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 limited.

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 increased odds of SI as marijuana use increases (especially at light use) with a notable plateau in odds at moderate use.

Conclusion: Marijuana use increases the odds of SI with a non-linear 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 (SAMHSA, 2023). Moreover, more than 10% of young adults experienced suicidal ideation (SI) in the past year (SAMHSA, 2023). 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 (SAMHSA, 2023). Furthermore, marijuana is the most commonly used illicit substance (> 61 million users in 2023) (SAMHSA, 2023). Thus, problems with marijuana use and SI among young adults merit further attention.

Building on this problem, the association between illicit substance use and mental health has been explored in prior research. Filbey *et al.* (2014), utilizing cross-sectional data on marijuana use and brain MRI data, found with parametric regression models that chronic marijuana users have lower OFC (orbitofrontal cortex) gray matter volume than non-using controls. Inversely, Cheetham *et al.* (2012), given longitudinal data, found also with parameter regression that smaller OFC volumes at age 12 years predicted initiation of cannabis use by age 16 years. 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.

In addition to this research, a random forest model was developed to evaluate risk of SI among adults who abuse marijuana (Choi *et al.*, 2021). Using the 2019 National Survey on Drug Use and Health (NSDUH), the study found the random forest model to have greater than 90% accuracy. Finally, another research article, using data from the 2015-2019 NSDUH on young adults, found illicit LSD use to be significantly positively associated with suicidal ideation (Han *et al.*, 2022). While all of these studies examine the relationship between illicit substance use and mental health, research specifically on the marijuana use-SI association from 2023 onward is limited. Moreover, little research has captured the non-linear relationship between marijuana use and SI.

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 study 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, these findings promote education for young adults on the connection between marijuana use and SI.

Addressing the aforementioned gaps, this research examines 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? Based on prior research, we hypothesize that past-year marijuana use is strongly, positively associated with past-year SI, even after adjusting for confounding factors. In addition, we hypothesize nervousness frequency in the past year to be an effect modifier of this association. Moreover, since past research has found an M-shaped relationship between alcohol use and depression (Qi *et al.*, 2024), we hypothesize that a non-linear association between marijuana use and SI exists. All that said, in the methodology section of this paper, we explain the procedures used from the data collection to the analytical methods to enable reproducibility. In the results section, we present raw results and some interpretation. Finally, in the discussion section of this paper, we discuss the findings, an explanation of the findings, limitations, public health implications, and next steps in their respective order.

Methodology

1. Research Overview

This research focused 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 was associated with illicit drug use and mental health problems among college students (Grant *et al.*, 2019). While several

confounding variables were controlled for in our analysis, this study did not establish causality because this was a cross-sectional study that cannot show temporal relationships.

2. Dataset Description

This study used secondary, public use data (confidential information eliminated) from the 2021-2023 National Survey on Drug Use and Health (NSDUH). The NSDUH data was the leading source of population-based statistics on behavioral health information like tobacco use, alcohol use, drug use, and mental health. The dataset was cross-sectional, and its unit of observation was 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, which revealed 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 were 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 due to their ability to produce more stable estimates. The imputation method of predictive mean neighborhood was used for the imputed variables, which was designed to minimize bias by identifying similar respondents to those with missing data, thereby preserving the data's original distribution and variable relationships (NSDUH Imputation Methods Evaluation, 2017). Subsequently, all the categorical variables were decoded (e.g. changing 0 to No and 1 to Yes) to ensure interpretable analysis in the exploratory stage. *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.

Before the logistic regression analysis, another dataset was used with all categorical variables recoded to enable generation of a correlation matrix as well as variance inflation factors. During the logistic regression analysis, in addition to a standardized yearly marijuana use variable, a

categorical variable for yearly marijuana use with equal length categories (e.g. 0-90, 90-180…) was created to identify non-linearities.

Regarding missing as well as imputed value handling, substance use responses coded as never or none (91, 93, 991, 993) were recoded to 0 for analysis to prevent those codes from skewing results (e.g. 991 or 993 raises the average yearly marijuana use in our data to an inaccurate value). Finally, values of -9 or 99 among covariates were replaced with NA.

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 Visual Studio Code software (Microsoft, 2025) and R software (R Core Term, 2024). 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).

