Cross-sectional study: Association between Marijuana Use and Suicidal Ideation among Young Adults aged 18-25

John Platt

Abstract

Background: As of 2023, about a quarter of young adults used marijuana in the past year, and more than 10% experienced suicidal ideation (SI). However, research from 2023 onward on the association, including non-linearities, between marijuana use and SI among young adults is lacking.

Objective: This study aimed to examine the association in-depth between past-year marijuana use and past-year SI among young adults aged 18-25.

Methods: Using data on 41,873 young adults from the 2021-2023 National Survey on Drug Use and Health (NSDUH), survey-weighted logistic regression adjusted for sociodemographic, substance use, and mental health covariates was applied. To capture non-linearities, unadjusted restricted cubic splines analysis was employed.

Results: The logistic regression model for standardized marijuana use showed that marijuana use increases the odds of SI (aOR = 1.15, 95% CI 1.09-1.23). Subsequently, the restricted cubic splines model revealed an L-shaped association between marijuana use and SI with a plateau in odds of SI at moderate use.

Conclusion: Marijuana use increases the odds of SI with an L-shaped pattern in the association. This shows that organizations like the Substance Abuse and Mental Health Services Administration (SAMHSA) and clinicians who are looking to provide timely marijuana use and SI treatment should aim for marijuana use prevention and minimization efforts.

Keywords: Young adults; Marijuana use; Suicidal ideation; National Survey on Drug Use and Health; Association; Logistic regression; Restricted cubic splines

Introduction

Substance use and mental health are issues that have become increasingly problematic in the US. In 2023, 14.1% of US young adults (age 18-25) had a co-occurring substance use and mental health disorder, which is more than other subgroups (Substance Abuse and Mental Health Services Administration, 2023). Moreover, more than 10% of young adults experienced suicidal ideation (SI) in the past year. Nearly a quarter of US civilians aged 12 or older used illicit drugs, and the percentage that used illicit drugs increased between 2021 and 2023. Furthermore, marijuana is the most commonly used illicit substance (> 61 million users in 2023). Thus, problems with marijuana use and SI among young adults merit further attention.

In research, illicit substance use exhibits a well-established relationship with mental health. Marijuana use can worsen mental health through changes in brain structure as well as function (Filbey *et al.*, 2014). Additionally, illicit LSD use is associated with increased risk of SI (Han *et al.*, 2022). Moreover, a predictive model was developed to evaluate risk of SI among adults who abuse marijuana (Choi *et al.*, 2021). While all of these studies examine the relationship between illicit substance use and mental health, this research fills in the recency gap, and provides an in-depth, novel perspective on the association between marijuana use and SI among young adults aged 18-25. Therefore, this research addresses the following research question: to what extent is past-year marijuana use associated with past-year suicidal ideation (SI) among young adults aged 18-25?

This research informs a variety of stakeholders, which include, but are not limited to: Substance Abuse and Mental Health Services Administration (SAMHSA), clinicians, and young adults. By uncovering insights on the marijuana use-SI association among young adults, we provide targeted recommendations to the SAMHSA agency on improvements in marijuana use and SI treatment. This research is also useful for clinicians who are looking to provide timely mental health treatment for those with substance use problems (Han *et al.*, 2022). Finally, this research promotes education for young adults on the connection between marijuana use and SI.

Methodology

1. Research Overview

This research focuses on inferential analysis, gaining a comprehensive understanding of the association between marijuana use and SI among young adults that is interpretable to a wider audience. By gaining an interpretable picture of this association, results of this research can be clearly communicated to public health officials and healthcare facilities that are looking to enhance behavioral treatment. However, there are many confounding variables that influence the marijuana use-SI association. For example, e-cigarette use is associated with illicit drug use and mental health problems among college students (Grant *et al.*, 2019). While several confounding variables are controlled for in our analysis, this study does not establish causality because this is a cross-sectional study that cannot show temporal relationships.

2. Dataset Description

This study uses secondary, public use data (confidential information eliminated) from the 2021-2023 National Survey on Drug Use and Health (NSDUH). The NSDUH data is the leading source of population-based statistics on behavioral health information like tobacco use, alcohol use, drug use, and mental health. The dataset is cross-sectional, and its unit of observation is the civilian, noninstitutionalized population aged 12 or older in the United States. The data was collected with web-based interviews, including a state-based, multistage, and stratified area probability sample. The 2021-2023 dataset has about 170,000 rows and 2,600 columns, revealing the importance of compressing the data to specific columns and population subsets.

3. Data Cleaning

To collect the data, the data was reduced from over 2600 columns to a total of 22 columns, including our independent variable, dependent variable, covariates, and survey design as well as weight variables (See Appendix I). All the columns selected from the data are imputed (missing values statistically imputed) and recoded (derived from one or more edited variables), which almost never contain missing values and are better for analysis. Subsequently, all the categorical variables were decoded (e.g. changing 0 to No and 1 to Yes). Then, substance use responses coded as never or none (91, 93, 991, 993) were recoded to 0 for analysis. Finally, values of -9 or

99 among covariates were replaced with NA. *While doing our analysis*, we excluded respondents from the data who are not young adults aged 18-25, making the final sample size 41,873.

4. Analytical Methods

The data was downloaded from the SAMHSA website (Substance Abuse and Mental Health Services Administration, 2025). All statistical analyses were performed in VS Code software (v1.104.2, https://code.visualstudio.com/) and R software (v4.5.1, https://www.r-project.org/). A significance level of $p < 0.05$ was considered statistically significant. To account for the complex design of the 2021-2023 NSDUH survey, strata, primary sampling units, and weights were incorporated anytime analysis was done. For statistical analysis, pseudo maximum likelihood (see Appendix II) logistic regression models, that fit our data types of numerical predictor and binary response, were applied, which examined the strength of the substance use and mental health association supported in previous research (Qi et al., 2024). These models captured the relationship between past-year marijuana use and past-year SI after controlling for sociodemographic, substance use, and mental health covariates. Two distinct models were done: an unadjusted model 1 with just our main predictor and a fully adjusted model 2 based on all covariates (see Appendix IV).

