# *The Impact of Recreational Marijuana Legalization on Binge Drinking in the U.S.: Evidence from a Difference-in-Differences Approach*

Maggie Ching, Katerina Folkin, Avi Gu, Anusha Kakileti
*Minerva University*
*1145 Market St, San Francisco, CA 94103*

Submitted: Spring 2025

**Abstract**

This paper investigates the impact of recreational marijuana legalization on binge drinking prevalence in the United States, addressing ongoing debates about whether cannabis acts as a substitute or complement to alcohol consumption. Drawing on the policy variation created by staggered state-level legalization between 2012 and 2016, we implement a quasi-experimental Difference-in-Differences (DiD) approach to estimate the causal effect of legalization on binge drinking prevalence, defined using CDC criteria. The study controls for demographic factors including educational attainment, median age, and urbanization rates. The results indicate that there is no statistically and practical (Cohen's d at 0.29) significant effect of recreational marijuana legalization on adult binge drinking rates during the study period. These findings suggest that in the years following legalization, changes in cannabis policy do not substantially influence binge-drinking behaviors at the population level. This evidence offers important insights for policymakers, indicating that fears of increased alcohol misuse or hopes of reduced alcohol consumption tied to marijuana legalization may not be supported by short-term data. Further research is recommended to explore potential differences across demographic subgroups and to assess longer-term trends as more states adopt legalization measures.

## 1 Introduction

Does recreational marijuana legalization affect the prevalence of binge drinking in the United States? This question centers on how shifting cannabis laws might influence alcohol consumption. Binge drinking here follows the CDC's definition—four or more drinks for women or five or more for men on a single occasion—and is measured by how many of those episodes occurred in the prior 30 days. The analysis spans U.S. states that legalized recreational marijuana after the first state's legalization in 2012 and those that have not, allowing a comparison of binge drinking prevalence across different policy contexts.

There is currently a field of research exploring whether alcohol and marijuana act

as substitutes or complementary products (Gunn et al. 2022, Subbaraman 2016). Although a slight majority of work suggests they can be substitutes, there is substantial evidence indicating they can also reinforce one another (Sewell et al. 2009). This study examines the substitution hypothesis, which proposes that increased access to recreational marijuana may lead to a reduction in alcohol consumption, as individuals replace alcohol with cannabis in certain social and personal contexts. Understanding whether relaxed marijuana policies drive people to reduce or increase alcohol consumption can therefore shed light on the broader consequences of legalization, including any unintended policy tradeoffs.

From a policy standpoint, a decline in alcohol use linked to marijuana legalization could be seen as beneficial, given that marijuana is often considered less harmful than alcohol. However, if legalization correlates with higher rates of binge drinking, it raises a fresh set of concerns regarding substance use and how regulations should respond. Recognizing such outcomes is important so that policymakers weigh both cannabis and alcohol dimensions when deciding on legalization and post-legalization strategies.

Finally, alcohol misuse has substantial public and personal health impacts, including increased healthcare expenses, a higher incidence of accidents, and serious long-term health risks. If recreational marijuana legalization reduces binge drinking, it could offer social benefits by curbing alcohol-related harms. Conversely, if it increases binge drinking, it may diminish or overshadow any positive outcomes, prompting the need for additional interventions and regulatory measures to address the negative effects. Therefore, this question has important implications.

## 2 Literature Review

The relationship between marijuana legalization and alcohol consumption has been widely debated, with studies offering mixed conclusions on whether cannabis acts as a substitute or a complement to alcohol. This distinction has important implications for public health, as co-use can amplify risks, while substitution may reduce alcohol-related harm.

A slight majority of research supports the substitution hypothesis, suggesting that expanded access to marijuana can lead to reductions in alcohol use. Subbaraman (2016) finds that cannabis may displace alcohol in certain contexts. Gunn et al. (2022) show that marijuana use, particularly for stress relief or medical reasons, may replace alcohol consumption for some users.

Other studies point to complementarity, where legalization is linked to increases in both marijuana and alcohol use. Sewell et al. (2009) note that simultaneous use is especially common among young adults, increasing risks such as impaired driving. Gonçalves et al. (2023) show that the effect of legalization on binge drinking differs by age, suggesting the relationship depends on demographic characteristics.

Contextual factors such as urbanization and education levels also appear to shape how legalization influences substance use. Steigerwald et al. (2020) find that both public support for legalization and actual use behaviors vary by socioeconomic background, making it important to control for these confounders in analysis.