Pseudo maximum likelihood (PML) logistic regression models are used for complex survey data, where standard maximum likelihood estimation (MLE) doesn't work (see Appendix II). They extend MLE by incorporating survey weights into the log-likelihood function, adjusting them to account for the sampling design and unequal selection probabilities. This ensures that the resulting parameter estimates represent the target population. Overall, these features make them suitable for survey data analysis (Dey *et al.*, 2025). 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).

PML 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, PML logistic regression analysis was stratified by the most significant confounder (how often felt nervous), which was determined by seeing the effect of removal of the confounder on the adjusted model results.

After PML 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 fit our data types of continuous predictor and binary response. Restricted cubic splines (RCS) is a method commonly used in health-related research that fits piecewise cubic polynomials together at knots (Desquilbet *et al.*, 2010). It can integrate directly with the PML logistic regression function while relaxing the linearity assumption. Largely, the sum of all RCS functions yields 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 public health research, making the odds ratios visualization derived from RCS more easily interpretable (Lopez-Ayala *et al.*, 2025).

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). For large sample sizes, 5 knots is the standard (Harrell, 2015), so these knots were incorporated at the 5th, 27.5th, 50th, 72.5th, and 95th percentile of past-year marijuana use. However, it's important to note that young adults with 0 past-year marijuana use (~30,000) were excluded from this

analysis due to values of 0 creating a huge imbalance in the data (see Appendix V). 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 (104 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%).

Concerning demographic characteristics, Table 1a-c showed that 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%).

Table 1a.

Young adults' demographic 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
Non-Hispanic White	52.4	56.6	51.8	
Hispanic	23.8	20.7	24.3	
Non-Hispanic Black/African American	13.5	11.8	13.7	
Non-Hispanic Asian	6.3	5.6	6.4	
Non-Hispanic more than one race	3.0	3.9	2.8	
Non-Hispanic Native American/Alaska Native	0.6	0.9	0.6	
Non-Hispanic Native Hawaiian/Other Pacific Islander	0.4	0.3	0.4	

[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

Table 1b.

Young adults' demographic characteristics according to the presence of suicidal ideation[a] (continued)

	Overall, n=41,873	**Suicidal Ideation, n=5303**	**No Suicidal Ideation, n=36570**	***P*-value[b]**
Highest Education Obtained, %				<.001
Some college/Associates Degree	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

[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

Table 1c.

Young adults' demographic characteristics according to the presence of suicidal ideation[a] (continued)

	Overall, n=41,873	**Suicidal Ideation, n=5303**	**No Suicidal Ideation, n=36570**	***P*-value[b]**
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	

[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

Regarding substance use patterns, Table 1d 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). Greater than one percent of young adults received inpatient substance use treatment in the past year (1.4%).

Table 1d.

Young adults' substance use patterns according to the presence of suicidal ideation[a]

	Overall, n=41,873	**Suicidal Ideation, n=5303**	**No Suicidal Ideation, n=36570**	***P*-value[b]**
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

[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

According to mental health patterns in Table 1e, 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 1e.

Young adults' mental health patterns according to the presence of suicidal ideation[a]

	Overall, n=41,873	**Suicidal Ideation, n=5303**	**No Suicidal Ideation, n=36570**	***P*-value[b]**
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[c], %				<.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

[c] Everything feeling like an effort refers to that even simple tasks like taking a shower feel overwhelming due to amotivation

With the exception of income, Table 1a-e 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 potential covariates that were accounted for in the survey-weighted logistic regression analysis.