Pseudo maximum likelihood logistic regression follows assumptions including binary outcome, low multicollinearity, correct model specification, linearity in the logit, sufficient sample size, and complex survey design. To satisfy these assumptions, data preparation methods involved dependent variable recoding, specifying complex survey design, and computing survey design measures. Methods to assess model quality contained generation of a correlation matrix as well as variance inflation factors, design-adjusted wald tests, and assessing model fit (see Appendix V). Nominal covariates with more than two levels were dummy coded. Logistic regression produced log odds (beta coefficients) and 95% confidence interval, which were exponentiated to obtain the odds ratio (OR). OR > 1 indicates higher odds in the exposed group compared to the unexposed group, while OR < 1 reveals lower odds. Finally, when a significant difference was seen between the unadjusted and adjusted models, logistic regression analysis was stratified by the most significant confounder.

After logistic regression, restricted cubic splines were utilized (see Appendix III), which modeled the non-linear relationship between substance use and mental health shown in previous research (Qi et al., 2024), and fits our data types of continuous predictor and binary response. For large sample sizes, 5 knots is the standard, so these knots were incorporated at the 5th, 27.5th, 50th, 72.5th, and 95th percentile of past-year marijuana use. A single unadjusted restricted cubic splines model (without covariates) was done with our continuous marijuana use predictor and binary response variable of SI (see Appendix IV).

After fitting the restricted cubic splines model, an interpretable visual was extracted from the model, plotting yearly marijuana use on the x-axis, and the odds ratios of SI on the y-axis. Odds ratios were extracted from exponentiating the difference in the log odds between any specified value and the reference value (0 past-year marijuana use). OR > 1 communicates that odds of the outcome are higher for the specified value of exposure compared to the reference value, and vice versa. However, it's important to keep in mind the sample size for each spline segment, which involves a key assumption of restricted cubic splines analysis: sufficient sample size per spline segment. In addition, this visual was interpreted based on notable places where the odds of SI decrease, stay constant, or increase.

Results

Among the 41,873 young adults, the overall prevalence of SI was greater than 10% (Weighted: 12.9%, Unweighted: 12.7%).

Overall, Table 1 showed a significant proportion of young adults are white (52.4%) or hispanic (23.8%), have completed some degree of college education (40.1%), are unemployed (9.6%) or employed full time (39.8%), have private health insurance (56.7%), have four people in household (24.5%), and have a household income of $75,000 or more (33.3%).

Regarding substance use and mental health patterns, Table 1 revealed that young adults used alcohol 44.1 days on average (± 68.3) in the past year, while nicotine vaping 4.6 days on average in the past month (± 9.9). Young adults most frequently felt nervous most of the time (33.9%) in

the past year, often mentioned everything feeling like an effort all of the time (27.7%) in the past year, and had severe difficulty concentrating a significant amount of the time (11.5%).

Table 1.

Young adults' characteristics according to the presence of suicidal ideation[a]

	Overall, n=41,873	**Suicidal Ideation, n=5303**	**No Suicidal Ideation, n=36570**	***P*-value[b]**
Gender, %				<.001
Male	50.1	40.99	51.5	
Female	49.9	59.01	48.5	
Race or Ethnicity, %				.003
Nonhisp White	52.4	56.6	51.8	
Hispanic	23.8	20.7	24.3	
NonHisp Black/Afr Am	13.5	11.8	13.7	
NonHisp Asian	6.3	5.6	6.4	
Nonhisp more than one race	3.0	3.9	2.8	
Nonhisp Native Am/AK Native	0.6	0.9	0.6	
NonHisp Native HI/Other Pac Isl	0.4	0.3	0.4	
Highest Education Obtained, %				<.001
Some coll/Assoc Dg	40.1	42.4	39.8	
High school grad	33.0	34.4	32.8	

College graduate	16.3	12.1	16.9	
Less high school	10.6	11.1	10.5	
Work Status, %				<.001
Employed full time	39.8	38.9	39.9	
Other (incl. not in labor force)	26.0	22.7	26.5	
Employed part time	24.7	27.8	24.2	
Unemployed	9.6	10.6	9.4	
Private Health Insurance, %	56.7	58.9	56.3	.04
Household Size, %				.006
Four people in household	24.5	22.6	24.8	
Three people in household	24.3	25.8	24.0	
Two people in household	20.3	22.2	19.9	
Five people in household	13.6	12.4	13.8	
6 or more people in household	11.3	9.9	11.5	
One person in household	6.1	7.2	5.9	
Household Income, %				0.12
$75,000 or More	33.3	31.1	33.6	
20,000 to 49,999	29.7	31.7	29.4	
Less than $20,000	22.6	23.2	22.5	

50,000 to 74,999	14.3	13.9	14.4	
Past Year Alcohol Use, days	44.1 ± 68.3	57.3 ± 80.8	42.1 ± 66.1	<.001
Past Month Binge Drinking, days	1.1 ± 3.0	1.6 ± 4.1	1.0 ± 2.7	<.001
Past Year Cigarette Use, days	1.5 ± 5.8	2.6 ± 7.5	1.3 ± 5.5	<.001
Past Month Nicotine Vaping, days	4.6 ± 9.9	7.2 ± 11.9	4.2 ± 9.6	<.001
Past Year Received Inpatient Substance Use Treatment, %	1.4	4.1	1.1	<.001
Past Year How Often Felt Nervous, %				<.001
Most of the time	33.9	33.9	33.9	
Some of the time	28.1	14.1	32.3	
All of the time	23.9	45.4	17.3	
A little of the time	11.4	5.3	13.2	
None of the time	2.8	1.3	3.2	
Past Year How Often Everything Felt Like an Effort, %				<.001
All of the time	27.7	55.3	19.3	
Most of the time	25.7	25.1	25.8	

Some of the time	23.0	12.3	26.3	
A little of the time	15.3	4.9	18.5	
None of the time	8.3	2.3	10.1	
Past Year Difficulty Concentrating, %				<.001
No difficulty	32.7	11.4	36.5	
Mild difficulty	32.4	24.1	33.8	
Moderate difficulty	23.4	34.6	21.5	
Severe difficulty	11.5	29.9	8.2	

[a] Suicidal ideation is defined as whether or not suicide was seriously considered in the past year.