Guttmannova et al. (2016) describe how the evolving legal and cultural landscape in the U.S. creates complex dynamics around substance use that are difficult to interpret without careful methodological design.

This study builds on these insights by using a difference-in-differences framework to estimate the effect of recreational marijuana legalization on binge drinking prevalence among U.S. adults, accounting for policy variation across time and states.

## 3 Data

This study uses a panel dataset made up of annual observations for all 50 U.S. states from 2011 through 2016. The unit of analysis is the state-year. The dataset was merged from several sources using state and year as the linking keys. The treatment variable is a binary indicator that equals one if a state legalized recreational marijuana in a given year and zero otherwise. Legalization dates vary by state, starting with Colorado in 2012. These dates were verified using multiple public sources, including infographics, state government websites, and well-known policy trackers like US News and World Report, which were cross-checked for consistency (Hansen et al. 2023). The variable was coded to change from 0 to 1 in the year a state's law took effect.

The main outcome variable is the percentage of adults who report binge drinking, defined by the CDC as four or more drinks for women or five or more drinks for men on a single occasion within the past 30 days. This data comes from the Behavioral Risk Factor Surveillance System (BRFSS), an annual nationwide survey conducted by the CDC (Centers for Disease Control and Prevention 2025). BRFSS collects responses through phone interviews using random-digit dialing across landlines and cell phones. The survey is designed to get a representative sample from each state. While the method allows for broad coverage and consistency over time, it relies on self-reports and may be affected by common survey issues like selective participation, memory problems, or people misreporting their behavior.

To control for possible confounding factors, the analysis includes three demographic variables: the percentage of people over age 25 with a bachelor's degree, the median age of the state population, and the share of the population living in urban areas. Educational attainment and median age come from the U.S. Census Bureau's American Community Survey (ACS), which is updated each year (U.S. Census Bureau, U.S. Department of Commerce 2023*b*,*a*). Urbanization rates come from the 2020 Decennial Census, which is more complete but only conducted once every ten years (U.S. Census Bureau 2020). Because of this, the urbanization variable is held constant across the full period and should be interpreted as a general structural feature of each state.

Table 1 shows basic descriptive statistics for all the key variables in the study. The table includes the mean, standard deviation, minimum, and maximum values. These statistics were calculated in R and displayed using the stargazer package Hlavac 2022. The summary helps give a sense of the variation across states and over time, especially in binge drinking rates and education levels. The average

value of the legalization variable is 0.20, which reflects the fact that only a portion of states passed laws during the years analyzed.

Table 1. Summary Statistics

| Statistic | Min | Mean | St. Dev. | Max |
|---|---|---|---|---|
| Binge_Drinking_Prevalence | 9.60 | 16.71 | 3.03 | 27.20 |
| Legalized | 0 | 0.20 | 0.40 | 1 |
| Bachelors_Rate | 18.50 | 31.80 | 6.79 | 65.90 |
| Median_Age | 29.60 | 38.42 | 2.41 | 45.10 |
| Urbanization_Rate | 0.31 | 0.72 | 0.15 | 1.00 |

All data cleaning and analysis were done in R, and the code is available at https://github.com/akakileti/ss154-final. The cleaned dataset can be downloaded from this Google Drive link. The cleaned dataset and variable codebook are also included. While the sources used are reliable and the dataset is structured to support causal inference, there are still some challenges. The binge drinking data is self-reported and may not perfectly reflect actual drinking behavior. The control variables are pulled from different surveys that follow different timelines, so year-to-year changes aren't always captured precisely. The legalization variable only marks whether a law was passed; it doesn't reflect differences in access or enforcement across states. These issues are addressed further in the Limitations section and tested in robustness checks.

## 4 Methodology

Estimating the causal effect of recreational marijuana legalization on adult binge drinking using a simple regression model would likely yield biased and unreliable results. This is due to confounding bias, where external variables influence both the likelihood of treatment (legalization) and the outcome (binge drinking prevalence). For example, states with younger populations tend to be more socially progressive, making them more likely to adopt marijuana legalization policies. At the same time, younger demographics also report higher rates of binge drinking. Similarly, states with higher education levels may be more supportive of legalization and simultaneously experience greater alcohol use among college-educated adults. Another critical factor is urbanization: urban states are more likely to legalize marijuana and have greater access to both cannabis and alcohol due to denser infrastructure, nightlife, and cultural attitudes. If these confounders are not properly controlled for, any estimated effect of legalization on binge drinking may reflect these underlying demographic or structural factors, not the causal effect of policy itself.

