliott, M. R., & Fleischer, N. L. (2023). Latent Classes of Tobacco and Cannabis Use among Youth and Young Adults in the United States. Substance Use & Misuse, 1-11.

McBride, L., & Nichols, A. (2018). Retooling poverty targeting using out-of-sample validation and machine learning. The World Bank Economic Review, 32(3), 531-550.

McCaffrey, D. F., Liccardo Pacula, R., Han, B., & Ellickson, P. (2010). Marijuana use and high school dropout: the influence of unobservables. Health economics, 19(11), 1281-1299.

Meier, M. H., Caspi, A., Ambler, A., Harrington, H., Houts, R., Keefe, R. S., ... & Moffitt, T. E. (2012). Persistent cannabis users show neuropsychological decline from childhood to midlife. Proceedings of the National Academy of Sciences, 109(40), E2657-E2664.

Mennis, J., McKeon, T. P., & Stahler, G. J. (2023). Recreational cannabis legalization alters associations among cannabis use, perception of risk, and cannabis use disorder treatment for adolescents and young adults. Addictive behaviors, 138, 107552.

Miech, R. A., O'Malley, P. M., Johnston, L. D., & Patrick, M. E. (2016). E-cigarettes and the drug use patterns of adolescents. Nicotine & Tobacco Research, 18(5), 654-659.

Miech, R. A., Johnston, L. D., Patrick, M. E., O'Malley, P. M., & Bachman, J. G. (2023). Monitoring the Future national survey results on drug use, 1975–2023: Secondary school students. Monitoring the Future Monograph Series. Ann Arbor, MI: Institute for Social Research, University of Michigan. Available at https://monitoringthefuture.org/results/annual-reports/

Mojtabai, R., Mauro, C., Wall, M. M., Barry, C. L., & Olfson, M. (2020). Private health insurance coverage of drug use disorder treatment: 2005–2018. Plos one, 15(10), e0240298.

Montgomery, L., McClure, E. A., Tomko, R. L., Sonne, S. C., Winhusen, T., Terry, G. E., ... & Gray, K. M. (2019). Blunts versus joints: Cannabis use characteristics and consequences among treatment-seeking adults. Drug and alcohol dependence, 198, 105-111.

Morrell, H. E., Cohen, L. M., & Dempsey, J. P. (2008). Smoking prevalence and awareness among undergraduate and health care students. American Journal on Addictions, 17(3), 181-186.

National Institute on Drug Abuse, 2020. Vaping & Cannabis Trends Among Young Adults (19–22). <https://www.drugabuse.gov/drug-topics/trends-statistics/infographics/ vaping-cannabis-trends-among-young-adults-19-22>

Navon, L., Jones, C. M., Ghinai, I., King, B. A., Briss, P. A., Hacker, K. A., & Layden, J. E. (2019). Risk factors for e-cigarette, or vaping, product use–associated lung injury (EVALI) among adults who use e-cigarette, or vaping, products—Illinois, July–October 2019. Morbidity and Mortality Weekly Report, 68(45), 1034.

NCSL (2023). June 22. State Medical Cannabis Laws. Retrieved from https://www.ncsl.org/research/health/state-medical-marijuana-laws.aspx on 2023, December 13.

Ng Fat, L., & Shelton, N. (2012). Associations between self‐reported illness and non‐drinking in young adults. Addiction, 107(9), 1612-1620.

NIDA (2019). December 24. Cannabis (Marijuana) DrugFacts. Retrieved from https://nida.nih.gov/publications/drugfacts/cannabis-marijuana on 2022, October 20

NIDA (2020). December 15. Study: Surge of teen vaping levels off, but remains high as of early 2020. Retrieved from https://nida.nih.gov/news-events/news-releases/2020/12/study-surge-of-teen-vaping-levels-off-but-remains-high-as-of-early-2020 on 2022, December 5

NIDA (2020). September 15. Vaping, marijuana use in 2019 rose in college-age adults. Retrieved from https://nida.nih.gov/news-events/news-releases/2020/09/vaping-marijuana-use-in-2019-rose-in-college-age-adults on 2022, December 22.

NIDA (2021). April 13. How does marijuana produce its effects?. Retrieved from https://nida.nih.gov/publications/research-reports/marijuana/how-does-marijuana-produce-its-effects on 2022, October 29

Notley, C., Gentry, S., Cox, S., Dockrell, M., Havill, M., Attwood, A. S., ... & Munafò, M. R. (2022). Youth use of e-liquid flavours—a systematic review exploring patterns of use of e-liquid flavours and associations with continued vaping, tobacco smoking uptake or cessation. Addiction, 117(5), 1258-1272.

Pacula, R. L., Powell, D., Heaton, P., & Sevigny, E. L. (2015). Assessing the effects of medical marijuana laws on marijuana use: the devil is in the details. Journal of policy analysis and management, 34(1), 7-31.

Parekh, T., & Fahim, F. (2021). Building risk prediction models for daily use of marijuana using machine learning techniques. Drug and alcohol dependence, 225, 108789.

Park, S., & Kim, J. (2018). Latent class analysis of substance use and predictors of latent class membership among adolescents in the Republic of Korea. Journal of Substance Use, 23(1), 20-28.

Patrick, M. E., Miech, R. A., Johnston, L. D., & O'Malley, P. M. (2023). Monitoring the Future Panel Study annual report: National data on substance use among adults ages 19 to 60, 1976-2022. Monitoring the Future Monograph Series. Ann Arbor, MI: Institute for Social Research, University of Michigan

Peacock, A., Leung, J., Larney, S., Colledge, S., Hickman, M., Rehm, J., ... & Degenhardt, L. (2018). Global statistics on alcohol, tobacco and illicit drug use: 2017 status report. Addiction, 113(10), 1905-1926.

Pericot-Valverde, I., Elliott, R. J., Priest, J. S., Barret, T., Yoon, J. H., Miller III, C. C., ... & Gaalema, D. E. (2019). Patterns of tobacco use among smokers prior to hospitalization for an acute cardiac event: Use of combusted and non-combusted products. Preventive medicine, 128, 105757.

Perrine, C. G., Pickens, C. M., Boehmer, T. K., King, B. A., Jones, C. M., DeSisto, C. L., ... & Lung Injury Response Epidemiology/Surveillance Group. (2019). Characteristics of a multistate outbreak of lung injury associated with e-cigarette use, or vaping—United States, 2019. Morbidity and Mortality Weekly Report, 68(39), 860.

Pokhrel, P., Fagan, P., Kawamoto, C. T., Okamoto, S. K., & Herzog, T. A. (2020). Predictors of marijuana vaping onset and escalation among young adults. Drug and Alcohol Dependence, 216, 108320.

Prowse, R., Sherratt, F., Abizaid, A., Gabrys, R. L., Hellemans, K. G., Patterson, Z. R., & McQuaid, R. J. (2021). Coping with the COVID-19 pandemic: examining gender differences in stress and mental health among university students. Frontiers in psychiatry, 12, 650759.

Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. Marketing Science, 12(1), 103−124.

Reiman, A. (2009). Cannabis as a substitute for alcohol and other drugs. Harm Reduction Journal, 6(1), 1-5.

Riehm, K. E., Young, A. S., Feder, K. A., Krawczyk, N., Tormohlen, K. N., Pacek, L. R., … & Crum, R. M. (2019). Mental health problems and initiation of e-cigarette and combustible cigarette use. Pediatrics, 144(1).

Sakr, S., Elshawi, R., Ahmed, A., Qureshi, W. T., Brawner, C., Keteyian, S., ... & Al-Mallah, M. H. (2018). Using machine learning on cardiorespiratory fitness data for predicting hypertension: The Henry Ford ExercIse Testing (FIT) Project. PLoS One, 13(4), e0195344.