[b] According to Weighted Mann-Whitney U test or the Chi-square test

With the exception of income, Table 1 also showed that the frequency distribution for all covariates differed between young adults who experienced SI and young adults who did not experience SI ($p<.05$). For instance, according to Weighted Mann-Whitney U tests, the averages for all the substance use indicators were significantly different for young adults that experienced SI relative to young adults that did not experience SI ($p<.05$). So, with the exception of income and everything felt like an effort (due to multicollinearity), these were all covariates that were accounted for in the survey-weighted logistic regression analysis.

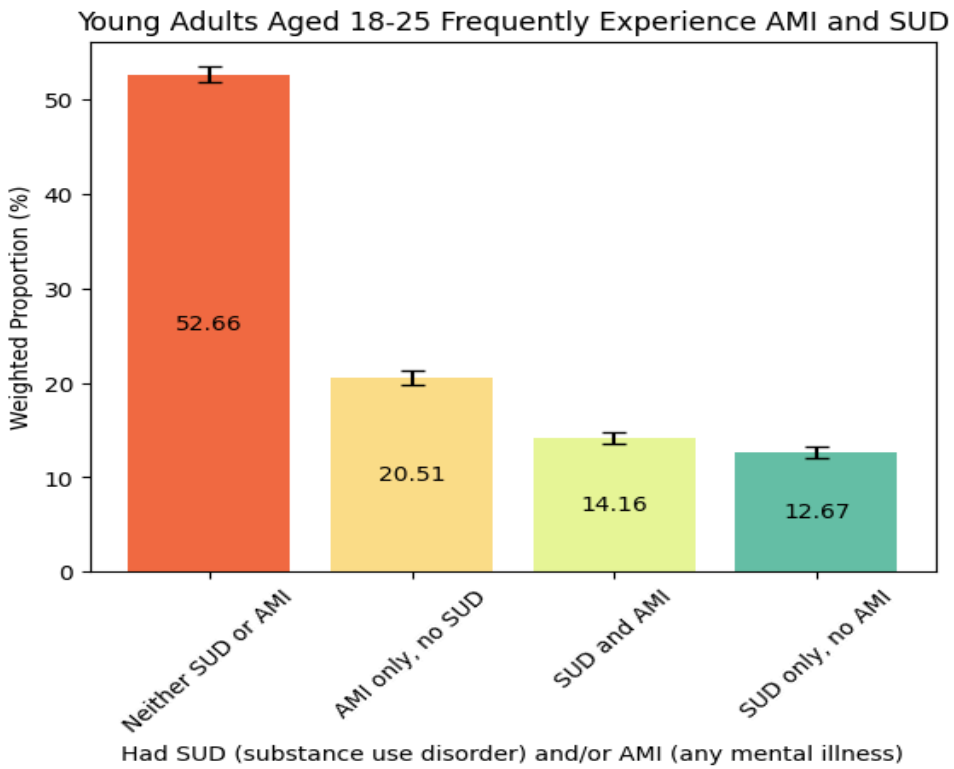

Figure 1. A significant percentage of young adults have a SUD and AMI relative to other age groups in the US (Substance Abuse and Mental Health Services Administration, 2024). Estimates on SUD and AMI from the 2021-2023 NSDUH data were derived using tabulation of frequencies, given survey weights, strata, and primary sampling units.

According to Figure 1, young adults reported experiencing co-occurring SUD and AMI 14.16% of the time, which is significantly more than other age groups in the US. Moreover, more than 20% of young adults reported experiencing AMI only, while more than 12% of young adults reported experiencing SUD. This further reveals why young adults specifically are important to be studied in substance use and mental health research.

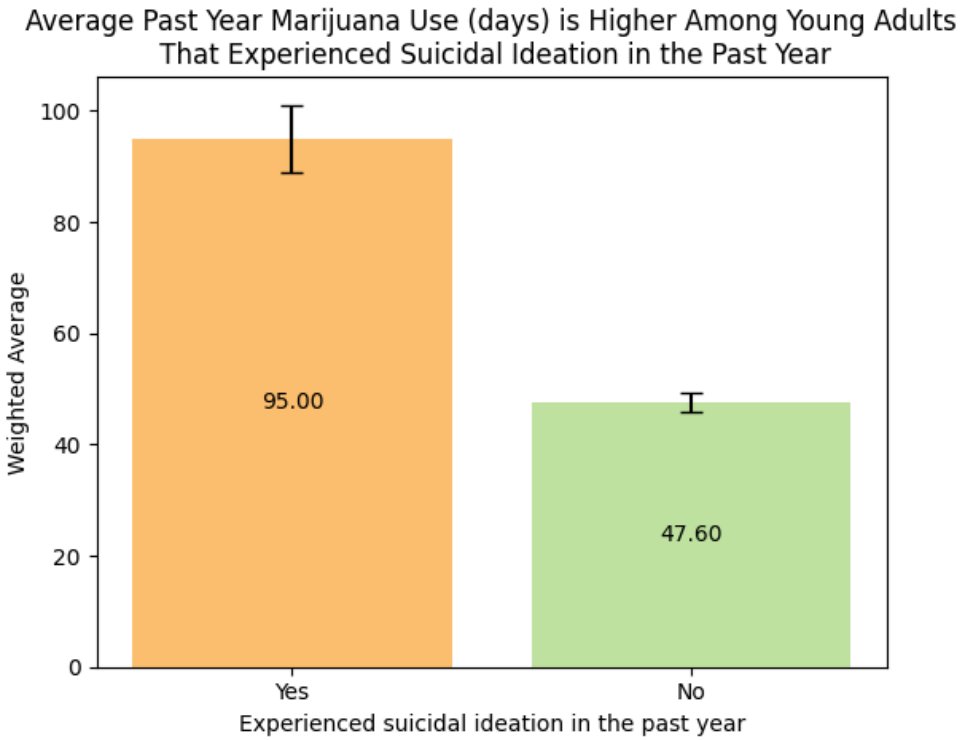

Figure 2. A weighted Mann-Whitney U test confirmed this significant association between past year marijuana use and suicidal ideation (p<.001). Weighted averages and error bars were derived from taylor linearization, which takes into account survey weights, strata, and primary sampling units of this data: 2021-2023 NSDUH.