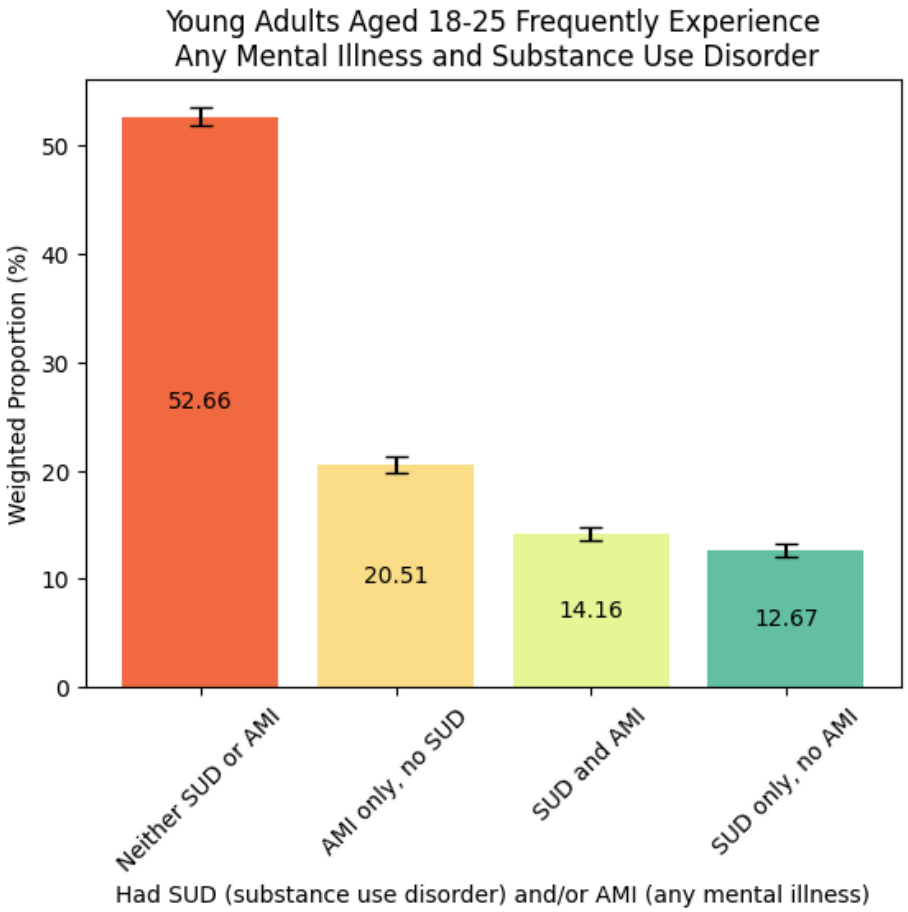

Figure 1. A significant percentage of young adults have a substance use disorder (SUD) and any mental illness (AMI) relative to other age groups in the US (Substance Abuse and Mental Health Services Administration, 2024). Estimates on SUD and AMI from the 2021-2023 NSDUH data were derived using tabulation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

According to Figure 1, young adults reported experiencing co-occurring substance use disorder (SUD) and any mental illness (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.

