# The Effect of Alcohol Consumption on Hard Drug Use

National Longitudinal Survey of Youth 1997

Saksham Arora, Foster Burnley, Griffin Kozlow, Sophie Kwon, and Anna Martin

Department of Economics, Dartmouth College

ECON 20: Econometrics

Professor Eric Chyn

March 12, 2021

# Abstract

This paper explores the effects of alcohol consumption on the use of hard drugs, especially in relation to the increase in access to alcohol at 21 years of age, the minimum legal drinking age (MLDA) in the US. To explore this relationship, we created a new data set from the National Longitudinal Study of Youth 1997 (NLSY97), limited to the years 1998-2008, which merges variables related to alcohol consumption and hard drug use. We limited the data to respondents who provided information about their alcohol consumption and hard drug use, meaning substances *other than* marijuana, such as crack, cocaine, and heroin. Initially, we ran OLS and fixed effects regressions to examine the effects of alcohol consumption on hard drug use, both of which were dummy variables equal to 1 if the respondent had consumed alcohol or used hard drugs since their last interview, the year prior. Some of these models provided statistically significant results, but we felt we didn't have the full story. Ultimately, by using a regression discontinuity model, we found that alcohol consumption sharply increased at 21 years of age, and hard drug use generally decreased around this same threshold. However, we could not conclude that there exists a discontinuity in hard drug use at 21 years of age, so we cannot definitively claim that alcohol and hard drugs are used as substitutes. All of our estimates are robust to various specifications and clustered by individuals, which will be expanded upon in our Data section.

# Introduction

Many studies across varied academic disciplines explore the consequences of alcohol consumption and hard drug use on individual health, work productivity, and interpersonal relationships. These studies also identify the particular consequences of underage drinking and hard drug use while the brain is still developing. One area of study that receives less

attention is the relationship *between* alcohol consumption and the use of hard drugs, particularly as that relationship relates to the minimum legal drinking age (MLDA) and how the relationship between alcohol and hard drug changes when one turns 21. How does alcohol consumption impact individuals' use of hard drugs? Are they complements for one another, or are they more commonly treated as substitutes? How does this relationship change for individuals as they are granted legal access to alcohol?

Our goal in this project is to identify the effects of alcohol consumption on the use of hard drugs. In order to explore this relationship, we generated a new panel dataset from the National Longitudinal Study of Youth 1997 (NLSY97). Our dataset includes variables related to alcohol consumption, hard drug use, and key demographic variables such as age, gender, and race, which we will further elaborate on in our Data section. Using these data, we leveraged three models to assess the relationship between alcohol consumption and hard drug use: ordinary least squares (OLS), fixed effects, and regression discontinuity (RD). As we will explain in our Methods section, each of these three models helps us to identify key components of the relationship between alcohol consumption and hard drug use that is of interest to us as a topic for research. In the Methods section, we will also compare our regression discontinuity model with an existing study which explores a similar research question to our own.

Through our three regression models, we found that alcohol and hard drugs tended to be treated as substitutes. We will elaborate on these findings in our Results section, but we found that as access to alcohol increases at the age of 21, a much higher proportion of respondents

reported alcohol consumption. In contrast, the proportion of respondents who reported using hard drugs had a decreasing slope around the threshold of 21.[1]

# Data

We examine the relationship between alcohol consumption and hard drug use using the National Longitudinal Survey of Youth 1997 (NLSY97) dataset, which collects panel data on a variety of factors on individual subjects each year from 1997 to 2015. The same individuals were interviewed each year, making this true panel data rather than pooled cross-sectional data. We limited this sample to the years 1998 to 2008. We didn't include 1997 because the survey and its questions changed quite a bit from 1997 to 1998, so it would have made our data much messier. We decided to cut off our data after 2008 because data was collected every *other* year starting then, which would have made our data much less informative and harder to work with. We are also limiting our data to respondents who included information about alcohol consumption and hard drug use. For the purposes of our study, hard drugs are defined as any illegal substance (such as crack, cocaine, heroine, etc.) *not including* marijuana.

In order to address our research question of the effects of alcohol consumption on hard drug use, we both merged existing variables from NLSY97 and generated new ones in our dataset. We created dummy variables for each of our controls—region, race, and gender—from categorical variables and merged in categorical variables corresponding to the highest year of education attained by each parent and the year the survey was conducted. The use of each of these controls is further explained in our Methods section. We also generated a dummy

[1] As a final introductory note, all figures, whether or not they are included in the text of our paper, are included in the Appendix section at the end.

variable for being over the age of 21 in order to address the changing relationship between alcohol consumption and hard drug use as a result of having legal access to alcohol.

Next, we generated dummy variables corresponding to a yes or no question on the survey asking if the respondent had ever consumed alcohol or used hard drugs. The dummy equals 1 if they responded "yes" and 0 otherwise. These variables are used in our OLS and fixed effects regression equations and experienced substantial nonresponse, which is explained in our Sample Selection subsection below and further elaborated upon in our Discussion and Conclusions section.

Finally, for our regression discontinuity model, we generated dummy variables corresponding to survey questions asking whether the participant had consumed alcohol or used hard drugs since their last interview, the year prior. These variables did not have the same nonresponse issue, which is also expanded upon in the Sample Selection subsection below. Lastly, we were able to interact our dummy variable for being over the age of 21 with our four age variables (*age*, *age2*, *age3*, and *age4*) to capture distinct functions for each side of the threshold in our RD regression, outlined in our Methods section.

## Sample Selection

Throughout the various models we created in this study, we used two unique samples. One of these samples was used for our OLS and fixed effects models, which will be discussed in greater detail in the Methods section of our paper. The second sample is larger, which allowed us to create a regression discontinuity model. As a result of these two separate samples, this section is split into two parts, each of which will cover one of these samples and the rationale behind our decisions.

## Sample 1: OLS and Fixed Effects Models

Our main challenge throughout the sample selection process was the fact that our data was hindered by a significant amount of nonresponse. Due to the illegality of hard drugs (no matter one's age) and alcohol (before one turns 21), surveyors were unable to get information on alcohol consumption and hard drug use from all participants. In order to address this issue, we limited our sample only to respondents who provided information on alcohol consumption *before* they turned 21. This brought our sample from 98,824 observations (8,894 subjects being interviewed each of 11 years) to just 4,381 observations and 1,961 subjects. The Table of Means for the entire dataset (98,824 observations) can be seen in *Figure 1* of the Appendix section at the end of our paper. Included below is the Table of Means which illustrates key demographic data for our smaller sample. We will compare this sample to the one used for our RD model in the subsection relating to regression discontinuity.