According to Figure 2, an association existed between SI and past-year marijuana use (days). Notably, not included in the figure, the weighted average past-year marijuana use (days) for young adults overall was 53.74, which was more frequent than alcohol use. This figure justified why we examined the association between past-year marijuana use and SI among young adults.

Table 2.

Survey-weighted unadjusted and adjusted odds ratios and 95% confidence intervals of suicidal ideation according to the primary studied factor

	Unadjusted OR (95% CI)[a]	**Adjusted OR (95% CI)[b]**
Past Year Marijuana Use		
0-90 days (ref)	1.00	1.00

90-180 days	2.04 (1.62, 2.57)*	1.68 (1.16, 2.44)*
180-270 days	2.13 (1.76, 2.58)*	1.55 (1.13, 2.13)*
270-360 days	2.72 (2.39, 3.09)*	1.48 (1.16, 1.90)*
Past Year Marijuana Use, Standardized[c]	1.33 (1.29, 1.37)*	1.15 (1.09, 1.23)*

* *p*-value < 0.05 indicating a significant association

[a] Unadjusted ORs: 95% CIs were estimated by a survey-weighted logistic regression model that contained only the main studied factor (yearly marijuana use).

[b] Adjusted ORs: 95% CIs were estimated by a survey-weighted logistic regression model that contained all significant covariates, which included sex, race/ethnicity, education level, work status, private health insurance, household size, alcohol use, binge alcohol use, cigarette use, nicotine vaping, receipt of in-patient substance use treatment, and mental health covariates (felt nervous, difficulty concentrating).

[c] The odds ratio for the standardized marijuana use variable represents how much higher the odds of suicidal ideation are for each 1 SD increment (~110.6 days) in yearly marijuana use.

In Table 2, past-year marijuana use was positively associated with SI. Compared with young adults that used marijuana 0-90 days in the past year (ref), those that used marijuana for 90-180 days were 104% more likely to experience SI (OR=2.04, 95% CI 1.62-2.57). The odds of SI plateaued between 180 and 270 days of marijuana use (OR=2.13, 1.76-2.58). Then, the odds of SI significantly increased between 270 and 360 days of marijuana use (OR=2.72, 2.39-3.09), suggesting a non-linear relationship between marijuana use and SI. Finally, for each 1 SD increment (~110.6) days in yearly marijuana use, the odds of SI increased by 33% (OR=1.33, 1.29-1.37).

According to the adjusted logistic regression analysis, past-year marijuana use was positively associated with SI, but the association was a lot weaker. Compared with young adults that used marijuana 0-90 days in the past year (ref), young adults that used marijuana for 90-180 days were 68% more likely to experience SI (aOR=1.68, 1.16-2.44). Similarly, young adults that used marijuana for 180-270 days were 55% more likely to experience SI (aOR=1.55, 1.13-2.13). Young adults that used marijuana for 270-360 days were 48% more likely to experience SI (aOR=1.48, 1.16-1.90). Finally, for each 1 SD increment in yearly marijuana use, the odds of SI

increased by 15% (aOR=1.15, 1.09-1.23). These weaker associations explained why the logistic regression analysis in Table 3 was stratified by the most significant confounder: past-year how often felt nervous.

Table 3.

Survey-weighted unadjusted and adjusted odds ratios and 95% confidence intervals of suicidal ideation according to the primary studied factor, stratified by how often felt nervous (past year)

	Felt Nervous at least most of the time		**Felt Nervous between some of the time and none of the time**	
	Unadjusted OR (95% CI)	**Adjusted OR (95% CI)**	**Unadjusted OR (95% CI)**	**Adjusted OR (95% CI)**
Past Year Marijuana Use				
0-90 days (ref)	1.00	1.00	1.00	1.00
90-180 days	1.57 (1.10, 2.23)*	1.48 (0.96, 2.27)	2.03 (1.38, 2.99)*	2.37 (1.43, 3.92)*
180-270 days	1.69 (1.28, 2.23)*	1.34 (0.89, 1.99)	2.51 (1.79, 3.53)*	2.36 (1.52, 3.66)*
270-360 days	1.82 (1.49, 2.23)*	1.59 (1.24, 2.04)*	2.12 (1.44, 3.11)*	1.38 (0.80, 2.40)

* p-value < 0.05 indicating a significant association

In Table 3, past-year marijuana use was significantly positively associated with SI among young adults, and a significant drop between the unadjusted and adjusted odds ratios was no longer seen. However, for young adults who felt nervous between some of the time and never, they had higher odds of SI at moderate levels (180-270) of marijuana use (aOR = 2.36, 1.52-3.66) compared to young adults who felt nervous frequently (aOR = 1.34, 0.89-1.99). Conversely, for young adults who felt nervous at least most of the time, they had higher odds of SI at heavy levels (270-360 days) of marijuana use (aOR = 1.59, 1.24-2.04) compared to young adults who felt nervous less frequently (aOR = 1.38, 0.80-2.40).