SAMHSA (2019). Center for Behavioral Health Statistics and Quality. 2019 National Survey on Drug Use and Health. Table 2.18B – Alcohol Use in Past Year among Persons Aged 12 or Older, by Age Group and Demographic Characteristics: Percentages, 2018 and 2019. Retrieved from https://www.samhsa.gov/data/sites/default/files/reports/rpt29394/NSDUHD on 2022, November 13.

SAMHSA (2019). Center for Behavioral Health Statistics and Quality. 2019 National Survey on Drug Use and Health. Table 2.20B – Alcohol Use in Past Year among Persons Aged 12 or Older, by Age Group and Demographic Characteristics: Percentages, 2018 and 2019. Retrieved from https://www.samhsa.gov/data/sites/default/files/reports/rpt29394/NSDUHD on 2022, November 13.

SAMHSA (2020). "Key substance use and mental health indicators in the United States: Results from the 2019 National Survey on Drug Use and Health," Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration, Rockville, MD.

Sansone, R. A., & Sansone, L. A. (2014). Marijuana and body weight. Innovations in clinical neuroscience, 11(7-8), 50.

Sawdey, M. D. (2017). Socioecological aspects of tobacco use in college populations.

Schauer, G. L., Rosenberry, Z. R., & Peters, E. N. (2017). Marijuana and tobacco co-administration in blunts, spliffs, and mulled cigarettes: A systematic literature review. Addictive behaviors, 64, 200-211.

Schulenberg, J., Johnston, L., O'Malley, P., Bachman, J., Miech, R., & Patrick, M. (2019). Monitoring the Future national survey results on drug use, 1975-2018: Volume II, college students and adults ages 19-60.

Shahandeh, N., Chowdhary, H., & Middlekauff, H. R. (2021). Vaping and cardiac disease. Heart, 107(19), 1530-1535.

Sheehan, B. E., Grucza, R. A., & Plunk, A. D. (2021, October). Association of racial disparity of cannabis possession arrests among adults and youths with statewide cannabis decriminalization and legalization. In JAMA health forum (Vol. 2, No. 10, pp. e213435-e213435). American Medical Association.

Shi, X., Huang, Y., Huang, J., & Ma, S. (2018). A Forward and Backward Stagewise algorithm for nonconvex loss functions with adaptive Lasso. Computational statistics & data analysis, 124, 235-251.

Shukla, S., & Doshi, H. (2021). Marijuana And Maternal, Perinatal, and Neonatal Outcomes. In StatPearls [Internet]. StatPearls Publishing

Spears, C. A., Jones, D. M., Weaver, S. R., Huang, J., Yang, B., Pechacek, T. F., & Eriksen, M. P. (2019). Sociodemographic correlates of electronic nicotine delivery systems (ENDS) use in the United States, 2016-2017. American Journal of Public Health, 109(9), 1224-1232.

Spindle, T. R., Hiler, M. M., Cooke, M. E., Eissenberg, T., Kendler, K. S., & Dick, D. M. (2017). Electronic cigarette use and uptake of cigarette smoking: A longitudinal examination of US college students. Addictive behaviors, 67, 66-72.

Squeglia, L. M., Jacobus, J., & Tapert, S. F. (2009). The influence of substance use on adolescent brain development. Clinical EEG and neuroscience, 40(1), 31-38.

Stanton, C. A., Sharma, E., Edwards, K. C., Halenar, M. J., Taylor, K. A., Kasza, K. A., ... & Hyland, A. (2020). Longitudinal transitions of exclusive and polytobacco electronic nicotine delivery systems (ENDS) use among youth, young adults and adults in the USA:

findings from the PATH Study Waves 1–3 (2013–2016). Tobacco Control, 29(Suppl 3), s147-s154.

Steigerwald, S., Wong, P. O., Cohen, B. E., Ishida, J. H., Vali, M., Madden, E., & Keyhani, S. (2018). Smoking, vaping, and use of edibles and other forms of marijuana among US adults. Annals of Internal Medicine, 169(12), 890-892.

Stormshak, E. A., Caruthers, A. S., Gau, J. M., & Winter, C. (2019). The impact of recreational marijuana legalization on rates of use and behavior: A 10-year comparison of two cohorts from high school to young adulthood. Psychology of Addictive Behaviors, 33(7), 595.

Suchting, R., Gowin, J. L., Green, C. E., Walss-Bass, C., & Lane, S. D. (2018). Genetic and psychosocial predictors of aggression: variable selection and model building with component-wise gradient boosting. Frontiers in Behavioral Neuroscience, 12, 89.

Sun, R., Mendez, D., & Warner, K. E. (2022). Use of electronic cigarettes among cannabis-naive adolescents and its association with future cannabis use. JAMA Network Open, 5(7), e2223277-e2223277.

Sun, R., Mendez, D., & Warner, K. E. (2022). Use of electronic cigarettes among cannabis-naive adolescents and its association with future cannabis use. JAMA Network Open, 5(7), e2223277-e2223277.

Taleb, Z. B., Kalan, M. E., Bahelah, R., Boateng, G. O., Rahman, M., & Alshbool, F. Z. (2020). Vaping while high: Factors associated with vaping marijuana among youth in the United States. Drug and Alcohol Dependence, 217, 108290.

Trivers, K. F., Phillips, E., Gentzke, A. S., Tynan, M. A., & Neff, L. J. (2018). Prevalence of cannabis use in electronic cigarettes among US youth. JAMA pediatrics, 172(11), 1097-1099.

Tsering, D., & Pal, R. (2009). Role of family and peers in initiation and continuation of substance use. Indian journal of psychological medicine, 31(1), 30-34.

Tufféry, S. (2011). Data mining and statistics for decision making. John Wiley & Sons.

Uddin, S. I., Osei, A. D., Obisesan, O. H., El-Shahawy, O., Dzaye, O., Cainzos-Achirica, M., ... & Blaha, M. J. (2020). Prevalence, trends, and distribution of nicotine and marijuana use in E-cigarettes among US adults: the behavioral risk factor surveillance system 2016–2018. Preventive Medicine, 139, 106175.

United States Department of Health and Human Services. National Institutes of Health. National Institute on Drug Abuse, and United States Department of Health and Human Services. Food and Drug Administration. Center for Tobacco Products. Population Assessment of Tobacco and Health (PATH) Study [United States] Restricted-Use Files. Inter-university Consortium for Political and Social Research [distributor], 2023-10-04. https://doi.org/10.3886/ICPSR36231.v37

Van den Bree, M. B., & Pickworth, W. B. (2005). Risk factors predicting changes in marijuana involvement in teenagers. Archives of General Psychiatry, 62(3), 311-319.

Vaughn, M. G., Beaver, K. M., DeLisi, M., Perron, B. E., & Schelbe, L. (2009). Gene-environment interplay and the importance of self-control in predicting polydrug use and substance-related problems. Addictive Behaviors, 34(1), 112-116.

Volkow, N. D., Baler, R. D., Compton, W. M., & Weiss, S. R. (2014). Adverse health effects of marijuana use. New England Journal of Medicine, 370(23), 2219-2227

Volkow, N. D., Koob, G., & Baler, R. (2015). Biomarkers in substance use disorders. ACS chemical neuroscience, 6(4), 522-525.

Vu, M., Getachew, B., Payne, J. B., Kirchner, T. R., & Berg, C. J. (2018). Initiation, continuation of use and cessation of alternative tobacco products among young adults: A qualitative study. Tobacco Prevention & Cessation, 4.

Wadhahi, A. A., Garvey, L., Edward, K. L., & Beasley, C. (2021). The lived experience of adherence to asthma medication in young adults (18–34 years). Journal of Asthma, 1-16.