Figure 2: OLS and Fixed Effects Table of Means

| OLS and Fixed Effects | | | |
|---|---|---|---|
| **VARIABLE** | **Overall** | **(1)** | **(2)** |
| white | 0.5062 | 0.7364 | 0.6095 |
| | (0.5000) | (0.4407) | (0.4879) |
| black | 0.2345 | 0.0742 | 0.1933 |
| | (0.4237) | (0.2621) | (0.3949) |
| hispanic | 0.2116 | 0.2109 | 0.2034 |
| | (0.4084) | (0.4080) | (0.4025) |
| female | 0.4881 | 0.4542 | 0.4919 |
| | (0.4999) | (0.4980) | (0.4999) |
| dad_hgc_pos | 12.5643 | 12.9500 | 12.8162 |
| | (3.2117) | (3.2342) | (3.1609) |
| mom_hgc_pos | 12.4381 | 13.0017 | 12.7435 |
| | (2.9133) | (2.8603) | (2.8851) |
| north_east | 0.1424 | 0.1753 | 0.1707 |
| | (0.3495) | (0.3803) | (0.3763) |
| north_central | 0.1895 | 0.2043 | 0.2325 |
| | (0.3919) | (0.4032) | (0.4224) |
| south | 0.3359 | 0.3118 | 0.3531 |
| | (0.4723) | (0.4633) | (0.4779) |
| west | 0.1930 | 0.3040 | 0.2331 |
| | (0.3946) | (0.4601) | (0.4228) |
| Observations | 98824 | 4381 | 55562 |

Standard deviations in parenthesis
(1) - Universe : Respondents not lost due to non-response for alcohol participation (alc_participation_month) and drug participation (drugs_dli)
(2) - Universe : Respondents not lost due to non-response for alcohol participation (alc_participation_month)

One concern with this method of sample selection is that it may produce some omitted variable bias, as non-responders could have unobservable traits in common that we are not currently controlling for in our regression. For example, one could imagine that those who don't respond to questions about alcohol consumption and drug use are using alcohol and drugs at a disproportionate rate. This isn't an unsolvable issue, but it *does* mean that we have to be careful about how we interpret our data, which we will address in our Discussion and Conclusions section. When we compared people who responded to both questions with those who didn't, we found that there remained a statistically significant difference in the gender breakdown of these two groups. When we looked at race, however, we found varying results, which are explained in the next section when we compare the two samples. The proportion of each race that chose not to respond to questions relating to alcohol and drug use can be seen in the table below.

**Table 1: Proportion of Subjects Who Did Not Respond >= 1 Time(s) by Race**

| | Alcohol Use | Hard Drug Use |
|---|---|---|
| White | 0.323 | 0.931 |
| Black | 0.537 | 0.982 |
| Hispanic | 0.459 | 0.952 |
| Other | 0.412 | 0.952 |

While we don't have a way to know exactly why discrepancies in reporting alcohol and drug use occur along racial lines, one reason could be that minorities are less likely to report illegal substance use due to historical discrimination in drug convictions. This intuition is supported by a study conducted by Michael Fendrich and Timothy P. Johnson studying the role of race and ethnicity in the validity of self-reported drug use.[2] This study found that African Americans were statistically significantly more likely to lie about marijuana and cocaine use

[2] Fendrich, Michael, and Timothy P. Johnson. "Race/Ethnicity Differences in the Validity of Self-Reported Drug Use: Results from a Household Survey." *Journal of Urban Health : Bulletin of the New York Academy of Medicine* 82, no. Suppl 3 (September 2005): iii67–81. https://doi.org/10.1093/jurban/jti065.

than White participants, and, when asked for a reason for their lack of response, they provided two main answers: social desirability and fear of discrimination.[3] This evidence supports our hypothesis that non-White subjects in this study may have been less likely to answer such questions due to social factors outside of their control.

To understand why we restricted our sample in this way, it will be helpful to see empirical evidence of the nonresponse problem. Included in *Table 2* are the percentages of total subjects who responded to various questions consistently throughout the entire 11-year time span.

**Table 2: Percent of Respondents that Responded Consistently Over Time**

| Category of Observation | Percent of respondents who responded consistently throughout the entire time period |
|---|---|
| Alcohol Observation | 0.562 |
| Drug Observation | 0.048 |
| Alcohol and Drugs Observation | 0.044 |

Despite these host of challenges, we still maintain that our chosen sample will allow us to find reliable results because our sample is sufficiently large—much larger than 30, the baseline sample size to find reliable and significant results.

## Sample 2: Regression Discontinuity Model

For the previous models, we were limited to a sample size of only 4,381 observations due to lack of response to questions about alcohol consumption and hard drug use. With the RD model, however, we were able to expand our sample. For this model, we looked at data asking participants whether or not they had used drugs or alcohol since their last interview. Likely due to the binary nature of such a question, more participants gave a definite answer, which allowed us to expand our sample to 84,478 observations (from 8,863 subjects), all of which we were able to use in our RD model.

[3] Ibid, pp. 67-81.

Included below is our Table of Means for the expanded sample we use in our RD model. As you can observe in columns (1) and (2) in *Figure 2*, the samples have significantly different racial demographics (especially for column (2) in *Figure 2* in which the share of White respondents jumps significantly while the share of Black respondents drops significantly), due to the limited sample size used in the OLS and fixed effects regressions. Ultimately, this means that the conclusions we draw may only apply to the specific samples being used to run each regression, rather than applying across regression types. More specifically, we can't confidently use our conclusions from OLS and fixed effects models to inform our decisions regarding the regression discontinuity model, and vice versa.

Figure 3: Regression Discontinuity Table of Means

| | Regression Discontinuity | | | |
|---|---|---|---|---|
| **VARIABLE** | **Overall** | **(1)** | **(2)** | **(3)** |
| white | 0.5062 | 0.5466 | 0.5458 | 0.5460 |
| | (0.5000) | (0.4978) | (0.4979) | (0.4979) |
| black | 0.2345 | 0.2517 | 0.2515 | 0.2522 |
| | (0.4237) | (0.4340) | (0.4339) | (0.4343) |
| hispanic | 0.2116 | 0.2119 | 0.2121 | 0.2118 |
| | (0.4084) | (0.4086) | (0.4088) | (0.4086) |
| female | 0.4881 | 0.4983 | 0.4975 | 0.4982 |
| | (0.4999) | (0.5000) | (0.5000) | (0.5000) |
| dad_hgc_pos | 12.5643 | 12.5887 | 12.5825 | 12.5880 |
| | (3.2117) | (3.2082) | (3.2125) | (3.2075) |
| mom_hgc_pos | 12.4381 | 12.4743 | 12.4713 | 12.4732 |
| | (2.9133) | (2.9347) | (2.9355) | (2.9340) |
| north_east | 0.1424 | 0.1647 | 0.1648 | 0.1648 |
| | (0.3495) | (0.3709) | (0.3710) | (0.3710) |
| north_central | 0.1895 | 0.2196 | 0.2192 | 0.2195 |
| | (0.3919) | (0.4139) | (0.4137) | (0.4139) |
| south | 0.3359 | 0.3870 | 0.3871 | 0.3871 |
| | (0.4723) | (0.4871) | (0.4871) | (0.4871) |
| west | 0.1930 | 0.2237 | 0.2237 | 0.2235 |
| | (0.3946) | (0.4167) | (0.4167) | (0.4166) |
| Observations | 98824 | 83742 | 84478 | 83984 |