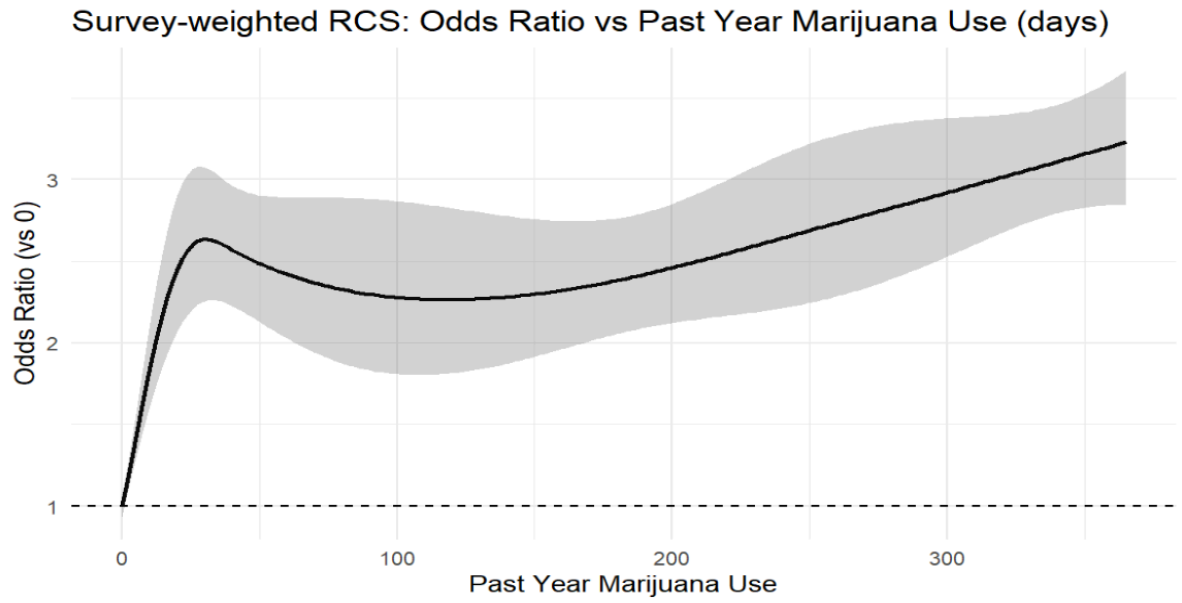

Figure 3. Past year marijuana use is associated with increased odds for suicidal ideation as we see a general increase in the odds ratio for suicidal ideation for each one unit increase in past-year marijuana use. However, the increase in odds of suicidal ideation is especially sharp between 0 and 25 days before decreasing slightly between roughly 25 days of marijuana use and 110 days. The odds of suicidal ideation then increases at a roughly constant rate between 110 days and 360 days. This visualization was extracted from an unadjusted restricted cubic splines model between past year marijuana use and suicidal ideation, which models non-linear relationships with flexible cubic functions joined together at knots.

According to Figure 3, without adjusting for covariates, a significant L-shaped association was found between marijuana use and SI. Overall, we observed a sharp rise in odds of SI for non-users vs users followed by a plateau for moderate users, and a slight increase in odds of SI for heavy users. This L-shaped association was supported by the 95% CI (gray area on the graph).

Discussion

1. Summary

The findings of this cross-sectional study among 41,873 young adults revealed a positive association between marijuana use and SI. This association, however, was mixed when stratified by nervousness frequency. Young adults who felt nervous at least most of the time in the past year had higher odds of SI at heavy levels of marijuana use, while young adults who felt nervous

between some of the time and none of the time had higher odds of SI at moderate levels of marijuana use. This association is also L-shaped with a sharp increase in odds of SI when going from no marijuana use to any use, a plateau in odds of SI at moderate use, and arguably a slight increase in the odds of SI at heavy levels of marijuana use.

2. Explanation

The positive association between marijuana use and SI among young adults aged 18-25 can be explained by changes in the brain that occur from chronic marijuana use. Filbey *et al.* (2014) finds that chronic marijuana users have lower OFC (orbitofrontal cortex) gray matter volume than non-using controls. The OFC gray matter volume plays an important role in emotional regulation, decision-making, reward processing, and behavioral control. So, when the brain has less OFC gray matter, emotional regulation, decision-making, reward processing, and behavioral control are all impaired.

However, the logistic regression results stratified by nervousness was unexpected. Research shows that stress-related symptoms are positively related to SI among young adults with depression and rumination being explanatory factors (Polanco-Roman *et al.*, 2016). Nevertheless, these findings can be explained by potential biases in this cross-sectional study, including self-report bias and recall bias. In addition, the definition for nervousness varies significantly from person to person based on coping skills. For example, someone with lots of nervousness can still have great coping skills, in turn reporting lower levels of nervousness relative to someone that experiences lots of nervousness with worse coping skills.

Secondly, the sharp rise in odds of SI between 0 and 25 days of yearly marijuana use was unexpected. These results can be attributed to the data being imbalanced. Most young adults in this data did not report using marijuana ever or in the past year, leading to there being lots of 0 in our data (see Appendix IV). Since the reference value for odds of SI is 0 marijuana use, it's plausible that massive imbalance in our data caused the spike in odds between 0 and 25 marijuana use.

3. Limitations

Additional limitations of this study include that it cannot establish causality, not all relevant confounders were accounted for, and there were imbalances with data on our covariates. First, this study cannot establish causality because we cannot tell whether our exposure of marijuana use happened before SI because the exposure and outcome are measured simultaneously. Secondly, potential confounders such as impulsivity (McHugh *et al.*, 2025) specific to college students and discrimination (Jones *et al.*, 2017) specific to young adults in racial minorities were not adjusted for in our logistic regression analysis. Third, for several of our covariates, the frequency of the outcome was not evenly distributed across each category, which likely led to some outliers greater than 6 in our deviated residuals vs fitted values plots (see Appendix IV).