Walss-Bass, C., Suchting, R., Olvera, R. L., & Williamson, D. E. (2018). Inflammatory markers as predictors of depression and anxiety in adolescents: statistical model building with component-wise gradient boosting. Journal of Affective Disorders, 234, 276-281.

Walters, S. T., Businelle, M. S., Suchting, R., Li, X., Hébert, E. T., & Mun, E. Y. (2021). Using machine learning to identify predictors of imminent drinking and create tailored messages for at-risk drinkers experiencing homelessness. Journal of Substance Abuse Treatment, 127, 108417.

Wang, G. S., Roosevelt, G., Le Lait, M. C., Martinez, E. M., Bucher-Bartelson, B., Bronstein, A. C., & Heard, K. (2014). Association of unintentional pediatric exposures with decriminalization of marijuana in the United States. Annals of Emergency Medicine, 63(6), 684-689

Wang, Y., Duan, Z., Self-Brown, S. R., Weaver, S. R., Spears, C. A., Zheng, P., ... & Huang, J. (2022). Longitudinal associations between e-cigarette use and onset of multiple modes of cannabis use among US adolescents. Addictive Behaviors, 131, 107316.

Whitehill, J. M., Harrington, C., Lang, C. J., Chary, M., Bhutta, W. A., & Burns, M. M. (2019). Incidence of pediatric cannabis exposure among children and teenagers aged 0 to 19 years before and after medical marijuana legalization in Massachusetts. JAMA Network Open, 2(8), e199456-e199456.

Wiles, N. J., Zammit, S., Bebbington, P., Singleton, N., Meltzer, H., & Lewis, G. (2006). Self-reported psychotic symptoms in the general population: results from the longitudinal study of the British National Psychiatric Morbidity Survey. The British Journal of Psychiatry, 188(6), 519-526.

Winters, K. C., & Arria, A. (2011). Adolescent brain development and drugs. The prevention researcher, 18(2), 21.

Winters, K. C., & Lee, C. Y. S. (2008). Likelihood of developing an alcohol and cannabis use disorder during youth: association with recent use and age. Drug and alcohol dependence, 92(1-3), 239-247.

Wong, S. W., Lohrmann, D. K., Middlestadt, S. E., & Lin, H. C. (2020). Is E-cigarette use a gateway to marijuana use? Longitudinal examinations of initiation, reinitiation, and persistence of e-cigarette and marijuana use. Drug and Alcohol Dependence, 208, 107868.

Xu, X., Pesko, M. F., Tynan, M. A., Gerzoff, R. B., Malarcher, A. M., & Pechacek, T. F. (2013). Cigarette price-minimization strategies by US smokers. American journal of preventive medicine, 44(5), 472-476.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. Perspectives on Psychological Science, 12(6), 1100-1122.

Yong, H. H., Karmakar, C., Borland, R., Kusmakar, S., Fuller-Tyszkiewicz, M., & Yearwood, J. (2020). Identifying smoker subgroups with high versus low smoking cessation attempt probability: A decision tree analysis approach. Addictive behaviors, 103, 106258.

Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American statistical association, 101(476), 1418-1429.

Zwerling, C., Ryan, J., & Orav, E. J. (1990). The efficacy of preemployment drug screening for marijuana and cocaine in predicting employment outcome. Jama, 264(20), 2639-2643.

APPENDIX A

HUMAN SUBJECTS APPROVAL

This dissertation research has been determined as not human subjects research by the Indiana University Human Research Protection Program (IU HRPP), thus no formal human subjects approval was required. An email proof is provided below:

4/4/23, 4:54 PM Gmail - 18942 - Not Human Research Determination

Gmail **Siyoung Choe <dschoe87@gmail.com>**

**18942 - Not Human Research Determination**

**Kuali Notifications** <no-reply@kuali.co> Tue, Apr 4, 2023 at 3:54 PM
To: sichoe@indiana.edu

*If you are not responsible for the IU Kuali Protocols submission for this protocol, this is for informational purposes only and no action is required.*

**NOTICE OF IRB REVIEW NOT REQUIRED**

**Protocol #: 18942**

**Protocol Title: Identifying predictors of marijuana vaping and investigating its polydrug use patterns, using machine learning and latent class analysis (doctoral dissertation research)**

**PI: Seo, Dong Chul**

The above submission was reviewed and IU HRPP staff determined the project is not human subjects research and does not require further review.

Please retain a copy of this email in your research records. You will not receive a separate approval letter.

If you have any questions or require further information, please contact the IU HRPP via email at irb@iu.edu or via phone at (317) 274-8289.

APPENDIX B

SELECTED TABLES AND FIGURES FROM PRELIMINARY ANALYSES

**Please view this section as reference only, as significant changes were made to final analytic model. Non-converging, computationally inaccurate, and irrelevant models are not included.**

**Table A.** PRELIMINARY: Weighted descriptive statistics for selected variables in sub-study 1

| Variable | No marijuana vaping (Weighted N = 1,208,820) | Marijuana vaping (Weighted N = 2,818,815) |
|---|---|---|
| Male | 641,416 (53.1%) | 1,487,564 (52.8%) |
| Currently enrolled in degree Program | 271,343 (22.5%) | 719,552 (25.5%) |
| Satisfaction with social activities:* | | |
| Extremely satisfied | 355,849 (29.7%) | 599,841 (21.5%) |
| Very satisfied | 435,396 (36.3%) | 859,796 (30.8%) |

| Variable | No marijuana vaping (Weighted N = 1,208,820) | Marijuana vaping (Weighted N = 2,818,815) |
|---|---|---|
| Moderately satisfied | 258,634 (21.6%) | 1,001,032 (35.8%) |
| A little satisfied | 100,284 (8.35%) | 197,746 (7.08%) |
| Not at all satisfied | 50,193 (4.18%) | 136,476 (4.88%) |
| Last time you were a bully or threatened others: | | |
| Past month | 55,832 (4.65%) | 58,287 (2.08%) |
| 2-12 months ago | 49,916 (4.16%) | 170,170 (6.06%) |
| Over a year ago | 202,059 (16.8%) | 455,912 (16.2%) |
| Never | 891,827 (74.3%) | 2,123,200 (75.6%) |
| Marijuana use* | 262,758 (31.8%) | 799,510 (57.1%) |
| Hookah use | 69,630 (5.76%) | 189,001 (6.70%) |

| Variable | No marijuana vaping (Weighted N = 1,208,820) | Marijuana vaping (Weighted N = 2,818,815) |
|---|---|---|
| Discounts or coupons for E-cigarettes | 79,714 (6.59%) | 185,429 (6.59%) |
| Perceived harmfulness of e-cigarettes | | |
| Not at all harmful | 57,003 (4.78%) | 89,095 (3.20%) |
| Slightly harmful | 178,303 (14.9%) | 605,712 (21.7%) |
| Somewhat harmful | 400,343 (33.5%) | 870,735 (31.3%) |
| Very harmful | 348,502 (29.2%) | 715,805 (25.7%) |
| Extremely harmful | 209,396 (17.5%) | 505,158 (18.1%) |
| Lax rules about non-combustible tobacco inside the home | | |
| Not allowed anywhere or at any time inside | 741,159 (61.9%) | 1,639,594 (58.5%) |
| Allowed in some places or | 215,741 | 532,875 |

| Variable | No marijuana vaping (Weighted N = 1,208,820) | Marijuana vaping (Weighted N = 2,818,815) |
|---|---|---|
| at some times inside | (18.0%) | (19.0%) |
| Allowed anywhere and | 239,810 | 631,765 |
| at any time inside | (20.0%) | (22.5%) |
| Health insurance | 980,833 | 2,348,810 |
| | (81.1%) | (84.1%) |
| Current housing | | |
| Apartment, condo, or house | 445,460 | 1,042,615 |
| | (84.2%) | (86.6%) |
| Campus housing, fraternity, | | |
| or sorority | 15,437 | 63,532 |
| | (2.92%) | (5.28%) |
| Someplace else | 68,138 | 98,025 |
| | (12.9%) | (8.14%) |
| BMI weight category | | |
| Underweight | 65,696 | 55,427 |
| | (5.54%) | (2.00%) |
| Normal weight | 510,973 | 1,147,259 |
| | (43.1%) | (41.5%) |
| Overweight | 264,681 | 839,996 |
| | (22.3%) | (30.4%) |
| Obesity | 344,058 | 723,194 |

| Variable | No marijuana vaping (Weighted N = 1,208,820) | Marijuana vaping (Weighted N = 2,818,815) |
|---|---|---|
| | (29.0%) | (26.2%) |