To overcome these limitations, this study applies a quasi-experimental approach

using the Difference-in-Differences (DiD) method, specifically the estimator proposed by Callaway & Sant'Anna (2021), which is designed for settings with staggered treatment adoption. This method compares the change in binge drinking in states that legalized marijuana (treatment group) to the change in states that did not legalize (control group) over the same time period. Crucially, it isolates the within-state change before and after legalization while accounting for time-specific shocks (like national economic events or federal campaigns) and fixed state-level characteristics (such as political culture or baseline health behaviors). Because recreational marijuana legalization occurred at different times across states between 2012 and 2023, the DiD framework is ideal—it allows the estimation of group-time average treatment effects that reflect variation in policy implementation timing. The model includes controls for three pre-identified confounders: median age, education level (percent with a bachelor's degree or higher), and urbanization rate.

This method is grounded in several key assumptions. First, the parallel trends assumption requires that, in the absence of legalization, binge drinking would have followed similar trends in both treated and control states. Second, the no anticipation assumption implies that individuals did not alter their drinking behavior in expectation of impending legalization. Third, the no spillover assumption assumes that marijuana legalization in one state does not affect alcohol use in neighboring control states—an assumption that could be violated if people cross state lines to access cannabis or if cultural shifts diffuse across borders. Lastly, the method assumes that all time-varying confounders are either accounted for or irrelevant to both treatment and outcome.

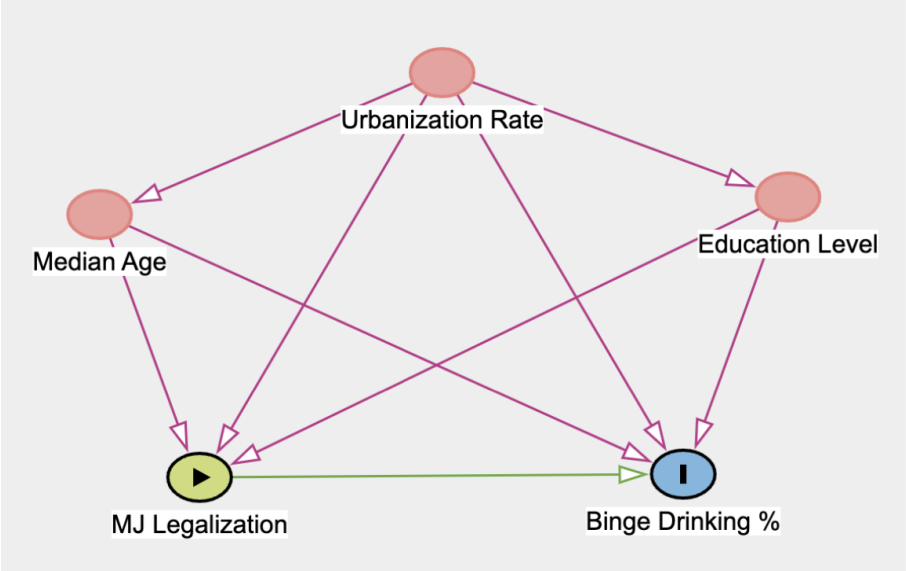


Fig. 1. DAG diagram showing the relationships between the treatment (MJ Legalization), the outcome (Binge Drinking ), and confounding variables

A causal diagram (DAG) (Figure 1) for this study places marijuana legalization as the treatment, with binge drinking prevalence as the outcome. From the backdoor

paths above, it is immediately clear that the minimal set would require all three variables, Median Age, Urbanization Rate, and Education Level, to be controlled for. This is because even if we controlled for any subset of two or one, at least one of the other variables would have a direct backdoor path between the treatment and outcome that would remain unblocked. Additionally, none of these three variables have two inward arrows pointing toward them in our causal diagram, so it is clear that none could be colliders. Thus, the full set of all three variables is minimal.

## 5 Analysis

The analysis revealed that marijuana legalization is associated with a small, statistically insignificant increase in binge drinking prevalence among adults. Group-time estimates of the treatment effect varied by cohort (i.e., states grouped by legalization year) and post-treatment year. However, most of these estimates were not statistically significant, as the confidence intervals typically included zero, and only a few years showed significance at conventional levels.