Standard deviations in parenthesis
(1) - Universe : Respondents not lost due to non-response for alcohol frequency (alc_freq) and drug frequency (drugs_freq)
(2) - Universe : Respondents not lost due to non-response for alcohol frequency (alc_freq)
(3) - Universe : Respondents not lost due to non-response for drug frequency (drugs_freq)

# Methods

In order to explore the relationship between alcohol consumption and hard drug use, we have decided to use three separate regression methods: OLS, fixed effects, and regression discontinuity. Each of these regressions is outlined below, with regression equations inserted directly below the corresponding paragraphs. Before we can discuss our specific regression equations, though, it is important to establish our assumptions of causal inference. Essentially, we must assume that our unobserved error term is uncorrelated with alcohol consumption and thus its impact on respondents' use of hard drugs. While this is fundamentally untestable, we can use balance tests to determine what third factors should be included in our regressions—those correlated with alcohol consumption and hard drug use, which might have an impact on our final estimates. We ran two separate balance tests because of our use of different variables for our different models. For our OLS and fixed effects models, we used dummy variables for whether the respondent has ever consumed alcohol or used hard drugs. In our regression discontinuity model, however, we use dummy variables for whether respondents had consumed alcohol or used hard drugs since their last interview, referred to as the *frequency* of alcohol consumption or hard drug use (*alc_freq* and *drug_freq*). These models and their associated variables will be expanded upon in the subsections to follow.

Our first balance test, shown below in *Figure 4*, corresponds with our OLS and fixed effects regressions. We regressed a variety of potential controls on our variables for participation in alcohol consumption and hard drug use in order to determine what controls to include in our regressions. When regressing third factors on whether a respondent consumed alcohol in the past month, we found that our gender and race dummies—female, White, Black, and

Hispanic—as well as our controls for age and parents' education (*dad_hgc_pos* and *mom_hc_pos*) were statistically significant, and thus we decided to include them in our regressions as controls. The dummy variables for region were insignificant, however, so we determined that they were unnecessary to include in our regression equations. When regressing third factors on the use of drugs since the last interview, we found our gender dummy and age variable to be statistically significant, as well as parents' education. The region dummy for the south was statistically significant, so we also decided to include the dummy variables for region as controls in our OLS and fixed effects regression equations. The dummy variables for race were not statistically significant, but we included them regardless. We are aware that adding unneeded controls can make our regressions less efficient, but we included them anyways in case there is some relationship that we were unable to capture through our balance test due to the limited number of minority races in this sample.

## Figure 4: OLS and Fixed Effects Balance Test

| | Universe: Respondents who answered yes to consuming VARIABLE | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| VARIABLES | female | age | white | black | hispanic | dad_hgc_pos | mom_hgc_pos | north_east | north_central | south |
| alc_participation_month | -0.0131 | 0.0928 | 0.0054 | -0.0032 | -0.0017 | 0.0261 | 0.0278 | -0.0002 | 0.0001 | -0.0003 |
| | (0.0005) | (0.0026) | (0.0006) | (0.0005) | (0.0005) | (0.0043) | (0.0035) | (0.0004) | (0.0005) | (0.0006) |
| Constant | 0.5630 | 20.7300 | 0.5800 | 0.2110 | 0.2120 | 12.6700 | 12.5900 | 0.1720 | 0.2320 | 0.3540 |
| | (0.0066) | (0.0255) | (0.0065) | (0.0053) | (0.0054) | (0.0462) | (0.0399) | (0.0048) | (0.0055) | (0.0061) |
| Observations | 55562 | 55562 | 55562 | 55562 | 55562 | 45518 | 51877 | 55562 | 55562 | 55562 |
| R-squared | 0.0280 | 0.0260 | 0.0050 | 0.0030 | 0.0010 | 0.0030 | 0.0040 | 0.0000 | 0.0000 | 0.0000 |
| F-stat | 572.1000 | 1287.0000 | 90.8700 | 47.4400 | 12.9800 | 36.8800 | 64.4000 | 0.3050 | 0.0240 | 0.2070 |
| p-value | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0003 | 0.0000 | 0.0000 | 0.5810 | 0.8770 | 0.6490 |
| drugs_dli | -0.0002 | -0.0013 | -0.0001 | 0.0000 | -0.0001 | -0.0037 | -0.0017 | -0.0001 | 0.0000 | 0.0002 |
| | (0.0001) | (0.0005) | (0.0001) | (0.0000) | (0.0001) | (0.0007) | (0.0005) | (0.0001) | (0.0001) | (0.0001) |
| Constant | 0.4610 | 20.3200 | 0.7280 | 0.0852 | 0.2170 | 13.0400 | 13.0000 | 0.1770 | 0.2040 | 0.3100 |
| | (0.0142) | (0.0634) | (0.0117) | (0.0062) | (0.0114) | (0.1050) | (0.0852) | (0.0108) | (0.0107) | (0.0125) |
| Observations | 4738 | 4738 | 4738 | 4738 | 4738 | 3880 | 4426 | 4738 | 4738 | 4738 |
| R-squared | 0.0020 | 0.0020 | 0.0000 | 0.0000 | 0.0000 | 0.0130 | 0.0040 | 0.0000 | 0.0000 | 0.0010 |
| F-stat | 6.5910 | 6.6780 | 1.2080 | 0.8220 | 1.7360 | 26.2500 | 12.2400 | 0.7330 | 0.0509 | 3.6560 |
| p-value | 0.0103 | 0.0098 | 0.2720 | 0.3650 | 0.1880 | 0.0000 | 0.0005 | 0.3920 | 0.8220 | 0.0560 |
| | Universe : Respondents who answered yes to consuming Alcohol and VARIABLE | | | | | | | | | |
| drugs_dli | -0.0002 | -0.0016 | -0.0001 | 0.0001 | -0.0001 | -0.0036 | -0.0019 | -0.0001 | 0.0000 | 0.0002 |
| | (0.0001) | (0.0005) | (0.0001) | (0.0001) | (0.0001) | (0.0007) | (0.0005) | (0.0001) | (0.0001) | (0.0001) |
| Constant | 0.4630 | 20.4600 | 0.7410 | 0.0710 | 0.2150 | 13.1000 | 13.0800 | 0.1780 | 0.2060 | 0.3040 |
| | (0.0147) | (0.0642) | (0.0120) | (0.0059) | (0.0118) | (0.1080) | (0.0876) | (0.0112) | (0.0113) | (0.0130) |
| Observations | 4381 | 4381 | 4381 | 4381 | 4381 | 3623 | 4102 | 4381 | 4381 | 4381 |
| R-squared | 0.0020 | 0.0030 | 0.0000 | 0.0010 | 0.0010 | 0.0130 | 0.0040 | 0.0000 | 0.0000 | 0.0020 |
| F-stat | 4.4580 | 10.0700 | 1.2890 | 2.2040 | 2.2650 | 27.4000 | 15.0100 | 0.6050 | 0.3040 | 4.7080 |
| p-value | 0.0349 | 0.0015 | 0.2560 | 0.1380 | 0.1320 | 0.0000 | 0.0001 | 0.4370 | 0.5820 | 0.0301 |