*Note.* Numbers are weighted frequencies with column percentages in parentheses. Statistical significance based on Pearson's χ test

*p<0.05

**Table B.** PRELIMINARY: Comparison of models and resulting variable set

| | Stepwise regression | | Ridge regression | | Standard LASSO | | CV LASSO | | Adaptive LASSO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Variable | \|β\| | Variable | Penalized \|β\| | Variable | Penalized \|β\| | Variable | Penalized \|β\| | Variable | Penalized \|β\| |
| 1 | R04R_A_CUR_ESTD_PIPE | 11.2 | R04R_A_CUR_ESTD_PIPE | 17.8 | R04R_A_CUR_ESTD~K | 0.16 | R04R_A_CUR_ESTD~K | 0.14 | R04_AX0085_12M | 0.98 |
| 2 | R04R_A_CUR_ESTD_CIGAR | 9.19 | R04R_A_CUR_ESTD_CIGAR | 15.3 | R04_AX0092 | 0.12 | R04_AX0092 | 0.09 | R04_AX0708_02 | 0.74 |
| 3 | R04_AX0708_02 | 5.33 | R04_AX0708_02 | 12.4 | R04_AX0085_12M | 0.10 | R04_AX0168 | 0.08 | R04R_A_CUR_ESTD_HOOK | 0.41 |
| 4 | R04_AX0708_01 | 2.94 | R04_AX0708_01 | 7.19 | R04_AX0168 | 0.10 | R04_AX0085_12M | 0.08 | R04_AX0168 | 0.25 |
| 5 | R04R_A_SEX_IMP | 2.84 | R04R_A_SEX_IMP | 6.09 | R04R_A_SEX_IMP | 0.09 | R04R_A_SEX_IMP | 0.05 | R04_AX0092 | 0.20 |

| | Stepwise regression | | Ridge regression | | Standard LASSO | | CV LASSO | | Adaptive LASSO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Variable | \|β\| | Variable | Penalized \|β\| | Variable | Penalized \|β\| | Variable | Penalized \|β\| | Variable | Penalized \|β\| |
| 6 | R04_AX0092 | 2.03 | R04_AX0092 | 3.21 | R04_AX0708_02 | 0.04 | R04_AX0708_02 | 0.02 | R04_AR1050 | 0.16 |
| 7 | R04_AM0019 | 1.73 | R04_AX0168 | 2.81 | R04_AM0019 | 0.03 | R04_AM0019 | 0.01 | R04_AM0019 | 0.09 |
| 8 | R04R_A_AM0026_V2 | 1.69 | R04_AR1050 | 2.51 | R04_AE9050 | 0.03 | | | R04_AE9050 | 0.07 |
| 9 | R04_AX0168 | 1.63 | R04_AX0165 | 2.49 | | | | | R04R_A_SEX_IMP | <0.01 |
| 10 | R04_AR1050 | 1.31 | R04_AE9050 | 2.48 | | | | | | |
| 11 | R04_AE9050 | 1.30 | R04_AM0019 | 2.37 | | | | | | |
| 12 | R04_AX0169 | 1.22 | R04R_A_AM0042 | 1.80 | | | | | | |
| 13 | R04_AX0164 | 1.01 | R04R_A_BMI | 1.56 | | | | | | |

| | Stepwise regression | | Ridge regression | | Standard LASSO | | CV LASSO | | Adaptive LASSO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Variable | \|β\| | Variable | Penalized \|β\| | Variable | Penalized \|β\| | Variable | Penalized \|β\| | Variable | Penalized \|β\| |
| 14 | R04_AX0165 | 1.01 | R04_AX0164 | 1.53 | | | | | | |
| 15 | R04_AX0163 | 0.98 | | | | | | | | |
| 16 | R04R_A_BMI | 0.69 | | | | | | | | |
| 17 | R04_AX0167 | 0.64 | | | | | | | | |
| 18 | R04R_A_AM0042 | 0.56 | | | | | | | | |
| 19 | R04_AM0040 | 0.30 | | | | | | | | |

*Note*. LASSO = Least Absolute Shrinkage and Selection Operator; CV = cross-validated

Stepwise regression used backward elimination at threshold = 0.20. Ridge regression used default settings of *plogit*. Standard LASSO and adaptive LASSO used λ parameter that minimize Akaike Information Criterion values (AIC). CV LASSO used 10-fold validation to find the smallest mean-squared prediction error (MSPE), and then estimated model with lambda that minimized MSPE.

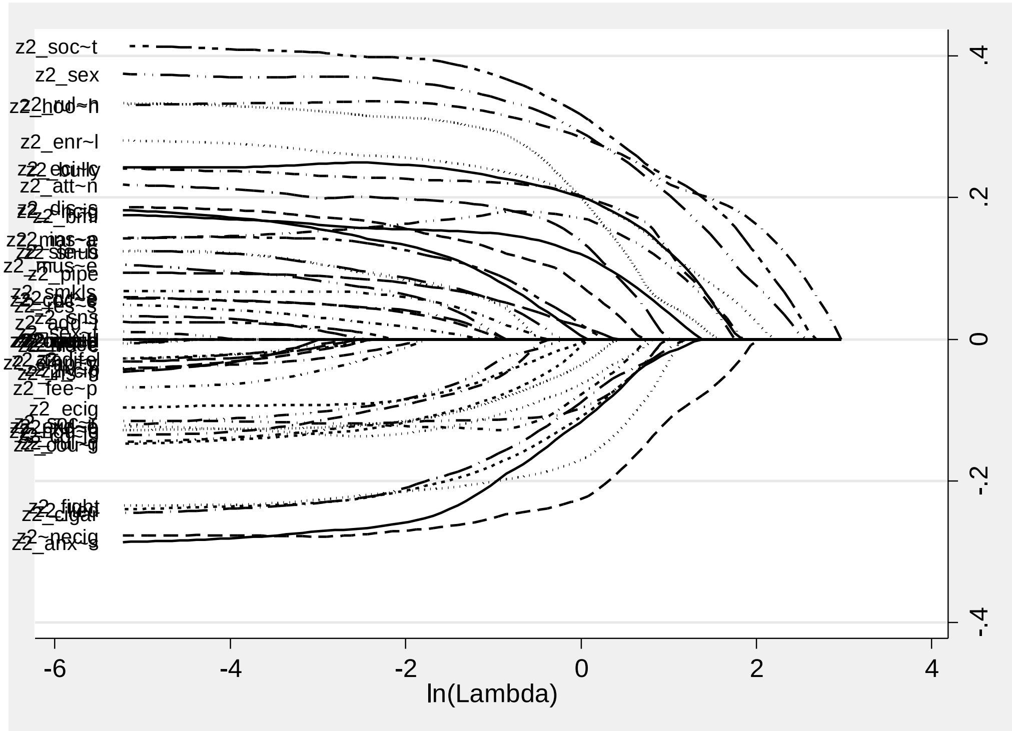
**Table C.** PRELIMINARY: Model performance diagnostics

| | Stepwise regression | Ridge regression | Standard LASSO | CV LASSO | Adaptive LASSO |
|---|---|---|---|---|---|
| Variable count | 19 | 14 | 8 | 7 | 9 |
| Training sample: | | | | | |
| RMSE | - | 1.010 | 0.978 | 0.976 | 0.978 |
| $R^2$ | - | 0.044 | 0.154 | 0.158 | 0.156 |
| Validation sample: | | | | | |
| RMSE | - | 1.020 | 1.063 | 1.064 | 1.062 |
| $R^2$ | - | 0.105 | 0.0967 | 0.085 | 0.108 |
| AUROC | 0.879 | 0.619 | 0.697 | 0.697 | 0.704 |

*Note.* RMSE = root-mean-square error. LASSO = least absolute shrinkage and selection operator; CV = cross-validated. Sqrt = square-root. AUROC = area under the receiver operating characteristic curve.