Table 2. Overall ATT Estimate

| Label | Estimate | Lower | Upper |
|---|---|---|---|
| Overall ATT | 0.879 | -0.920 | 2.678 |

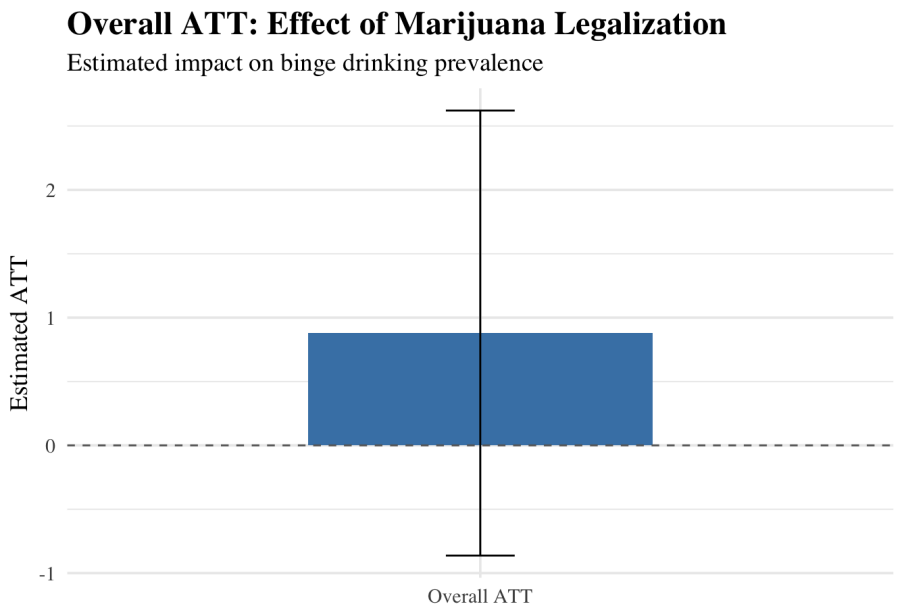


Fig. 2. Overall ATT Estimate with Confidence Intervals

The overall average treatment effect on the treated (ATT) was estimated to be

approximately 0.88 percentage points (see Table 2). This means that, on average, states that legalized marijuana experienced a 0.88 percentage point higher binge drinking rate in the years following legalization compared to states that did not legalize. However, this estimate came with a standard error of 0.86 and a 95 percent confidence interval ranging from -0.81 to 2.57, indicating that the result was not statistically significant. The effect size is also small in practical terms: given that the standard deviation of binge drinking prevalence is around 3 percentage points, this translates to a Cohen's d of about 0.29, which is conventionally considered a small effect.