4. Public Health Implications

Nevertheless, these findings reveal for clinicians, SAMHSA, healthcare facilities, and young adults the importance of marijuana use prevention programs, or at least limiting marijuana use to moderate use to minimize the prevalence of SI. Preventing marijuana use significantly reduces the odds of SI relative to limiting use to moderate.

5. Next Steps

To address the data imbalance limitation, future studies might decide to look at the association between marijuana use and SI again, but restrict the data to just marijuana users and see what the association looks like among users. Furthermore, future studies, to establish a temporal relationship, may want to look at longitudinal data between marijuana use and SI. To add in the possibility of causation, prospective cohort or experimental studies could be done. Finally, future studies could do an adjusted restricted cubic splines model (including covariates) between marijuana use and SI, but those studies should recognize that interpretability is sacrificed with the addition of more predictor variables to an RCS model.

References

Choi, J., Chung, J., & Choi, J. (2021). Exploring Impact of Marijuana (Cannabis) Abuse on Adults Using Machine Learning. *International Journal of Environmental Research and Public Health*, *18*(19), 10357. https://doi.org/10.3390/ijerph181910357

Filbey, F. M., Aslan, S., Calhoun, V. D., Spence, J. S., Damaraju, E., Caprihan, A., & Segall, J. (2014). Long-term effects of marijuana use on the brain. *Proceedings of the National Academy of Sciences*, *111*(47), 16913–16918. https://doi.org/10.1073/pnas.1415297111

Grant, J. E., Lust, K., Fridberg, D. J., King, A. C., & Chamberlain, S. R. (2019). E-Cigarette Use (Vaping) is Associated with Illicit Drug Use, Mental Health Problems, and Impulsivity in University Students. *Annals of Clinical Psychiatry : Official Journal of the American Academy of Clinical Psychiatrists*, *31*(1), 27. https://pmc.ncbi.nlm.nih.gov/articles/PMC6420081/

Han, B., Blanco, C., Einstein, E. B., & Compton, W. M. (2022). Mental health conditions and receipt of mental health care by illicit lysergic acid diethylamide (LSD) use status among young adults in the United States. *Addiction*. https://doi.org/10.1111/add.15789

Jones, D. M., Masyn, K. E., & Spears, C. A. (2017). Discrimination, psychological functioning, and substance use among U.S. young adults aged 18–28, 2017. *Experimental and Clinical Psychopharmacology*. https://doi.org/10.1037/pha0000502

McHugh, R., McLafferty, M., Brown, N., Ward, C., Walsh, C. P., Bjourson, A. J., McBride, L., Brady, J., O'Neill, S., & Murray, E. K. (2025). The mediating role of impulsivity on suicidal behaviour among higher education students with depression and substance abuse disorders. *Alcohol*, *124*, 89–96. https://doi.org/10.1016/j.alcohol.2025.01.002

Polanco-Roman, L., Gomez, J., Miranda, R., & Jeglic, E. (2016).

Stress-related symptoms and suicidal ideation: The roles of rumination and depressive symptoms vary by gender. *Cognitive Therapy and Research, 40*(5), 606–616. https://doi.org/10.1007/s10608-016-9775-3

Qi, P., Huang, M., & Zhu, H. (2024). Association between alcohol drinking frequency and depression among adults in the United States: a cross-sectional study. *BMC Psychiatry*, *24*(1). https://doi.org/10.1186/s12888-024-06296-9

Substance Abuse and Mental Health Services Administration. (2024, July). *Key substance use and mental health indicators in the United States: Results from the 2023 National Survey on Drug Use and Health* (HHS Publication No. PEP24-07-021, NSDUH Series H-59).

Substance Abuse and Mental Health Services Administration. (2025, February 13). *Download NSDUH data files*. SAMHSA. https://www.samhsa.gov/data/data-we-collect/nsduh-national-survey-drug-use-and-health/datafiles/2021-2023

Appendix

I. Data Documentation

Variable Description	Values	Data Type (numeric, categorical, text, date)	% Missingness	Role (Predictor, Covariate, Response)
Sex at birth - Imputation Revised	1 = Male, 2 = Female	Categorical (binary)	NA	Covariate
Race/Hispanicity recode	1 = NonHisp White, 2 = NonHisp Black/Afr Am, 3	Categorical (nominal)	NA	Covariate

	= NonHisp Native Am/AK Native, 4 = NonHisp Native HI/Other Pac Isl, 5 = NonHisp Asian, 6 = NonHisp more than one race, 7 = Hispanic			
Recoded-Education Categories	1 = Less high school, 2 = High school grad, 3 = Some coll/Assoc Dg, 4 = College graduate, 5 = 12 to 17 year olds	Categorical (ordinal)	NA	Covariate
Employment Status 18+ - Imputation Revised	1 = Employed full time, 2 = Employed part time, 3 = Unemployed, 4 = Other (incl. not in labor force), 99 = 12-17 year olds	Categorical (nominal)	NA	Covariate
Private Health Insurance - Imputation Revised	1 = Yes, R does have private health insurance,	Categorical (binary)	NA	Covariate

	2 = No, R does not have private health insurance			
Recode - Imputation -Revised # Persons in Household	1 = One person in household, 2 = Two people in household, 3 = Three people in household, 4 = Four people in household, 5 = Five people in household, 6 = 6 or more people in household	Categorical (ordinal)	NA	Covariate
RC-Total Family Income Recode	1 = Less than $20,000, 2 = $20,000 - $49,999, 3 = $50,000 - $74,999, 4 = $75,000 or More	Categorical (ordinal)	NA	Covariate
Alcohol Frequency Past Year - Imputation Revised	Range = 1 - 365, 991 = Never Used Alcohol, 993 = Did Not Use Alcohol Past Year	Numerical (discrete)	NA	Covariate