Our second balance test, shown in *Figure 5*, corresponds to our RD regressions. The difference between this balance test and the first one is that this one regressed potential controls on our variables for *frequency* of alcohol consumption and hard drug use, rather than participation. We used the frequency variables in our RD model because they did not have the issue of nonresponse bias and thus enabled us to expand our sample, as was discussed in our Data section. When we regressed the dummy variable for the frequency of alcohol consumption on third factors, we determined that our gender and race dummies (female, White, Black, and Hispanic), age, parents' education (*dad_hgc_pos* and *mom_hc_pos*), and our dummy variables for region were all statistically significant, and thus we decided to include them in our RD regressions as controls. Regressing the same third factors on the dummy variable for frequency of hard drug use also gave us statistically significant results. As shown in *Figure 5*, the dummy variables for Hispanic and the northeast region were not statistically significant, but are still included in the regression equations as this simply means they were not significantly different from the excluded variables for race (non-White non-Hispanic minorities) and region (West).

Figure 5: Regression Discontinuity Balance Test

| | Universe: Respondents who answered yes to consuming VARIABLE | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | (1) female | (2) age | (3) white | (4) black | (5) hispanic | (6) dad_hgc_pos | (7) mom_hgc_pos | (8) north_east | (9) north_central | (10) south |
| alc_freq | -0.0170 | 1.5730 | 0.1920 | -0.1740 | -0.0226 | 0.7510 | 0.8300 | 0.0211 | 0.0426 | -0.0968 |
| | (0.0077) | (0.0364) | (0.0074) | (0.0070) | (0.0062) | (0.0569) | (0.0474) | (0.0053) | (0.0060) | (0.0073) |
| Constant | 0.5090 | 19.7400 | 0.4170 | 0.3680 | 0.2270 | 12.0700 | 11.9100 | 0.1510 | 0.1910 | 0.4520 |
| | (0.0075) | (0.0321) | (0.0073) | (0.0074) | (0.0063) | (0.0563) | (0.0466) | (0.0051) | (0.0056) | (0.0073) |
| Observations | 84478 | 84478 | 84478 | 84478 | 84478 | 67186 | 78190 | 84478 | 84478 | 84478 |
| R-squared | 0.0000 | 0.0460 | 0.0330 | 0.0360 | 0.0010 | 0.0120 | 0.0180 | 0.0010 | 0.0020 | 0.0090 |
| F-stat | 4.9130 | 1870.0000 | 682.3000 | 622.4000 | 13.1300 | 173.8000 | 306.6000 | 16.1300 | 50.4000 | 174.6000 |
| p-value | 0.0267 | 0.0000 | 0.0000 | 0.0000 | 0.0003 | 0.0000 | 0.0000 | 0.0001 | 0.0000 | 0.0000 |
| drugs_freq | -0.0477 | -0.5220 | 0.1840 | -0.1710 | 0.0024 | 0.2920 | 0.4610 | 0.0098 | -0.0197 | -0.0717 |
| | (0.0134) | (0.0610) | (0.0113) | (0.0069) | (0.0108) | (0.1020) | (0.0802) | (0.0101) | (0.0099) | (0.0119) |
| Constant | 0.5010 | 20.8300 | 0.5350 | 0.2620 | 0.2120 | 12.5700 | 12.4500 | 0.1640 | 0.2210 | 0.3910 |
| | (0.0056) | (0.0172) | (0.0055) | (0.0049) | (0.0045) | (0.0399) | (0.0340) | (0.0040) | (0.0044) | (0.0052) |
| Observations | 83984 | 83984 | 83984 | 83984 | 83984 | 66793 | 77762 | 83984 | 83984 | 83984 |
| R-squared | 0.0000 | 0.0010 | 0.0070 | 0.0080 | 0.0000 | 0.0000 | 0.0010 | 0.0000 | 0.0000 | 0.0010 |
| F-stat | 12.7500 | 73.3000 | 264.4000 | 626.2000 | 0.0493 | 8.1720 | 33.0100 | 0.9520 | 3.9400 | 36.2000 |
| p-value | 0.0004 | 0.0000 | 0.0000 | 0.0000 | 0.8240 | 0.0043 | 0.0000 | 0.3290 | 0.0472 | 0.0000 |
| | Universe : Respondents who answered yes to consuming Alcohol and VARIABLE | | | | | | | | | |
| drugs_freq | -0.0479 | -0.5180 | 0.1850 | -0.1710 | 0.0024 | 0.2930 | 0.4610 | 0.0098 | -0.0195 | -0.0721 |
| | (0.0134) | (0.0610) | (0.0113) | (0.0068) | (0.0108) | (0.1020) | (0.0803) | (0.0101) | (0.0099) | (0.0119) |
| Constant | 0.5010 | 20.8300 | 0.5360 | 0.2620 | 0.2120 | 12.5700 | 12.4500 | 0.1640 | 0.2210 | 0.3910 |
| | (0.0056) | (0.0172) | (0.0055) | (0.0049) | (0.0045) | (0.0399) | (0.0340) | (0.0040) | (0.0044) | (0.0052) |
| Observations | 83742 | 83742 | 83742 | 83742 | 83742 | 66632 | 77544 | 83742 | 83742 | 83742 |
| R-squared | 0.0000 | 0.0010 | 0.0070 | 0.0080 | 0.0000 | 0.0000 | 0.0010 | 0.0000 | 0.0000 | 0.0010 |
| F-stat | 12.8200 | 72.2100 | 265.5000 | 629.9000 | 0.0481 | 8.2750 | 32.9600 | 0.9410 | 3.8620 | 36.5300 |
| p-value | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.8260 | 0.0040 | 0.0000 | 0.3320 | 0.0494 | 0.0000 |

Finally, we also clustered all of our regressions by individuals to account for time invariant, unobservable characteristics that are particular to each respondent.