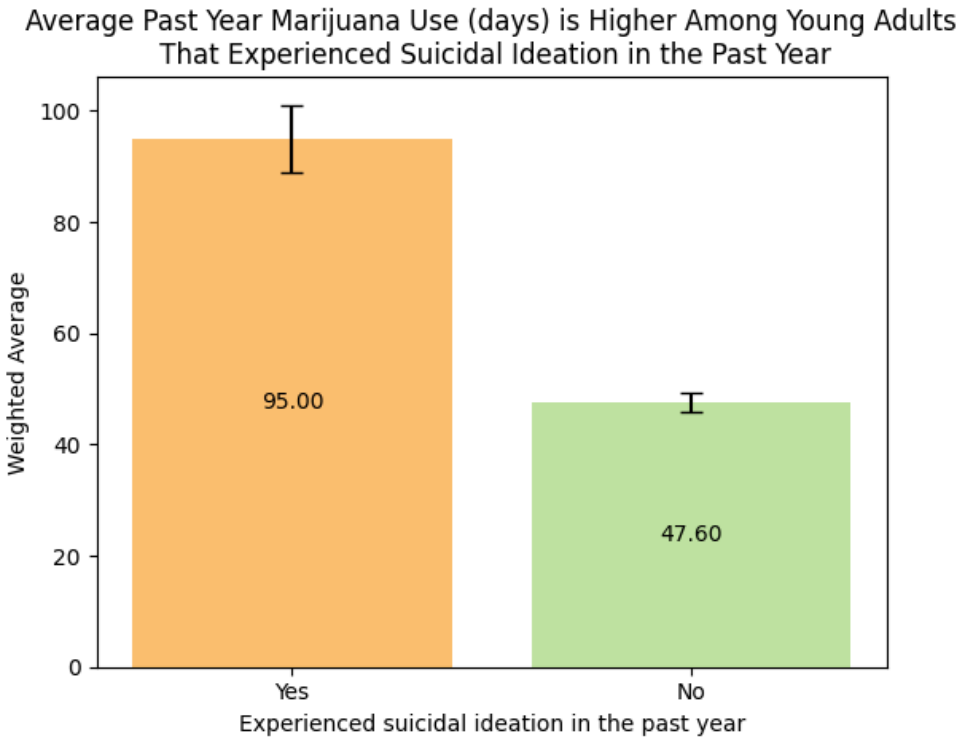

Figure 2. A weighted Mann-Whitney U test confirmed this significant association between past year marijuana use and suicidal ideation (p<.001). Weighted averages and error bars were derived from taylor linearization, which takes into account survey weights, strata, and primary sampling units of this 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, n=10073		**Felt Nervous between some of the time and none of the time, n=7796**	
	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). With that being said, nervousness did not appear to be an effect modifier of past-year SI, and this was confirmed by a design-adjusted Wald test (not shown), finding the interaction term added to the logistic regression model (Table 2) to be insignificant (p>0.05).

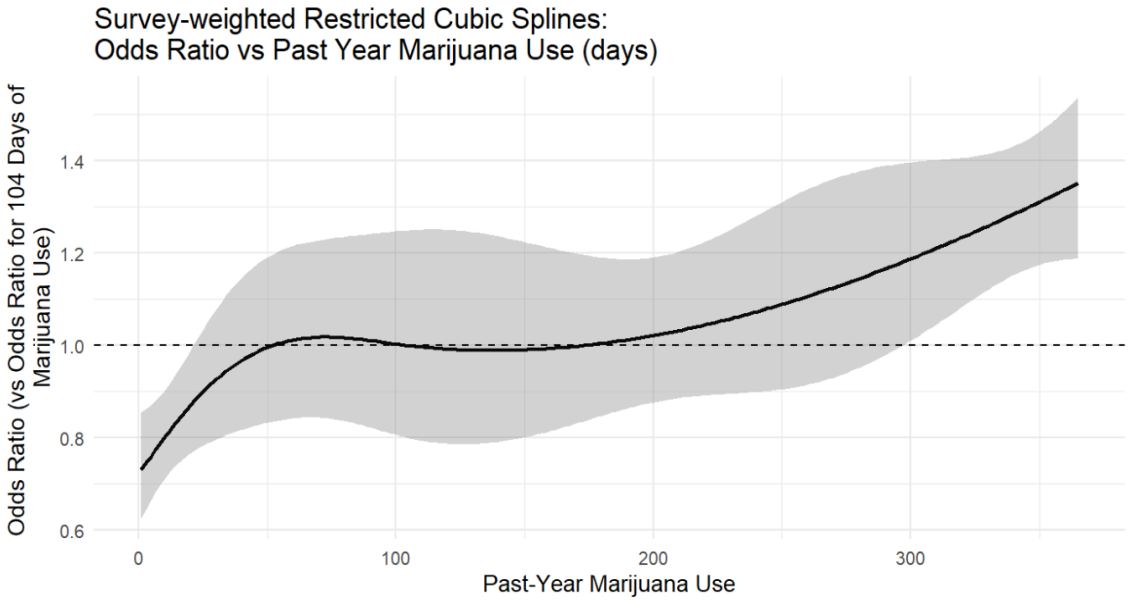

Figure 3. This visualization was extracted from an unadjusted restricted cubic splines model between past-year marijuana use and past-year suicidal ideation *among past-year marijuana users only.* Past-year marijuana use is associated with increased odds for suicidal ideation as we see a general increase in the odds ratio for suicidal ideation for each one unit increase in past-year marijuana use. An increase in odds of suicidal ideation is especially shown between 0 and 50 days before plateauing with a slight dip between 50 and 200 days. Subsequently, the odds slightly increase again between 200 and 365 days.

According to Figure 3, without adjusting for covariates, a non-linear association was found between marijuana use and SI. Overall, a rise in odds of SI for light marijuana users relative to those that used marijuana for 104 days (median value) was observed, followed by a plateau with a slight dip in odds for moderate users, and an increase in odds for heavy users. This non-linear 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 aged 18-25 revealed a positive association between marijuana use and SI, meaning that those who use marijuana more frequently are more likely to experience SI relative to infrequent users. 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 non-linear with an increase in odds of SI among light users, a plateau in addition to a slight dip in odds of SI at moderate use, and an increase in the odds of SI at heavy levels of marijuana use. With one exception of nervousness not being an effect modifier, these results align with our hypotheses.