Cig Frequency Past Month - Imputation Revised	Range = 1 - 30, 91 = Never Used Cigarettes, 93 = Did Not Use Cigarettes Past Month	Numerical (discrete)	NA	Covariate
Nicotine Vaping Frequency Past Month - Imputation Revised	Range = 1 - 30, 91 = Never Vaped Nicotine, 93 = Did Not Vape Nicotine Past Month, -9 = Data not collected for the year	Numerical (discrete)	NA	Covariate
Binge Alcohol Frequency Past Month - Imputation Revised	Range = 0 - 30, 91 = Never Used Alcohol, 93 = Did Not Use Alcohol Past Month	Numerical (discrete)	NA	Covariate
Recoded-Received Substance Use Treatment As An Inpatient - Past Year	0 = No, 1 = Yes, -9 = Data not collected for the year	Categorical (binary)	NA	Covariate
How Often Felt Nervous Worst Month in Past Year -	1 = All of the time, 2 = Most of the time, 3 =	Categorical (ordinal)	NA	Covariate

Imputation Revised	Some of the time, 4 = A little of the time, 5 = None of the time, 99 = Legitimate Skip			
How Often Felt Everything Effort Worst Month in Past Year - Imputation Revised	1 = All of the time, 2 = Most of the time, 3 = Some of the time, 4 = A little of the time, 5 = None of the time, 99 = Legitimate Skip	Categorical (ordinal)	NA	Covariate
Difficulty Concentrating One Month in Past 12 Months - Imputation Revised	1 = No difficulty, 2 = Mild difficulty, 3 = Moderate difficulty, 4 = Severe difficulty, 99 = Legitimate Skip	Categorical (ordinal)	NA	Covariate
Marijuana Frequency Past Year - Imputation Revised	Range = 1 - 365, 991 = Never Used Marijuana, 993 = Did Not Use Marijuana Past Year	Numerical (discrete)	NA	Predictor
Adult Seriously Thought About	. = Aged 12-17, 0 = No, 1 = Yes	Categorical (binary)	NA	Response

Killing Self Past Year - Imputation Revised				
Final person-level sample weight	Len: 8 (e.g. 3276.46987)	Numeric (discrete)	NA	NA
Variance stratum	Len: 5 (e.g. 40031)	Numeric (discrete)	NA	NA
Variance primary sampling unit	Len: 2 (e.g. 2)	Numeric (discrete)	NA	NA
Year in which data was collected	Len: 4 (2021-2023)	Numeric (discrete)	NA	NA
Recoded-Only AMI, Only SUD, Both, or Neither - PY-DSM-5-ANY	. = Aged 12-17, 1 = SUD only, no AMI, 2 = AMI only, no SUD, 3 = SUD and AMI, 4 = Neither SUD or AMI	Categorical (nominal)	NA	NA - exploratory purposes
Recoded-Age category	1 = 12-17 Years Old, 2 = 18-25 Years Old, 3 = 26-34 Years Old, 4 = 35 or Older	Categorical (ordinal)	NA	NA - subpop var

II. Pseudo Maximum Likelihood Logistic Regression

a) Context

Pseudo maximum likelihood (PML) is a method used in logistic regression for complex survey data, where standard maximum likelihood estimation (MLE) doesn't work. It extends MLE by incorporating survey weights into the log-likelihood function, adjusting it to account for the sampling design and unequal selection probabilities. This ensures that the resulting parameter estimates represent the target population. PML logistic regression inherently accounts for survey weights by including them directly in the adjusted log-likelihood, making it suitable for survey data analysis.

b) Equation

$$\ell_w(\boldsymbol{\beta}) \;=\; \sum_{i=1}^{n} w_i \Big[\, y_i \log p_i \;+\; (1 - y_i) \log(1 - p_i) \,\Big].$$

The equation above is the pseudo log-likelihood function for logistic regression. What it intuitively does is figure out the beta parameter (B) that maximizes the log-likelihood function to where the predicted probabilities (Pi) are as close as possible to the outcomes (yi), while accounting for survey weights (wi). These beta parameters, because they are selected based on survey weights, reflect population-level relationships between predictors and the binary outcome. These beta parameters (will be more than one in our case) are used in interpreting the output of our logistic regression model.

III. Restricted Cubic Splines

a) Context

Restricted cubic splines (RCS) is a method commonly used in health-related research such as epidemiology to model the non-linear relationship between exposure and outcome. It can integrate directly with the pseudo maximum likelihood logistic regression function while relaxing the linearity assumption of PML logistic regression. Largely, the sum of all RCS functions gives us a locally adaptive slope that can bend between knots. Its "restriction" comes from preventing the tails (boundary knots) from having extreme values by assuming linearity at

the tails. This restriction makes RCS especially useful for health-related research, making the odds ratios visualization derived from RCS more easily interpretable.

b) Equation

$$f(X_1) = \beta_1 X + \beta_2 (X - \kappa_1)^3 + \beta_3 (X - \kappa_2)^3 + \beta_4 (X - \kappa_3)^3 + \beta_5 (X - \kappa_4)^3$$

The equation above is the restricted cubic splines function used in our restricted cubic splines model. What this function intuitively does is join piecewise cubic polynomials together at 5 knots to model the non-linear relationship between past-year marijuana use and past-year suicidal ideation. The k parameters in this function are the knots used to build the restricted cubic splines model, and the visualization between past-year marijuana use and past-year suicidal ideation. Meanwhile, the beta parameters are determined with maximum likelihood estimation, choosing the beta parameters that lead to the best model fit.

IV. Model Formulations

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_{11}(X_{11} \times X_{12}) + \beta_{12}(X_{11} \times X_{13}) + \cdots + \beta_{21} X_{21}$$

Equation 1. Adjusted Logistic Regression Model Formulation (including interactions). The model contains a total of 22 variables (including interactions and dummy variables), with our main predictor, our main response variable, and 14 covariates.

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + f(X_1)$$

Equation 2. Restricted Cubic Splines model between our main predictor and response variable. F(x) is the restricted cubic splines function.