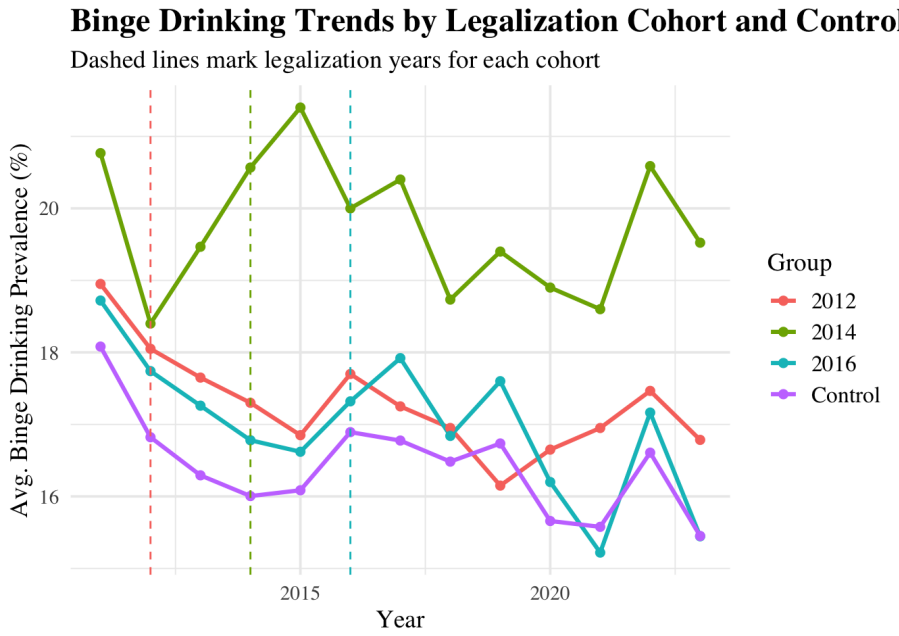


Fig. 3. Binge Drinking Trends by Legalization Cohort and Control Group

These results were supported by robustness checks. A visual and statistical inspection of pre-treatment trends showed no significant differences between treated and control groups prior to legalization, supporting the parallel trends assumption. This was confirmed by a p-value of approximately 0.75, suggesting no systematic divergence in trends before the policy was enacted. Additionally, results were compared with other analytic strategies. These consistent findings across methods reinforce the main result: no robust evidence of a causal effect between marijuana legalization and changes in binge drinking prevalence. Figure 2 visually presents the ATT estimate alongside its confidence intervals, confirming the non-significance of these effects.

An event study analysis, which plots treatment effects for each year relative to the legalization date, further confirmed the absence of systematic shifts in drinking behavior. Most treatment effects hovered near zero, with no clear upward or downward trajectory post-legalization. Figure 3 displays the binge drinking trends for treated and control cohorts over time, visually reinforcing these findings. This

visual evidence complements the statistical results, reinforcing the conclusion that the relationship between legalization and binge drinking is, at most, weak.

## 6 Limitations

This analysis is not without limitations. First, the BRFSS survey data is self-reported, which can introduce bias due to misreporting or non-response. Second, while three key confounders are included, variables like policy strength, marketing restrictions, or local alcohol tax laws are omitted, potentially leading to residual confounding.

Third, state-level aggregation masks within-state heterogeneity. Effects may differ by race, gender, or socioeconomic status, which cannot be detected with this design. Fourth, the binary treatment variable does not capture the nuances of marijuana policy implementation. States varied significantly in enforcement, regulation, and accessibility, which could influence behavior in more complex ways.

Finally, the findings are U.S.-specific. Cultural and regulatory contexts in other countries could produce different results, limiting the generalizability of this study.

## 7 Conclusion

This study finds no strong evidence that recreational marijuana legalization causes an increase in binge drinking prevalence among adults. Although the ATT is directionally positive (0.88 percentage points), it is not statistically significant. In practical terms, this effect is very small compared to the average binge drinking rate of 16.7 percent and the standard deviation of 3.03 percentage points. With a Cohen's d of 0.29, the effect size is classified as small.

Policymakers concerned that marijuana legalization might trigger increases in alcohol misuse can take cautious reassurance. At the same time, this study does not support the hypothesis that cannabis serves as a substitute for alcohol use either. Rather, both behaviors appear to evolve independently at the macro level.

Future research should explore individual-level or subgroup-specific effects, particularly among younger adults, and may consider structural equation modeling to investigate multiple behavioral channels simultaneously.

## Appendix A. Data, Code, and Codebook Access

The full dataset, variable codebook, and reproducible code are publicly available on GitHub: github.com/akakileti/ss154-final. The dataset includes binge drinking prevalence from the BRFSS, demographic data from the U.S. Census and ACS, and marijuana legalization years by state. The codebook explains all variable definitions, sources, and units. An R Markdown file was used for the analysis and visualizations, which can be downloaded via the GitHub repository.

**Links**
GitHub Repository: github.com/akakileti/ss154-final

Codebook (Google Sheets): Google Sheets Link
Dataset (Google Drive): Google Drive Link

**Data Sources**
Marijuana Legalization: US News Legalization Guide
Binge Drinking (BRFSS): CDC BRFSS Portal
Median Age (ACS): Census ACS Table S0101
Education Level (ACS): Census ACS Table S1501
Urbanization Rate (Census): Census Decennial P2 Table

## Appendix B. AI Use Statement

Grammarly was used to proofread the assignment. ChatGPT was used to assist with writing code, formatting in Latex, and making the text more concise.

## Appendix C. LO Appendix

**#ss154-CausalAssumptions:** In the study, we address the core statistical and causal assumptions that underpin the fixed-effects DiD model. While we use visual inspection to assess pre-treatment trends and ensure reasonable similarity between treated and control states, we also critically examine possible violations such as anticipation effects and spillovers. Our discussion of high standard errors and limited significance further reinforces the need for caution, recognizing that even if statistical assumptions are met, power and variance issues still constrain interpretation.

**#ss154-CausalStudyDesign:** The methodology section has clearly explained the justification for using the Difference-in-Differences model based on the structure of our panel dataset, the timing of marijuana legalization, and the availability of untreated control states, over simple regression or other models. Although we use a standard estimator rather than an advanced one, we account for heterogeneity in timing through the interaction specification and acknowledge, in the limitations, the constraints of this approach.