2. Explanation

The overall positive association between marijuana use and SI among young adults aged 18-25 can be explained by aforementioned changes in OFC gray matter volume that occur from chronic marijuana use (Filbey *et al*., 2014). 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 (e.g. respondents of NSDUH survey reporting inaccurate personal information on nervousness level) and recall bias (e.g. survey respondents likely had trouble recalling how often they felt nervous in the past). 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. Finally, the non-linear association seen between marijuana use and SI was expected, and supports non-linear findings in prior research that found a rise in odds of depression at light alcohol use as well as heavy use (Qi *et al.*, 2024).

3. Limitations

In addition to self-report bias and recall bias, limitations of this study include that it cannot establish causality, not all relevant confounders were accounted for, data imbalances, misclassification bias, and overadjustment bias. 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). Fourth, by recoding our marijuana use predictor values of 991 or 993 to 0, we introduced a slight amount of misclassification bias in this study. In other words, no use of marijuana ever and no use of marijuana in the past year are two entirely different categories that we treated the same (as 0). Nonetheless, it must be re-emphasized that restricted cubic splines requires a continuous predictor, and recoding 991 or 993 marijuana use to any value other than 0 would skew the results. Finally, it's likely that the mental health covariates adjusted for in this study (nervousness, difficulty concentrating) were mediators and not confounders, making us potentially underestimate the true association.

4. Public Health Implications

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. These findings enable SAMHSA to provide actionable recommendations towards marijuana use and SI treatment. Moreover, clinicians and healthcare facilities are given more knowledge on how to treat SI in times of crisis. Young adults are informed on methods of marijuana use prevention and minimization.

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

Cheetham, A., Allen, N. B., Whittle, S., Simmons, J. G., Yücel, M., & Lubman, D. I. (2012). Orbitofrontal Volumes in Early Adolescence Predict Initiation of Cannabis Use: A 4-Year Longitudinal and Prospective Study. *Biological Psychiatry*, *71*(8), 684–692. https://doi.org/10.1016/j.biopsych.2011.10.029

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

Dey, D., Haque, M. S., Islam, M. M., Umme Iffat Aishi, Sajida Sultana Shammy, Mayen, A., Noor, A., & Uddin, M. J. (2025). The proper application of logistic regression model in complex survey data: a systematic review. *BMC Medical Research Methodology*, *25*(1). https://doi.org/10.1186/s12874-024-02454-5

Desquilbet, L., & Mariotti, F. (2010). Dose-response analyses using restricted cubic spline functions in public health research. *Statistics in Medicine*, n/a-n/a. https://doi.org/10.1002/sim.3841

EVALUATION OF IMPUTATION METHODS FOR THE NATIONAL SURVEY ON

DRUG USE AND HEALTH. (2017). https://www.samhsa.gov/data/sites/default/files/NSDUH-PMNImputationMethods2015.pdf

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

Harrell, F. E. (2015). Regression Modeling Strategies. In Springer Series in Statistics. Springer International Publishing. https://doi.org/10.1007/978-3-319-19425-7

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

Lopez-Ayala, P., Riley, R. D., Collins, G. S., & Zimmermann, T. (2025). Dealing with continuous variables and modelling non-linear associations in healthcare data: practical guide. BMJ, 390, e082440–e082440. https://doi.org/10.1136/bmj-2024-082440

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

Microsoft. (2025). *Visual Studio Code*. Visualstudio.com; Microsoft. https://code.visualstudio.com/

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/s10608-016-9782-0

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

R Core Team. (2024). *R: The R Project for Statistical Computing*. R-Project.org. https://www.r-project.org

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/2021-202

Appendix

I. Data Documentation

Variable Description	Values	Data Type (numeric, categorical, text, date)	% Missingness	Role (Predictor, Covariate, Response)
Adult Seriously Thought About Killing Self Past Year - Imputation Revised	. = Aged 12-17, 0 = No, 1 = Yes	Categorical (binary)	0%	Response
Marijuana Frequency Past Year - Imputation Revised	Range = 1 - 365, 991 = Never Used Marijuana, 993 = Did Not Use Marijuana Past Year	Numerical (discrete)	0%	Predictor
Sex at birth - Imputation Revised	1 = Male, 2 = Female	Categorical (binary)	0%	Covariate
Race/Hispanicity recode	1 = NonHisp White, 2 = NonHisp Black/Afr Am, 3 = NonHisp Native Am/AK Native, 4 = NonHisp Native	Categorical (nominal)	0%	Covariate