## OLS

Our first OLS model simply looks at the effects of alcohol consumption on hard drug use. Here we regress *drugs_dli*, a dummy for whether or not a respondent has used hard drugs since the last interview, on *alc_participation_month*, a dummy for whether or not a respondent has consumed alcohol in the past month. We included year effects in this regression, but did not control for any third factors.

$$drugs\_dli_{it} = \beta_0 + \beta_1 alc\_participation\_month_i + u_{it}$$

Because our first regression is very simple and could suffer from omitted variable bias, our next OLS regression added controls for race (*black, hispanic,* and *other_nonwhite_participants*), gender (*female*), and parents' education (*dad_hgc_pos,* and *mom_hgc_pos*), all of which are represented by the term $X_{it}$. We decided to control for race, gender, and parents' education because they are time invariant and can show different effects across demographics. Additionally, papers by Monica Deza[4] and Benjamin Crost and Daniel Rees[5] which served as inspiration for our own paper include gender, race, and education attainment controls. However, it is important to note that the paper by Crost and Rees includes other controls such as student status, employment status, marital status, and income. We chose not to include employment status and income because we are interested in analyzing respondents around the age of 21, which would include a wide range of employment statuses and incomes. Additionally, the inclusion of marital status for respondents around the age of 21 did not seem productive or necessary. We chose to include the educational attainment of one's *parents* instead of individual educational attainment because there were no clear variables for individual's educational attainment that could be used without major manipulation of the data, which would have been unrealistic given our timeframe. Finally, we believe that because many other papers include the same controls (marital status, income, education, employment, etc.) our paper will contribute a new analysis and perspective on the issue.

$$drugs\_dli_{it} = \beta_0 + \beta_1 alc\_participation\_month_i + X_{it} + u_{it}$$

Because we are attempting to understand more about the relationship between consumption of alcohol and use of hard drugs, we thought it could be helpful to analyze the effects of being over 21 on the respondents' frequency of using hard drugs (a dummy variable called

[4] Deza, Monica. "The effects of alcohol on the consumption of hard drugs: regression discontinuity evidence from the National Longitudinal Study of Youth, 1997." *Health economics* 24.4 (2015).

[5] Crost, Benjamin, and Daniel I. Rees. "The Minimum Legal Drinking Age and Marijuana Use: New Estimates from the NLSY97." *Journal of Health Economics* 32, no. 2 (March 1, 2013): 474–76. https://doi.org/10.1016/j.jhealeco.2012.09.008.

*drugs_freq*). We decided to run a regression of drug use since the last survey on a dummy variable (*over_21*) for being of legal drinking age. This regression includes all of our controls, as well as year effects.

$$drugs\_freq_{it} = \beta_0 + \beta_1 over\_21_{1it} + X_{it} + u_{it}$$

The next regression we ran was similar to the regression above, but this time we used a dummy variable for frequency of alcohol consumption (*alc_freq*) rather than for hard drug use. Here we looked at whether or not an individual was more likely to consume alcohol after turning 21 years old. We ran this regression with all of our controls as well as year effects.

$$alc\_freq_{it} = \beta_0 + \beta_1 over\_21_{1it} + X_{it} + u_{it}$$

## Fixed Effects

Because our dataset was in the form of panel data, we believe fixed effects regressions are important to include. Fixed effects will help us capture any fixed biases in our dataset and hopefully get our analysis closer to the true relationship between alcohol and hard drugs. Our first fixed effects regression controls for individual fixed effects ($a_{it}$) as well as year effects, and looks at alcohol consumption in the past month related to hard drug use since the last survey. Controlling for individual effects allowed us to control for any differences in individual results that were due to these "family effects." Our year fixed effects will control for all factors changing each year that are common to all respondents in a given year.

$$drugs\_dli_{it} = \beta_0 + \beta_1 alc\_participation\_month_{1it} + a_i + u_{it}$$

Our next consideration was the effect of regions on this relationship between alcohol consumption and hard drug use. It is plausible to think that depending on the region of the country, these effects could change. Therefore, our next fixed effects regression looks instead at the regional fixed effects ($r_{it}$) and includes the set of controls used in the above OLS

regressions (*black, hispanic, other_nonwhite_participants, female, parent_education*). This model again looks again at the relationship between alcohol consumption in the past month and hard drug use since the last interview.

$$drugs\_dli_{it} = \beta_0 + \beta_1 alc\_participation\_month_{1it} + r_{it} + X_{it} + a_i + u_{it}$$

The final fixed effects regression we analyzed is a first-stage regression controlling for individual fixed effects and year effects. Here we are attempting to analyze a similar relationship to our OLS regression of alcohol frequency on turning 21, while making sure to control for the fixed effects. We think this is particularly important to include because of the MLDA of 21. Since hard drugs are never legal (even at 21), we do not believe this regression would be helpful to include for hard drug use.

$$alc\_participation\_month_{it} = \gamma_0 + \gamma_1 over\_21_{1it} + \gamma_{it} + a_i + u_{it}$$

## Regression Discontinuity

After working through the fixed effects model we decided to explore the use of a regression discontinuity model to further analyze the effects of alcohol consumption on hard drug use. An RD model would allow us to examine the differences in alcohol consumption and hard drug use before and after a respondent turns 21. We would expect to see a jump in alcohol consumption, but are curious what effect, if any, turning 21 would have on hard drug use. We believe that this model could provide us with interesting insights about how the relationship between alcohol consumption and hard drug use changes when a respondent turns 21. We are using age as our running variable, and controlling for race, gender, region, year effects, and highest grade of education completed by each parent. Our estimating equation is:

$$a^P = \beta_1 a_{it} + \ldots + \beta_4 a_{it}^4$$

$$Y_{it} = \alpha + \rho D_{it} + \gamma a + a^{P} D_{it} + e_{it}$$

Our RD model is similar to a 2014 study[6] by Monica Deza that uses the NLSY 1997 data set to examine the effects of alcohol consumption on the initiation to and frequency of hard drug use. Deza's use of a regression discontinuity model is similar to ours, as she describes the purpose is to "disentangle the causal effect of alcohol consumption on other risky behaviors" and to "identify a research design that involves an exogenous variation in alcohol consumption."[7] Using the same age-based threshold in an RD design, we are able to explore the relationship between alcohol consumption and age. This allowed us to assess whether the change in the relationship between alcohol consumption and hard drug use is affected in any way by the relationship between alcohol consumption and age. A key difference in our study compared to Deza's is our sample size; as described in our Data section, our data was collected from 1998 to 2008, while hers cover the range of 1997 to 2009. Another difference is that we focus on the effects of consuming alcohol on whether or not one uses hard drugs, rather than *also* considering its effect on the frequency of hard drug use. Another useful aspect of Deza's study is the introduction of a falsification test which we will use to strengthen our justification of the RD model, elaborated upon in our Discussion and Conclusions section.