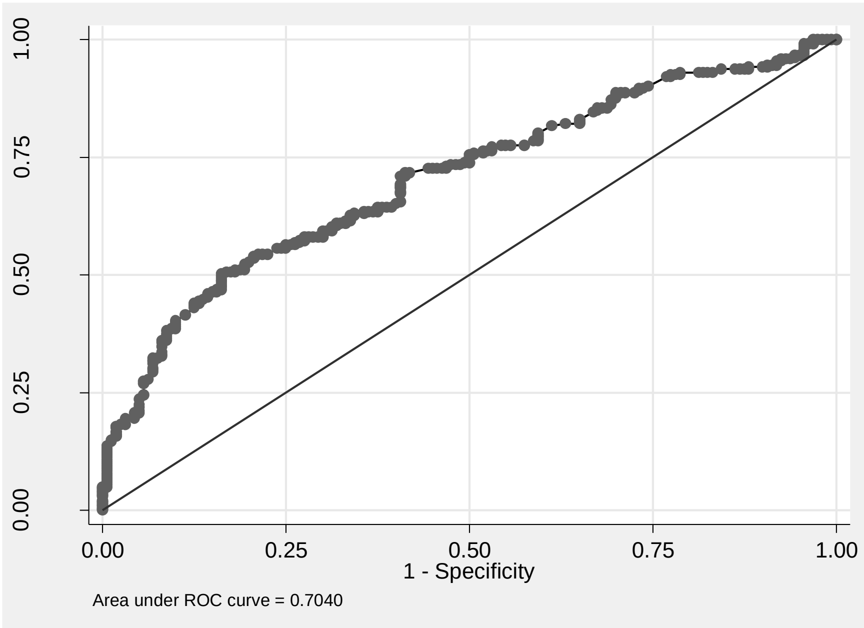
Ratio of training sample to validation sample was 3:1. Square-root LASSO was performed in addition to models above, but its results were similar to that of standard LASSO, and are not presented separately.

**Figure A.** PRELIMINARY: Coefficient path of standardized variables in adaptive LASSO




*Note.* LASSO = least absolute shrinkage and selection operator.

**Figure B.** PRELIMINARY: Overall performance of the best performing model (adaptive LASSO regression) for marijuana vaping continuation among U.S. young adults based on AUROC




*Note.* AUROC = area under the receiver operating characteristic curve; LASSO = Least Absolute Shrinkage and Selection Operator.

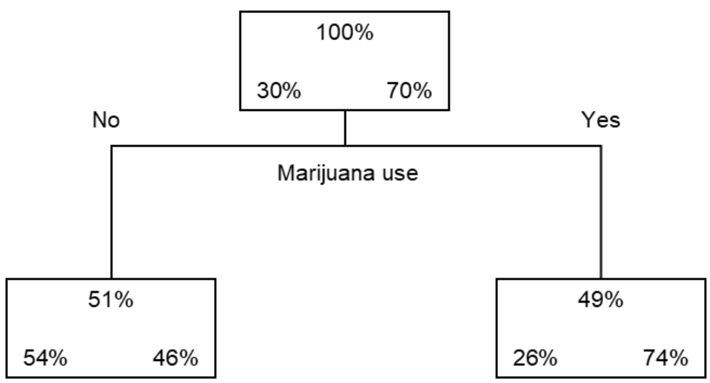
AUROC = 0.704

**Table D.** PRELIMINARY: LASSO regression of marijuana vaping with penalized coefficients and scaled importance

| | Variable | Variable description | Penalized β | Penalized \|β\| | Scaled importance |
|---|---|---|---|---|---|
| 1 | R04_AX0085_12M | Marijuana use | 0.981 | 0.981 | 1.000 |
| 2 | R04_AX0708_02 | Discounts or coupons for E-Cigarettes | -0.736 | 0.736 | 0.750 |
| 3 | R04R_A_CUR_ESTD_HOOK | Hookah use | 0.410 | 0.410 | 0.418 |
| 4 | R04_AX0168 | Not threatening others | 0.252 | 0.252 | 0.257 |
| 5 | R04_AX0092 | Dissatisfied with social activities: | 0.204 | 0.204 | 0.207 |
| 6 | R04_AR1050 | Lax rules about non-combustible tobacco inside the home | 0.162 | 0.162 | 0.165 |
| 7 | R04_AM0019 | Currently enrolled in degree Program | -0.093 | 0.093 | 0.095 |
| 8 | R04_AE9050 | Perceived harmfulness of E- | 0.065 | 0.065 | 0.067 |

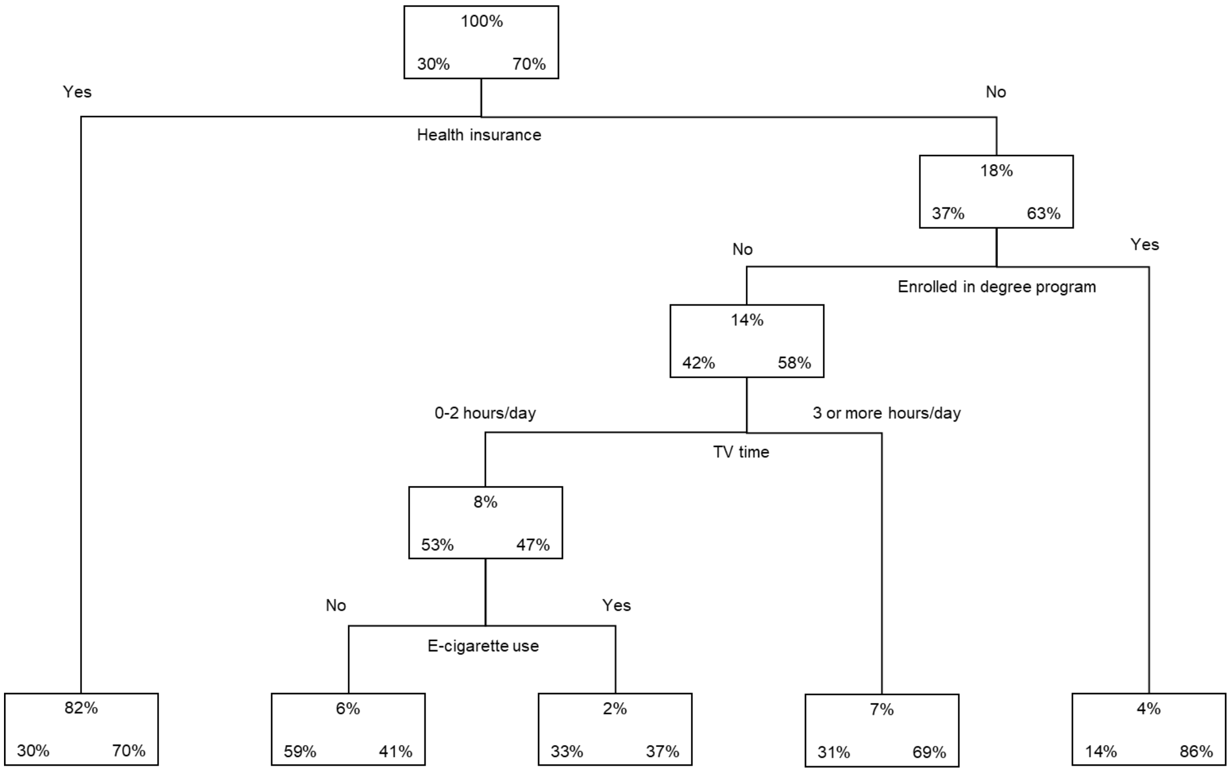
| | Variable | Variable description | Penalized β | Penalized \|β\| | Scaled importance |
|---|---|---|---|---|---|
| | | Cigarettes | | | |
| 9 | R04R_A_SEX_IMP | Sex | -0.002 | 0.002 | 0.002 |