**#ss154-InterpretingResults:** Adjusting with feedback from previous feedback, we have interpreted the estimated treatment effect both statistically and practically by incorporating Cohen's d. The initial result—with a positive (0.136 percentage points) estimated treatment effect—is not statistically significant. The coefficient is also reported in percentage points rather than percentage change, maintaining clarity in interpreting the direction and magnitude of effects. Practically, the Cohen's d is calculated at 0.29, suggesting a small effect size relative to the sample's standard deviation. Our conclusion reflects these nuances and avoids overgeneralizing from the results.

**#ss154-CausalGraphs:** The construction and inclusion of the DAG diagram have clearly established the assumed directional causal pathways between variables. Taking feedback from previous assignments, we have supported them with relevant

literature. It also serves as primary evidence for the selection of our study design and covariate selection.

**#ss154-CausalQuestions:** Our causal question is defined as: *Does recreational marijuana legalization affect the prevalence of binge drinking in the United States?* The question has indicated both theoretical motivation and our choice of a fixed-effects DiD model, with consideration of geographic and temporal variation in policy adoption.

## Appendix D. HC Appendix

**#organization:** We applied #organization by structuring the content to ensure clarity and ease of understanding for the intended audience. We began by outlining the key ideas and arranging them logically, ensuring each section, such as the Introduction, Data Analysis, and Conclusion, was clearly numbered and easy to follow. Throughout the paper, we made sure to reference and number all tables, figures, and appendices, which were incorporated to support the main arguments and were directly linked to the text for easy navigation. This approach was designed to help the audience easily identify and absorb the most important points, while also providing a streamlined structure for the overall presentation.

**#professionalism:** We applied #professionalism by adhering to established academic guidelines and ensuring proper formatting and attribution. We followed the expected conventions for citing sources, including data and quotations, and ensured that all references were accurately attributed in APA format. Additionally, we linked to the relevant code and data via a GitHub repository, providing transparency. Throughout the paper, we maintained a formal tone and carefully proofread the document to eliminate errors, ensuring it met the professional standards expected in academic and research communication.

**#responsibility:** In this paper, we applied responsibility by ensuring that each team member followed through on their assigned tasks and communicated proactively about their progress. We adhered to the commitments made at the start of the project, ensuring that all sections were completed on time, including code, data, and the final document. When setbacks arose, we took responsibility for any delays and worked together to adapt and find solutions, such as reviewing each other's work and ensuring all requirements were met. Additionally, we communicated any challenges (including external time commitments and other final assignments) in advance, allowing the team to adjust plans and manage expectations effectively.

**#dataviz:** We ensured that all graphs and tables were clear, well-labeled, and self-explanatory. We created visualizations with clear axis labels and titles, such as the Directed Acyclic Graph (DAG) showing the relationships between marijuana legalization, binge drinking, and confounding variables. In addition to graphs, we recreated tables in the text editor rather than copying and pasting R tables, ensuring that each table had a clear caption and enough description for the reader to understand it without needing to refer to the main text. This approach helped present the data in a way that was visually accessible and easy to interpret.

**#hypothesisdevelopment:** We applied #hypothesisdevelopment by clearly outlining the expectations for the treatment effect based on existing theories and previous research. Specifically, we hypothesized that marijuana legalization would reduce binge drinking rates, supporting this with current literature and theoretical frameworks linking substance use and policy changes. Our hypothesis was grounded in observations from past studies that indicated shifts in substance use patterns following changes in marijuana policy. Additionally, we discussed how the treatment effect might manifest and referred to relevant theories, such as those related to behavioral changes in response to policy interventions, which helped shape our research design and analysis approach.

**#significance:** In this study, we have distinguished between statistical and practical significance. We started with the exploration of estimated average treatment effect (ATT) that yield at 0.88 percentage points. However, the result is not statistically significant. To access a practical significance, we calculate Cohen's d using the sample's standard deviation in binge drinking prevalence. The result then brings us to the conclusion that the observed difference is unlikely to be meaningful from a policy or behavioral standpoint.