In order for our RD model to be valid, it must satisfy the identification assumptions. The first of these assumptions states that all variables that might affect the outcome variable be continuous at the threshold. In other words, no other variable which affects alcohol consumption or drug use should experience a jump at 21. There is no reason that any other

[6] Deza, Monica. "The effects of alcohol on the consumption of hard drugs: regression discontinuity evidence from the National Longitudinal Study of Youth, 1997." *Health economics* 24.4 (2015).
[7] Ibid., pp 419-438.

variables associated with alcohol consumption would experience a significant jump at age 21. 21 is a pretty arbitrary age, and the only other significant things which become legal in the United States at 21 are gambling and recreational marijuana use in certain states. Gambling should not have any major effect on our results, as it isn't extremely pervasive in our culture. Recreational marijuana use might be a variable to consider if our data were more recent, but in 2014, the last year of this panel data, only 2 states had legalized the recreational use of marijuana. As a result, the effect of marijuana becoming legal should not have too much of an impact on our regression discontinuity model, or any of our models, for that matter. Also, we can say confidently that there are no unobservable differences between people who are 20 and 21, leading us to conclude that any observed jump in our regression at the threshold is a result of the increased access to alcohol at the MLDA of 21.

The second of the identification assumptions is that there must be a clear jump in the outcome variable at the threshold. This assumption is easier to prove, as we have empirical evidence backing up this "jump." This evidence comes from running our regression, which will be expanded upon in our Results section.

# Results

Below, we report the results from each of the regressions we ran, as well as some insights into our findings and our thought processes while analyzing the results. First, we will talk through our various OLS regressions, the distinctions between each model, and our findings. Then, we will look at our fixed effects models, pointing out any distinctions between the regressions and laying out our findings. Lastly, we will cover our regression discontinuity models, and we will touch on some concerns we have regarding the validity of this model.

# OLS

Our first OLS model simply looks at the effects of alcohol consumption on hard drug use. Here we regress *drugs_dli*, a dummy for whether or not a respondent has used hard drugs since the last interview, on *alc_participation_month*, a dummy for whether or not a respondent has consumed alcohol in the past month. We included year effects in this regression, but didn't control for any third factors. The results are shown in column (2) of *Figure 6*. We found that if a respondent consumed alcohol, there was a 39.2% increase in probability of doing hard drugs. This result is significant at an 11.3% level which does not satisfy our requirement of significance at a 5% level. This makes sense because it is likely that there are many third factors associated with both alcohol consumption and drug use that we have not yet controlled for.

$$drugs\_dli_{it} = \beta_0 + \beta_1 alc\_participation\_month_{1it} + u_{it}$$

Our next OLS regression added controls for race and gender, all of which is represented by the term $X_i$. The results are shown in *Figure 6* column 4. Here we see that, when controlling for race, gender, and parents' education, if a respondent consumed alcohol in the past month, there was a 41.6% increase in probability that the respondent consumed hard drugs. This result is significant at a 13% level. Again, this does not satisfy our requirement of a 5% significance level and therefore we do not consider this result statistically significant.

$$drugs\_dli_{it} = \beta_0 + \beta_1 alc\_participation\_month_{1it} + X_{it} + u_{it}$$

We then decided to run a regression of drug use on a dummy variable for being of legal drinking age. Here we are looking at *frequency* of drug use. This regression includes all of our controls, as well as year effects. The results are in *Figure 7*. Here, we found that if a respondent is over 21, there is a -0.8% decrease in the probability that the respondent will use hard drugs. This regression was statistically significant at a 5.5% level. This result is not

statistically significant at a 5% level which is our threshold, however it is very close which is important to note.

$$drugs\_freq = \beta_0 + \beta_1 over\_21_{1it} + X_{it} + u_{it}$$

The next regression we ran was similar to the regression above, but this time we used alcohol consumption rather than hard drug use. Here we looked at whether or not an individual was more likely to consume alcohol after turning 21 years old. We ran this regression with all of our controls as well as year effects. The results are in *Figure 7*. Here we saw that being over 21 is associated with a 12.2 percentage point increase in probability that the respondent drank alcohol. This result is significant at a 0.00% level.

$$alc\_freq_{it} = \beta_0 + \beta_1 over\_21_{1it} + X_{it} + u_{it}$$

## Fixed Effects

After running our OLS regressions with and without controls, we ran regressions controlling for fixed effects. First, we looked at the effects of alcohol consumption on hard drug use while controlling for individual fixed effects and for year effects, but without explicitly including controls for race, gender, and parents' education. Here we found that if a respondent consumed alcohol in the past month, there was a 13.9% increase in the probability of doing hard drugs. With a p-value of 0.751, this result is not significant at a 5% significance level. Below, we explore other fixed effects regressions to learn more about this relationship.

$$drugs\_dli_{it} = \beta_0 + \beta_1 alc\_participation\_month_{1it} + a_i + u_{it}$$

Next, we looked at the effects of alcohol consumption on hard drug use while absorbing for region fixed effects, as well as controlling for year effects, race, gender, and parents' education. We found that if the participant consumed alcohol in the past month, there was a 33.6% increase in the probability of hard drug use compared to those that did not consume

alcohol in the past month. With a p-value of 0.172, this result is not statistically significant at a 5% significance level. This is consistent with other regressions we have run, and it makes sense that regional factors are unlikely to change the relationship between alcohol consumption and hard drug use. After running fixed effect regressions on alcohol consumption and drug use, controlling first for individual effects and year effects and then for region and year effects, we did not find a very strong correlation between alcohol consumption and drug use.

$$drugs\_dli_{it} = \beta_0 + \beta_1 alc\_participation\_month_{1it} + r_i + X_{it} + u_{it}$$

We also ran a first-stage fixed effects regression controlling for individual effects, as well as year effects, on the relationship between a dummy for whether a participant is over 21 years old and alcohol consumption in the past month. We found that a person who is over 21 years old will consume 156.7 percentage points more alcohol (# of drinks) than a person who is under 21 years old, which is shown in *Figure 20*. This gives a significance at the 1% level and beyond. It makes sense that we found statistical significance given that the US minimum drinking age is 21. This could be because people under 21 years of age are genuinely not consuming alcohol, or given that it is illegal, respondents under 21 are purposely under-reporting alcohol consumption. This is potentially indicative of non-classical measurement error because misreporting is not done at random. As it impacts more than just our fixed effects model, we will further discuss the issue of nonresponse and misreporting in the Discussion and Conclusions section of our paper.

$$alc\_participation\_month_{it} = \gamma_0 + \gamma_1 over\_21_{1it} + a_i + u_{it}$$

# Regression Discontinuity

## General

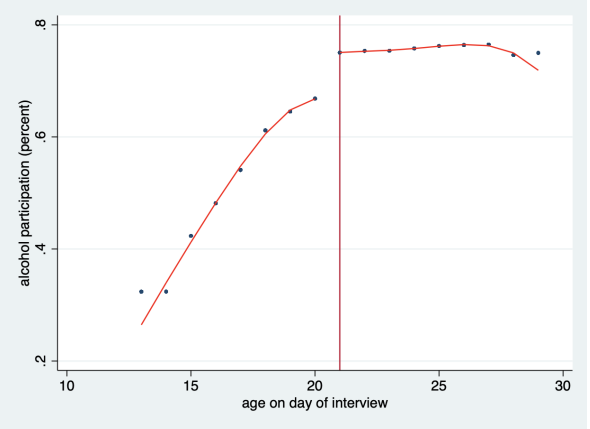
The results from our OLS and fixed effects regression did not tell the full story of the relationship between alcohol consumption and hard drug use. While we included a dummy for being over the age of 21 in some of our earlier models, they didn't provide us with enough information to determine how alcohol consumption and hard drug use are impacted when a respondent turns 21 years of age. In order to further explore this, we decided to implement a regression discontinuity model. This model would confirm whether or not there is a clear spike in alcohol consumption at the legal drinking age of 21. Further, the increased access to alcohol (and unchanged access to hard drugs) at age 21 might have an effect on the relationship between alcohol consumption and hard drug use. A regression discontinuity model can give extremely credible estimates if it uses enough data, as ours does. As discussed in our Data section, this model enabled us to address the large amounts of nonresponse bias that hindered our OLS and fixed effects regressions.