V. Supplemental Figures

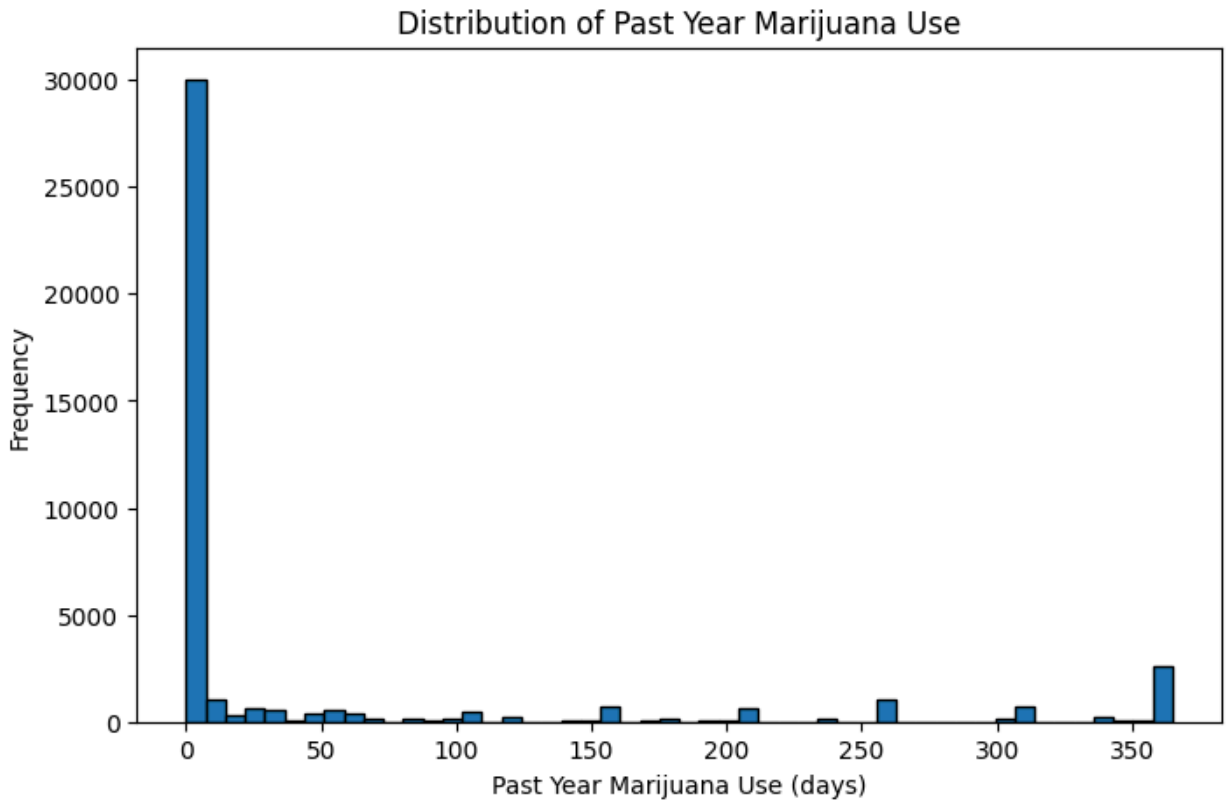

Figure 4. This figure shows that roughly 75% of the young adults in our data did not report using marijuana in the past year, revealing a significant imbalance in our data. Moreover, the ~25% of young adults that did use marijuana in the past year appear to have moderate to high use on average.

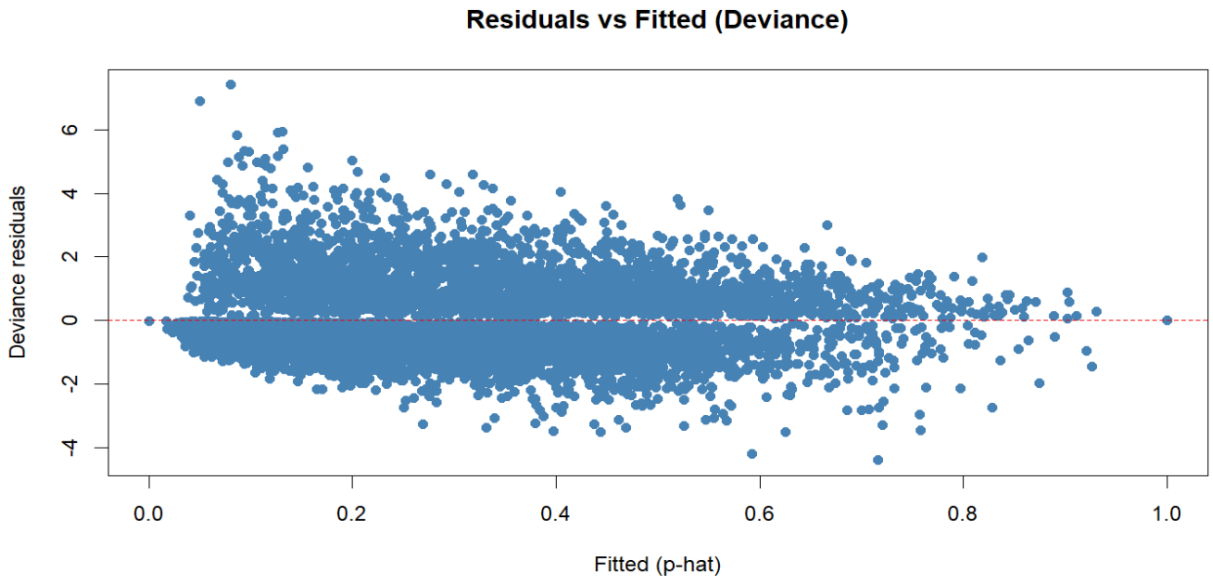

Figure 5. This figure shows the deviated residuals (integrate directly with the log-likelihood function) versus fitted values plot for our adjusted logistic regression model (see Table 2) between categorical marijuana use and suicidal ideation. With the exception of some outliers for fitted values near 0, this plot demonstrates a random scatter of deviated residuals around 0, so our model fit appears to be good overall.