**Figure C.** PRELIMINARY: Overall CART model of marijuana vaping continuation




*Note.* The percentage on the top row of each rectangle indicates the proportion of the node from the entire sample. The percentage on the bottom left of each rectangle indicates the proportion of the respondents within that node who did not vape marijuana, while the percentage on the bottom right indicates the proportion of the respondents within that node who continued marijuana vaping at wave 5.

**Figure D.** PRELIMINARY: Results of CART model of marijuana vaping after manual splitting by respondents' health insurance status.




*Note.* The percentage on the top row of each rectangle indicates the proportion of the node from the entire sample. The percentage on the bottom left of each rectangle indicates the proportion of the respondents within that node who did not vape marijuana, while the percentage on the bottom right indicates the proportion of the respondents within that node who continued marijuana vaping at wave 5.

**Table E.** PRELIMINARY: Weighted descriptive statistics for all variables in sub-study 2

| Variable | No marijuana vaping (Weighted N = 9,360,442) | Marijuana vaping (Weighted N = 6,421,045) |
|---|---|---|
| Cigarette | 2,411,857 | 1,553,138 |

| Variable | No marijuana vaping (Weighted N = 9,360,442) | Marijuana vaping (Weighted N = 6,421,045) |
|---|---|---|
| | (25.8%) | (24.2%) |
| E-cigarette | 1,741,097 | 1,015,139 |
| | (18.6%) | (15.8%) |
| Cigar | 424,934 | 319,912 |
| | (4.54%) | (4.98%) |
| Pipe | 92,695 | 29,046 |
| | (0.99%) | (0.45%) |
| Hookah | 256,452 | 155,082 |
| | (2.74%) | (2.42%) |
| Snus | 86,892 | 71,681 |
| | (0.93%) | (1.12%) |
| Smokeless tobacco | 403,107 | 218,280 |
| | (4.31%) | (3.40%) |
| Alcohol* | 7,716,604 | 5,556,793 |
| | (82.4%) | (86.5%) |
| Marijuana | 1,171,877 | 4,101,348 |
| | (12.5%) | (63.9%) |
| Adderall | 220,564 | 486,865 |
| | (2.36%) | (7.58%) |
| Painkiller* | 320,506 | 418,519 |
| | (3.42%) | (6.52%) |

| Variable | No marijuana vaping (Weighted N = 9,360,442) | Marijuana vaping (Weighted N = 6,421,045) |
|---|---|---|
| Cocaine* | 187,639 (2.00%) | 511,180 (7.96%) |
| Methamphetamine* | 56,038 (0.60%) | 94,776 (1.48%) |
| Heroin* | 104,935 (1.12%) | 303,022 (4.72%) |
| Discounts or coupons for E-cigarettes | 544,814 (5.82%) | 444,833 (6.95%) |
| Male | 4,849,595 (51.8%) | 3,392,626 (52.9%) |
| Currently enrolled in degree program | 2,269,034 (24.2%) | 1,781,880 (27.8%) |
| Satisfaction with social activities:* | | |
| Extremely satisfied | 2,336,351 (25.0%) | 1,389,600 (21.7%) |
| Very satisfied | 3,893,936 (41.7%) | 2,461,301 (38.4%) |
| | 2,320,904 | 1,880,431 |

| Variable | No marijuana vaping (Weighted N = 9,360,442) | Marijuana vaping (Weighted N = 6,421,045) |
|---|---|---|
| Moderately satisfied | (24.8%) | (29.3%) |
| A little satisfied | 596,488 | 501,784 |
| | (6.38%) | (7.82%) |
| Not at all satisfied | 200,635 | 184,244 |
| | (2.15%) | (2.87%) |
| Last time you were a bully or threatened others:* | | |
| Past month | 194,782 | 161,103 |
| | (2.08%) | (2.51%) |
| 2-12 months ago | 232,523 | 223,009 |
| | (2.49%) | (3.48%) |
| Over a year ago | 1,291,740 | 972,137 |
| | (13.8%) | (15.2%) |
| Never | 7,628,588 | 5,055,146 |
| | (81.6%) | (78.9%) |

*Note.* Numbers are weighted frequencies with column percentages in parentheses. Statistical significance based on Pearson's $\chi$ test

*p<0.05

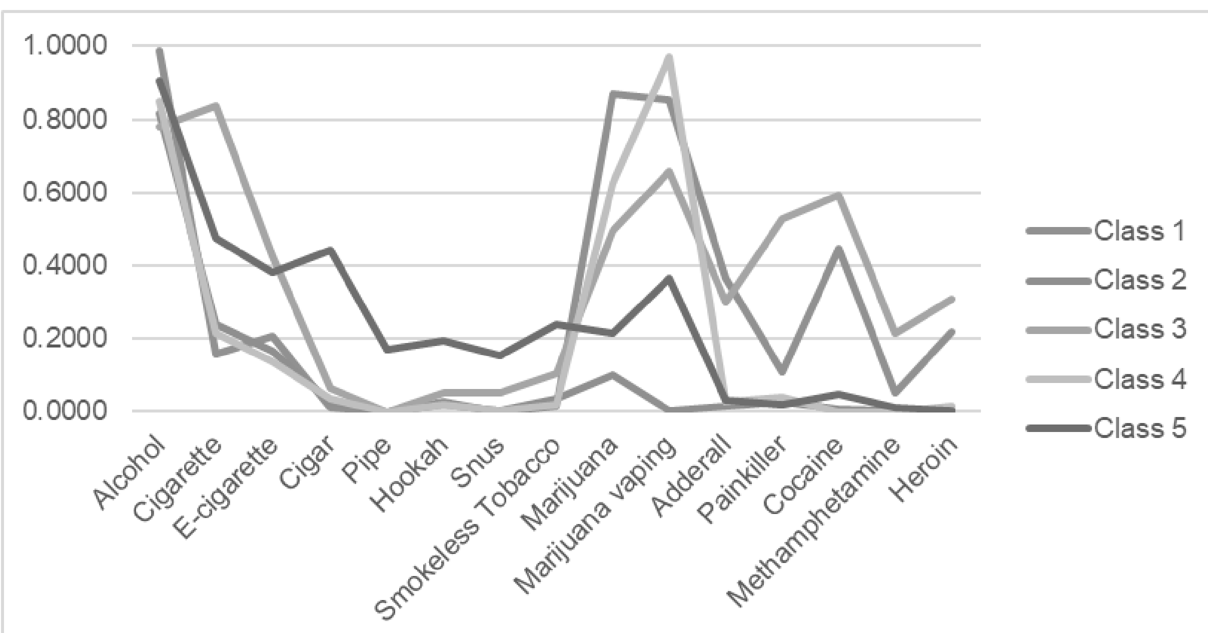
**Table F.** PRELIMINARY Latent class model fit indices

| Classes | Log likelihood | Parameters (ρ) | BIC | Adjusted BIC | AIC | Entropy $R^2$ |
|---|---|---|---|---|---|---|
| 1 | -17174.23 | 15 | 4063.54 | 4015.87 | 3968.37 | N/A |
| 2 | -16324.82 | 30 | 2498.24 | 2399.73 | 2301.56 | 0.71 |
| 3 | -16120.66 | 45 | 2223.42 | 2074.08 | 1925.24 | 0.80 |
| 4 | -15998.49 | 60 | 2112.58 | 1912.40 | 1712.89 | 0.78 |
| **5** | **-15959.79** | **75** | **2168.69** | **1917.67** | **1667.50** | **0.84** |
| 6 | -15915.14 | 90 | 2212.91 | 1911.04 | 1610.20 | 0.73 |

*Note.* AIC = Akaike information criterion; BIC = Bayesian information criterion.