We set up two RD models to address this question: one for alcohol consumption and one for hard drug use. We established age as our running variable, measured in years. If we had more precise data, we would have liked to scale our age variable to months—or even days—and look at the data points only closely surrounding the threshold. Unfortunately, our data did not provide such precise measurements, nor was there a way to find such data due to confidentiality constraints. Included below are the regression discontinuity models for hard drug use and alcohol consumption, along with our interpretations of these models.

## Alcohol

This regression discontinuity model allowed us to investigate whether there was a jump in alcohol consumption at the minimum legal drinking age (MLDA) of 21. We regressed alcohol frequency, a dummy variable for whether respondents reported having consumed alcohol in the last year, on age, with a threshold at 21. Our RD model showed a sharp jump upward at the threshold, as shown in *Figure 10*. The upward trend in alcohol consumption before the threshold—which we know is likely an accurate trend—followed by the clear jump at the threshold, leads us to conclude that while alcohol consumption generally increases before the age of 21, there is a significant, discontinuous jump upward at the MLDA of 21.

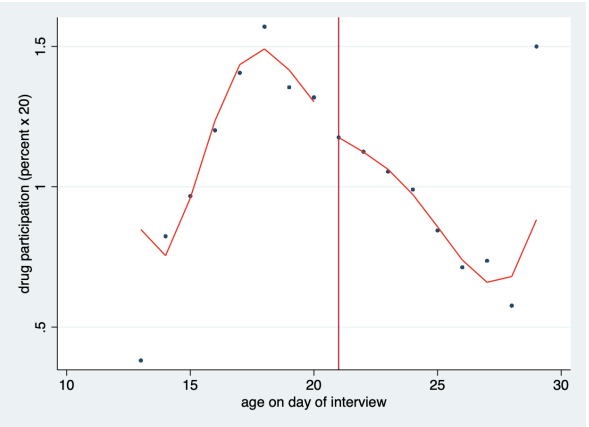
Figure 10: Alcohol Frequency RD




## Hard Drugs

We next ran a regression discontinuity model on hard drug frequency, which is also a dummy variable measuring whether or not participants had used hard drugs since their last interview (which was a year ago in this case). Age, measured in years, is still our running variable. We were interested in finding whether there was an arbitrary jump in the regression when respondents turned 21 (even though hard drugs remain illegal). As seen in *Figure 11*, we also see a jump in the

Figure 11: Hard Drug Frequency RD



regression at the threshold with this RD model. However, this jump is not nearly as sharp as in the alcohol RD model, and as a result, we are not able to draw any major conclusions from this model. This inconclusivity stands in stark contrast to our alcohol consumption RD model, which showed a clear jump in alcohol consumption when the subject became of legal drinking age.

## Interpretation

When we look at the alcohol RD model, we can see a sharp upward jump in alcohol consumption at the threshold. In the hard drug RD model, however, there is no clear, significant jump at the threshold—only a gradual decrease in hard drug use as age increases. The empirical data in *Figure 13* supports these interpretations. The alcohol consumption regression in column 2 shows a statistically significant jump at the age threshold of 21, with a p-value of less than 0.01, significant at the 1% level. In addition, when we create alcohol consumption regressions for age thresholds of 20 and 22, we observe that there is no clear jump in alcohol consumption, and the estimate is arbitrarily close to zero. This implies that alcohol consumption jumps discontinuously at age 21, and that the jump is statistically significant. On the other hand, when we look at the hard drug use regression in column 1 of *Figure 13*, we cannot make the same interpretations about the

Figure 13: Regression Discontinuity Robustness

| VARIABLES | (1) drugs_freq | | | (2) alc_freq | | |
|---|---|---|---|---|---|---|
| | Age 20 | Age 21 | Age 22 | Age 20 | Age 21 | Age 22 |
| Di | -0.114*** | -0.0667*** | 0.0836*** | -0.00412*** | 0.0517*** | -0.00739*** |
| | (0.0010) | (0.0010) | (0.0015) | (0.0002) | (0.0002) | (0.0002) |
| Constant | -27.62*** | -36.26*** | -37.99*** | 2.225*** | -0.793*** | 1.651*** |
| | (0.6790) | (0.6380) | (0.6710) | (0.0833) | (0.1170) | (0.0666) |
| Controls | Y | Y | Y | Y | Y | Y |
| Year Effects | Y | Y | Y | Y | Y | Y |
| Cluster (id) | Y | Y | Y | Y | Y | Y |
| age-poly interaction | Y | Y | Y | Y | Y | Y |
| Observations | 64,646 | 64,646 | 64,646 | 64,646 | 64,646 | 64646 |
| R-squared | 0.937 | 0.932 | 0.934 | 0.988 | 0.996 | 0.988 |
| F-stat | 11940 | 4783 | 2939 | 275.7 | 67658 | 1993 |
| p-value | 0 | 0 | 0 | 0 | 0 | 0 |

discontinuity. Even though there appears to be a discontinuous drop in hard drug consumption at age 21, when we run the same regression but with thresholds at 20 and 22 years of age, the estimates are indistinguishable from those in the 21-year threshold regression. As a result, we decided not to draw any major conclusions from the hard drug regression discontinuity model. However, the downward trend of the hard drug RD model (negative slope coefficient) coupled with the alcohol RD model's upward trend and sharp jump suggests that hard drugs and alcohol trend in opposite directions, and thus may be substitutes, though we do not have nearly enough evidence to make such a strong claim.