	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)	0%	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)	0%	Covariate
Private Health Insurance - Imputation Revised	1 = Yes, R does have private health insurance, 2 = No, R does not have private health insurance	Categorical (binary)	0%	Covariate

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)	0%	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)	0%	Covariate
Alcohol Frequency Past Year - Imputation Revised	Range = 1 - 365, 991 = Never Used Alcohol, 993 = Did Not Use Alcohol Past Year	Numerical (discrete)	0%	Covariate
Cig Frequency Past Month - Imputation Revised	Range = 1 - 30, 91 = Never Used Cigarettes, 93 = Did Not Use	Numerical (discrete)	0%	Covariate

	Cigarettes Past Month			
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)	33.4%	Covariate
Binge Alcohol Frequency Past Month - Imputation Revised	Range = 0 - 30, 91 = Never Used Alcohol, 93 = Did Not Use Alcohol Past Month	Numerical (discrete)	0%	Covariate
Recoded-Received Substance Use Treatment As An Inpatient - Past Year	0 = No, 1 = Yes, -9 = Data not collected for the year	Categorical (binary)	33.4%	Covariate
How Often Felt Nervous 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 =	Categorical (ordinal)	71.6%	Covariate

	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)	71.6%	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)	36.7%	Covariate
Final person-level sample weight	Len: 8 (e.g. 3276.46987)	Numeric (discrete)	0%	NA - survey weights
Variance stratum	Len: 5 (e.g. 40031)	Numeric (discrete)	0%	NA - survey strata
Variance primary sampling unit	Len: 2 (e.g. 2)	Numeric (discrete)	0%	NA - survey primary sampling units

Year in which data was collected	Len: 4 (2021-2023)	Numeric (discrete)	0%	NA - data collection year
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)	0%	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)	0%	NA - subpop var

---

II. Pseudo Maximum Likelihood Logistic Regression

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

Equation 1. Pseudo log-likelihood function.

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 (are more than one in our case) are used in interpreting the output of our logistic regression model.

III. Restricted Cubic Splines

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

Equation 2. Restricted cubic splines function.

The equation above is the restricted cubic splines function (5 knots) 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 an exposure (X) and outcome (f(X)). The k parameters in this function are the knots used to build the restricted cubic splines model, and the visualization between exposure and outcome. 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 3. 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 4. Restricted Cubic Splines model between our main predictor and response variable. F(x) is the restricted cubic splines function.