Numbers are based on pseudo-likelihood incorporating weights. Entropy>0.80 indicates good separation of identified groups, based on Ramaswamy et al (1993)

**Figure E.** PRELIMINARY: Item-response probabilities of substance use classes



*Note*. y-axis represents probability. Labels assigned to each class are as follows: Class 1 = conventional marijuana users; Class 2 = low risk users; Class 3 = high risk users; Class 4 = marijuana vapers; Class 5 = tobacco users

**Table G.** PRELIMINARY: Estimated odds ratios (OR) of substance use class membership based on covariates

| | Reference class: Low risk | | | |
|---|---|---|---|---|
| | Conventional marijuana users | High risk | Marijuana vapers | Tobacco users |
| | Odds Ratio (95% CI) | Odds Ratio (95% CI) | Odds Ratio (95% CI) | Odds Ratio (95% CI) |
| Discounts or coupons for E-cigarettes | 2.86 (0.27, 30.1) | 0.99 (0.12, 8.11) | 3.66 (0.55, 24.2) | **252*** (20.4, 3105) |
| Male | 1.67 (0.84, 3.33) | **3.12*** (1.49, 6.56) | 1.33 (0.76, 2.31) | 2.09 (0.96, 4.56) |
| Currently enrolled in degree program | 0.72 (0.45, 1.15) | 0.90 (0.45, 1.79) | 1.08 (0.64, 1.82) | **0.02*** (0.01, 0.10) |
| Dissatisfaction with Social activities | **15.7*** (5.44, 45.2) | **6.20*** (3.03, 12.7) | **6.32*** (3.38, 11.8) | **6.43*** (2.25, 18.4) |

| | Reference class: Low risk | | | |
|---|---|---|---|---|
| | Conventional marijuana users | High risk | Marijuana vapers | Tobacco users |
| | Odds Ratio (95% CI) | Odds Ratio (95% CI) | Odds Ratio (95% CI) | Odds Ratio (95% CI) |
| Not threatened others | **0.68*** | 1.04 | **0.71*** | **0.60*** |
| | (0.53, 0.87) | (0.76, 1.42) | (0.54, 0.93) | (0.43, 0.83) |

*$p<.05$

CURRICULUM VITAE

SIYOUNG CHOE, PHD, MPH, CHES

## EDUCATION

Doctor of Philosophy 2024

Health Behavior (Minor: Epidemiology)

Indiana University School of Public Health

Bloomington, IN

Master of Public Health 2013

Public Health Administration

Indiana University School of Public Health

Bloomington, IN

Bachelor of Science 2010

Biology (Minor: Chemistry)

Indiana University College of Arts and Science

Bloomington, IN

## PROFESSIONAL CERTIFICATION

Health Education Specialist 2017

Certified Health Education Specialist (CHES) #27827

National Commission for Health Education Credentialing (NCHEC)

USA

## RESEARCH ACTIVITIES

### *Published Articles (Refereed)*

1. Shon EJ, **Choe S**, Lee L, & Ki Y. Influenza vaccination among US college or university students: A systematic review. Am J Health Promot. 2021;35(5):708-719.
2. Kwak H, Sa J, **Choe S**, Chaput J, & Chung J. Correlates of Highly Caffeinated Beverage Consumption among Korean Adolescents. Osong Public Health Res Perspect. 2021;12(6):374–384.
3. Sa J, **Choe S**, Seo J, Chaput JP, Cho B, Gazmararian J,… & Moen J. Sleep duration and weight gain among students at a historically black university. Health Behav Policy Rev. 2021;8(1):71-82.
4. Sa J, Cho BY, Chaput JP, Chung J, **Choe S**, Gazmararian J,… & Han T. Sex and racial/ethnic differences in the prevalence of overweight and obesity among US college students, 2011–2015. J Am Coll Health. 2021;69(4):413-421.
5. Sa J, Kwon E, Seo J, **Choe S**, Chaput JP, Gazmararian J,… & Kim Y. Obesity-related behaviors of students at historically black colleges and universities and students at non-historically black colleges and universities. Health Behav Policy Rev. 2020;7(6):570-583.
6. Sa J, **Choe S**, Cho BY, Chaput JP, Kim G, Park CH,… & Kim Y. Relationship between sleep and obesity among US and South Korean college students. BMC Public Health. 2020;20(1):1-11.
7. Sa J, **Choe S**, Cho BY, Chaput JP, Lee J, & Hwang S. Sex and racial/ethnic differences in suicidal consideration and suicide attempts among US college students, 2011-2015. Am J Health Behav. 2020;44(2):214-231.

8. Seo D-C, **Choe S**, & Torabi M. Is waist circumference ≥102/88cm better than body mass index ≥30 to predict hypertension and diabetes development regardless of gender, age group, and race/ethnicity? Meta-analysis. Prev Med. 2017.
9. Seo D-C, Torabi MR, Kim N, Lee CG, & **Choe S**. Smoking among East Asian college students: Prevalence and correlates. Am J Health Behav. 2013;37(2):199-207.

***Published Article (Invited)***

1. **Choe S**, Sa J, Chaput J, & Kim D. Effectiveness of obesity interventions among South Korean children and adolescents and importance of the type of intervention component: a meta-analysis. Clin Exp Pediatr. 2022;65(2):98-107.

***Oral Presentations (Refereed)***

1. Sa J, **Choe S**, Chaput JP, Chung J, Nelson B, & Hwang S. "Sleep duration and weight gain among students at a historically black university," The 148th Annual Meeting and Exposition of the American Public Health Association (APHA), San Francisco, CA, 2020.
2. Sa J, **Choe S**, Kim G, To B, & Lewis S. "Sex and racial/ethnic differences in the prevalence of overweight and obesity among US college students, 2011 to 2015," The 148th Annual Meeting and Exposition of the American Public Health Association (APHA), San Francisco, CA, 2020.
3. Sa J, **Choe S**, Cho B, Hwang S, & To B. "Sex and racial/ethnic differences in suicidal consideration and suicide attempts among US college students, 2011-2015," The 148th

Annual Meeting and Exposition of the American Public Health Association (APHA), San Francisco, CA, 2020.

4. Sa J, **Choe S**, Cho B, Kim G, & Choi Y. "Relationship between sleep and obesity among US and South Korean college students," The 148th Annual Meeting and Exposition of the American Public Health Association (APHA), San Francisco, CA, 2020.