**#descriptivestats:** In our study, we applied descriptive statistics by calculating the mean, standard deviation, minimum, and maximum values for key variables related to marijuana legalization and binge drinking. As shown in Table 1, we used these statistics to provide an overview of the data's variation across states and over time, particularly focusing on binge drinking rates and educational attainment. For instance, the average binge drinking prevalence was 16.71%, with a standard deviation of 3.03, indicating some variation in binge drinking levels across states, with values ranging from a minimum of 9.60% to a maximum of 27.20%. The legalization variable had a mean of 0.20, showing that only a portion of states legalized marijuana during the analyzed years, with values between 0 (no legalization) and 1 (full legalization). These descriptive statistics, calculated using R's stargazer package (Hlavac 2022), set the stage for understanding the broader patterns in our data, enabling us to interpret the effects of marijuana legalization on binge drinking behavior.

**#complexcausality:** From the start, we have identified and addressed confounding bias through a DAG diagram (Figure 1); a directed acyclic graph have effectively defined the structure between our treatment—recreational marijuana legalization, and our outcome, binge drinking prevalence with clear representation of directional relationships between variables. The 3 key backdoor paths, supported by literature, are then included as covariates in our fixed effects regression. This is to ensures that omitted variables bias is minimized by blocking all non-causal paths.

## References

Callaway, B. & Sant'Anna, P. H. (2021), 'Difference-in-differences with multiple time periods', *Journal of Econometrics* **225**(2).

Centers for Disease Control and Prevention (2025), 'Behavioral risk factor surveillance system (brfss) prevalence data (2011 to present)', Data.cdc.gov. Available at https://data.cdc.gov/Behavioral-Risk-Factors/Behavioral-Risk-Factor-Surveillance-System-BRFSS-P/dttw-5yxu/about_data.

Gonçalves, P. D., Bruzelius, E., Levy, N. S., Segura, L., Livne, O., Gutkind, S., Boustead, A. E., Hasin, D. S., Mauro, P. M., Silver, D., Macinko, J. & Martins, S. S. (2023), 'Recreational cannabis legislation and binge drinking in u.s. adolescents and adults', *International Journal of Drug Policy* **118**, 104085.

Gunn, R. L., Aston, E. R. & Metrik, J. (2022), 'Patterns of cannabis and alcohol co-use: Substitution versus complementary effects', *Alcohol Research: Current Reviews* **40**(2).

Guttmannova, K., Lee, C. M., Kilmer, J. R., Fleming, C. B., Rhew, I. C., Kosterman, R. & Larimer, M. E. (2016), 'Impacts of changing marijuana policies on alcohol use in the united states', *Alcoholism, Clinical and Experimental Research* **40**(1), 33–46.

Hansen, C., Alas, H. & Davis Jr., E. (2023), 'Where is marijuana legal? a guide to marijuana legalization'. US News & World Report.
**URL:** *https://www.usnews.com/news/best-states/articles/where-is-marijuana-legal-a-guide-to-marijuana-legalization*

Hlavac, M. (2022), 'stargazer: Well-formatted regression and summary statistics tables', https://CRAN.R-project.org/package=stargazer. R package version 5.2.3.

Sewell, R. A., Poling, J. & Sofuoglu, M. (2009), 'The effect of cannabis compared with alcohol on driving', *The American Journal on Addictions* **18**(3), 185–193.

Steigerwald, S., Cohen, B. E., Vali, M., Hasin, D., Cerda, M. & Keyhani, S. (2020), 'Differences in opinions about marijuana use and prevalence of use by state legalization status', *Journal of Addiction Medicine* **14**(4), 337–344.

Subbaraman, M. S. (2016), 'Substitution and complementarity of alcohol and cannabis: A review of the literature', *Substance Use & Misuse* **51**(11), 1399–1414.

U.S. Census Bureau (2020), 'Urban and rural', Decennial Census, DEC 118th Congressional District Summary File, Table P2. Retrieved March 4, 2025, from https://data.census.gov/table/DECENNIALCD1182020.P2?q=urban+rate.

U.S. Census Bureau, U.S. Department of Commerce (2023*a*), 'Age and sex', American Community Survey, ACS 1-Year Estimates Subject Tables, Table S0101. Retrieved March 4, 2025, from https://data.census.gov/table/ACSST1Y2023.S0101?q=ACS.

U.S. Census Bureau, U.S. Department of Commerce (2023*b*), 'Educational attainment', American Community Survey, ACS 1-Year Estimates Subject Tables, Table S1501. Retrieved March 4, 2025, from https://data.census.gov/table/ACSST1Y2023.S1501?q=education+acs.