V. Supplemental Figures

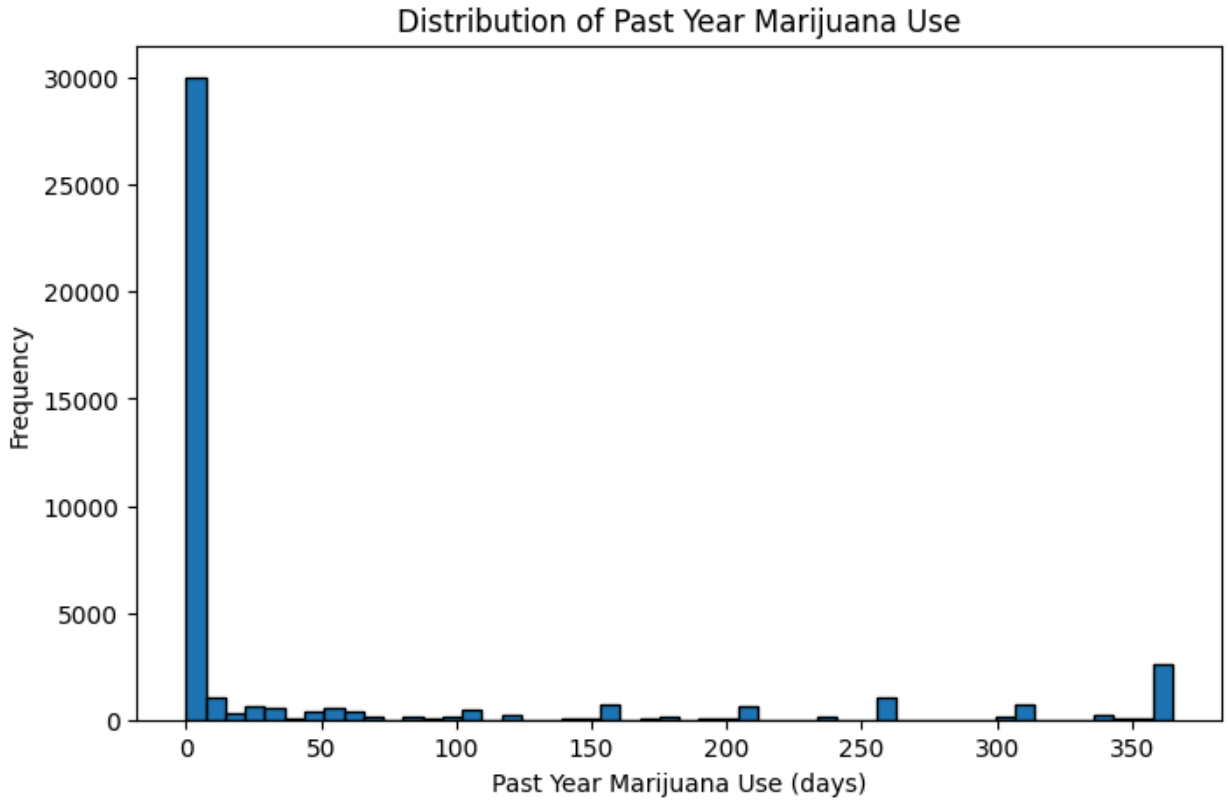

Figure 4. This figure shows that roughly 75% of the young adults in our data did not report using marijuana in the past year, revealing a significant imbalance in our data. Moreover, the ~25% of young adults that did use marijuana in the past year appear to have moderate to high use on average.

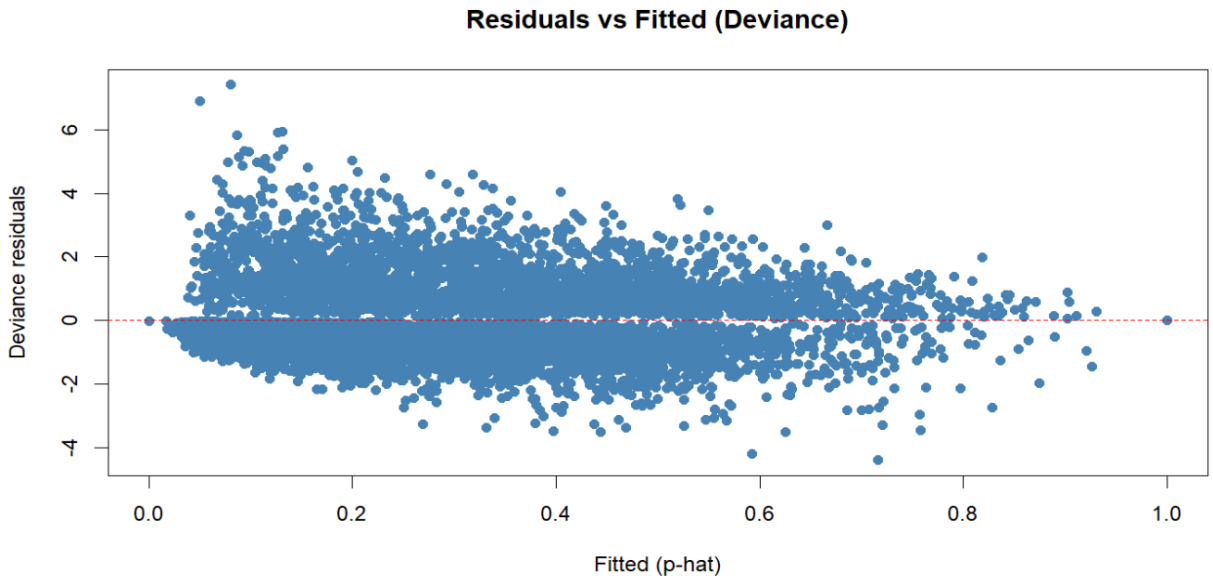

Figure 5. This figure shows the deviated residuals (integrate directly with the log-likelihood function) versus fitted values plot for our adjusted logistic regression model (see Table 2) between categorical marijuana use and suicidal ideation. With the exception of some outliers for fitted values near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