***Poster Presentations (Refereed)***

1. Sa J, **Choe S**, Hwang S. "A meta-analysis of obesity interventions among South Korean children and adolescents," 2023 Doswell Health Informatics Conference, Denton, TX, 2023.
2. Sa J, Chaput J, Chung J, **Choe S**, & Yu J. "Sex, racial, ethnic differences in sleep quality & its relationship with BMI in US college students," The Obesity Society 37th Annual Scientific Meeting, Las Vegas, NV, 2019
3. **Choe S**, Lin H-C, & Seo D-C. "Internet overuse as a predictor of traditional bullying victimization in the United States," The 17th Annual Scientific Meeting of the American Academy of Health Behavior (AAHB), Tucson, AZ, 2017.
4. Choe AY, Jones VF, Leslie K, **Choe S**, & Seo JM. "Perceptions of Healthcare in Populations of Korean Descent in a Midwest Urban Setting," The American Academy of Pediatrics (AAP) National Conference & Exhibition, San Francisco, CA, 2016.
5. **Choe S**, Lin H-C, & Seo D-C. "Patient utilization of diabetes screening services under the Affordable Care Act of 2010," The 143rd American Public Health Association (APHA) Annual Meeting and Exposition, Chicago, IL, 2015.

6. **Choe S** & Lin H-C. “Effect of the Early Retiree Reinsurance Program under the Affordable Care Act on insurance status among elderly Medicare ineligible in the US,” The 143rd American Public Health Association (APHA) Annual Meeting and Exposition, Chicago, IL, 2015.
7. **Choe S** & Lin H-C. “Utilization of obesity screening and counseling under the Affordable Care Act of 2010,” The 14th Annual Scientific Meeting of the American Academy of Health Behavior (AAHB), Charleston, SC, 2014.
8. Lin H-C, **Choe S**, & Chen B. “Removing the barrier of cost to smoking cessation medications under the Affordable Care Act of 2010,” The 14th Annual Scientific Meeting of the American Academy of Health Behavior (AAHB), Charleston, SC, 2014.
9. Seo D-C, Torabi MR, Kim N, Lee CG, & **Choe S**. “Smoking among East Asian college students: Prevalence and correlates,” The 13th Annual Scientific Meeting of the American Academy of Health Behavior (AAHB), Santa Fe, NM, 2013.

***Other Presentations (Seminar/Workshop)***

1. **Choe S**, Kim T, Wong S-W, & Hu Y-H. “Secondary dataset resources: Medical Expenditure Panel Survey,” TobWell Meeting, Bloomington, IN, Sep 9, 2016.
2. **Choe S**. "IOM Obesity Prevention Guideline," TobWell Meeting, Bloomington, IN, Mar 28, 2014.
3. **Choe S**. "Disparities in Obesity Treatment," TobWell Meeting, Bloomington, IN, Feb 14, 2014.

4. **Choe S** & Siela J. "The Youth Risk Behavior Survey (YRBS)," Quarterly Meeting of the Indiana State Department of Health Division of Epidemiology, Indianapolis, IN, Mar 27, 2013.

## TEACHING ACTIVITIES

### *Instructor Of Record*

Visiting Instructor 2017 – 2020

**MBI 131** Community Health Perspectives *(undergraduate-level)*

**KNH 125** Introduction to Public Health *(undergraduate-level)*

**KNH 321** National and Global Health Policy *(undergraduate-level)*

**KNH 340** Public Health Internship *(undergraduate-level)*

**KNH 441/541** Environmental Public Health *(undergraduate- & graduate-level)*

Miami University

Oxford, OH

Associate Instructor 2014 – 2016

**SPH H263** Personal Health *(undergraduate-level)*

Indiana University School of Public Health

Bloomington, IN

## FUNDING ACTIVITIES

### *Teaching Grant Proposal*

EHS Interdisciplinary Teaching Grant 2018

Principal Investigators: En-Jung Shon, PhD, Lena Lee, PhD, Yoon Ki, PhD

**Role: Grant Developer & Consultant** *(funded; $10,000)*

Miami University College of Education, Health, and Society

Oxford, OH

***Research Grant Proposal***

EHS Interdisciplinary Teaching Grant 2018

Principal Investigators: En-Jung Shon, PhD, Lena Lee, PhD, Yoon Ki, PhD

**Role: Grant Developer & Consultant** *(funded; $10,000)*

Miami University College of Education, Health, and Society

Oxford, OH

NIH Academic Research Enhancement Award (AREA) Program (R15) 2015

Principal Investigator: Hsien-Chang Lin, PhD

**Role: Grant Developer** *(unfunded)*

Indiana University School of Public Health

Bloomington, IN

Korea Research Foundation Grant 2013

Principal Investigator: Dong-Chul Seo, PhD

**Role: Grant Developer** *(unfunded)*

Indiana University School of Public Health

Bloomington, IN

Patient-Centered Outcomes Research Institute (PCORI) Research Grant 2012 – 2013

Principal Investigators: Dong-Chul Seo, PhD & Hsien-Chang Lin, PhD

**Role: Grant Developer** *(unfunded)*

Indiana University School of Public Health

Bloomington, IN

***Program Grant Proposal***

CDC Strategy 1: School-Based Surveillance Grant (CDC-RFA-PS-13-1308) 2013

Principal Investigator: Jeena Siela, MPH

**Role: Grant Developer** *(funded; amount unknown)*

Indiana State Department of Health

Indianapolis, IN

Nina Mason Pulliam Charitable Trust (NMPCT) 501(c)(3) Program Grant 2012

Principal Investigator: Mary Byrne

**Role: Grant Developer** *(unfunded)*

The Indiana Youth Group

Indianapolis, IN

**SERVICE ACTIVITIES**

***Professional Service***

Ad-Hoc Peer-Reviewer for Academic Journals 2014 – Present

Health Education & Behavior

Journal of School Health

American Journal of Health Behavior

### ***Institutional Service***

Academic Advisor for the Department of Kinesiology and Health 2017 – 2020

Miami University College of Education, Health, and Society

Oxford, OH

Faculty Club Advisor for Body Language in Business Club 2017 – 2020

Miami University

Oxford, OH

Contact person for new and potential public health students 2017 – 2019

Miami University

Oxford, OH

Ad-Hoc Member of the Public Health Steering Committee 2017 – 2019

Miami University

Oxford, OH

### ***Community Service***

Career Consultant 2014

Indiana University Science Association for Koreans

Bloomington, IN

Sunday School Teacher & Principal 2005 – 2009, 2013

Korean Ministry at St. Paul Catholic Center

Bloomington, IN

International Student Orientation Volunteer 2007

Indiana University Office of International Services

Bloomington, IN

## OTHER POSITIONS/ACTIVITIES

### *Employments*

Research Consultant 2020 – 2021

Miami University College of Education, Health, and Society

Oxford, OH

Surveillance Program Coordinator 2013

Indiana State Department of Health (Contractor; Knowledge Services, Inc.)

Indianapolis, IN

Youth Risk Behavior Survey (YRBS) Administrator 2013

Indiana State Department of Health

Indianapolis, IN

### *Internships*

Graduate Intern 2013

Indiana State Department of Health

Indianapolis, IN

Graduate Intern 2012

Monroe County Health Department

Bloomington, IN

## ORGANIZATIONAL MEMBERSHIPS

### *Professional Organizations*

| | |
|---|---|
| American Public Health Association (APHA) | 2012 – Present |
| Society for Public Health Education (SOPHE) | 2016 – 2022 |
| American Academy of Health Behavior (AAHB) | 2015 – 2022 |
| American School Health Association (ASHA) | 2012 – 2017 |
| Indiana Public Health Association (IPHA) | 2011 – 2017 |
| American Bar Association (ABA) | 2010 – 2011 |

### *Other Organizations*

| | |
|---|---|
| Tobacco, Obesity, and Behavioral Wellness (TobWell) Research Group | 2014 – Present |
| Indiana University Asian Culture Center (ACS) | 2005 – 2009 |