## Robustness

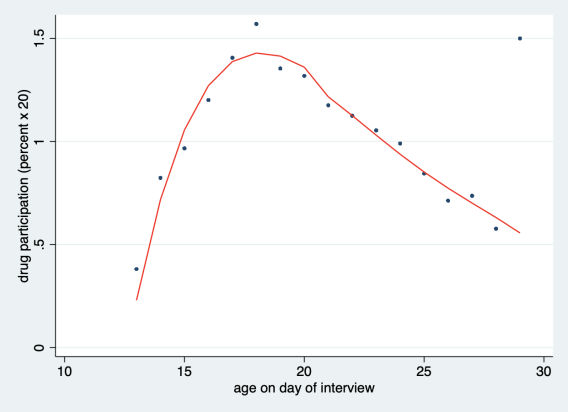
We also examined whether there was random sorting of demographics at the age threshold of 21, shown in a visual balance test in *Figures 14-19*. These graphs provide evidence that there is a smooth distribution of race and gender demographics and parents' education around the age threshold of 21. As a result, we can attribute any changes in alcohol and/or drug consumption that occur at the age of 21 years to the exogenous decrease in the consequences of alcohol consumption or the increase in the accessibility of alcohol that occurs at the MLDA.

One concern we have with our analysis of alcohol consumption and its relation to turning 21 is the potential of the estimate being biased due to birthday celebratory effects. Our inability to draw a stronger inference from our model is due to the fact that we do not have access to more granular age and date of birth data from each respondent due to confidentiality concerns. Having access to month-wise age data would allow us to be more precise with our running variable in the RD model, while access to dates of birth for each respondent would allow us to control for birthday celebratory effects. Because our dataset does not include

participants' birthdays, it is impossible to rule this out as a source of bias for our analyses. However, in the similar paper by Monica Deza, she is able to rule out birthday celebratory effects from significantly affecting the discontinuity in alcohol and drug regressions due to her access to confidential data from the NLSY97.[8]

## Challenges

Figure 12: Non-RD Hard Drug Participation Model




A challenge we faced with regression discontinuity was distinguishing between a discontinuity and a nonlinearity in our regression. This is more easily addressed in the case of alcohol consumption, where there was a clear jump in its relationship with age at the threshold of 21. Alcohol consumption consistently increases with age before the age of 21 and then experiences an upward jump in alcohol consumption that remains relatively constant as age increases from 21. This could be as a result of a genuine increase in alcohol consumption, but it could also be due to non-classical measurement error in our data. Given that it is illegal to consume alcohol before the age of 21, participants younger than 21 most likely under-reported their alcohol consumption. This kind of under-reporting is also a potential issue with hard drug use as most hard drugs are illegal in the United States regardless of age.

The question of nonlinearity is more complicated when concerning the hard drug use RD model. When we regressed hard drug use on age *without* a discontinuity model, our regression showed a curve connecting the data points on both sides of 21 with no major

[8] Deza, Monica. "The effects of alcohol on the consumption of hard drugs: regression discontinuity evidence from the National Longitudinal Study of Youth, 1997." *Health economics* 24.4 (2015).

jumps, as shown in *Figure 12*. The lack of any discontinuity in this regression is a significant cause for concern and is the primary reason we won't be drawing major conclusions from this model. As such, we are not able to conclude that the RD model is an appropriate method to consider the relationship between alcohol consumption and hard drug use, and we should therefore rely on other methods to address our hypothesis. It is still worth noting, however, that our RD model showed a distinct jump in alcohol consumption at the MLDA threshold.

# Discussion and Conclusions

This paper considers the effects of alcohol consumption on hard drug use. Further, it considers the discontinuous changes in alcohol consumption and hard drug use at the age of 21. We see that the probability of consumption of alcohol increases significantly (significance level of 0.000%) by 12.2 percentage points (column (1) of *Figure 7*) after a respondent turns 21. We also found that a person who is over 21 is 156.7 percentage points (*Figure 20*) percentage points more alcohol (# of drinks) than a person who is under 21 years old, also statistically significant at a 1% level (*Figure 20*). These results fall in line with a similar paper by Monica Deza[9] that finds a statistically significant positive relationship between being 21 and consuming alcohol (7 to 8 percentage points) as well as a statistically significant negative relationship between being 21 and using hard drugs (1.5 to 2 percentage points). Similarly to us, she also finds that there is a statistically significant change in alcohol consumption after the age of 21.

An important note to mention about our findings is the issue of nonresponse bias. The nonresponse found in questions concerning alcohol consumption and hard drug use affected

[9] Deza, Monica. "The effects of alcohol on the consumption of hard drugs: regression discontinuity evidence from the National Longitudinal Study of Youth, 1997." *Health economics* 24.4 (2015).

every regression model we ran, impacting any conclusions we were able to draw from them. There is no way to definitively determine the cause of this nonresponse bias, but there are a host of unobservable factors that could cause such bias. For example, we found that respondents who didn't identify as White were less likely to answer questions relating to alcohol and drugs. This is likely due to the inequities in the drug conviction process which disproportionately affect minorities,[10] but again, there is no way to empirically verify whether or not this is the cause of our nonresponse issue—see our discussion of this issue in the Methods section for further information. As a result of this bias, our study's conclusions apply only to the population that provided information about their alcohol consumption and hard drug use, rather than to any population.

The issue of nonresponse also raises validity questions in our models, specifically concerning our regression discontinuity model. One concern is that the jump we observed in the RD model for alcohol consumption was likely due to the fact that survey participants didn't respond to questions about alcohol consumption until they were 21 and above the MLDA. We can confirm this is false based on our robustness check in *Figure 13*, where we observe that there are the exact same number of respondents providing information about their frequency of alcohol consumption at ages 20, 21, and 22. Another possibility is that respondents lied about their alcohol consumption until they were 21, afraid of the legal consequences that might result from admitting to underage drinking. The latter scenario would be more problematic, as non-classical measurement error in alcohol consumption would lead to potential bias in our model. Because this is unmeasurable, we must make the assumption that our dataset does not suffer from such measurement error, allowing us to

[10] Ferrer, Barbara, and John M Connolly. "Racial Inequities in Drug Arrests: Treatment in Lieu of and After Incarceration." *American journal of public health* vol. 108,8 (2018): 968-969. doi:10.2105/AJPH.2018.304575

make the conclusions outlined below. Moreover, a 2009 paper by Carpenter and Dobkin[11] argues that because there is a large, discrete jump in alcohol-related deaths at age 21, there is compelling evidence of a change in alcohol consumption, ruling out nonresponse bias as the sole reason for this jump. This is compelling evidence that the discontinuous jump in alcohol consumption at age threshold of 21 is not driven by the underreporting of alcohol use by respondents under the age of 21.

Through our research, we were able to draw the following conclusions:

First, alcohol consumption and hard drug use trend in opposite directions. We see from the OLS model of alcohol consumption on being 21 or older being positive (with a slope of 0.119) and the OLS model of hard drug use on being 21 or older being negative (with a slope of -0.008) that there is an increase in alcohol consumption and a decrease in hard drug use after 21. Additionally, in the RD model for alcohol and age (*Figure 10*), we were able to see on the graph a clear and somewhat consistent increase in alcohol consumption from the youngest age up until the threshold of 21. After age 21, there seems to be a plateau, ending with a slight decrease of alcohol consumption for the highest ages. When looking at the RD model for hard drugs and age (*Figure 11*) there is a less clear story. Here, we find an increase in the use of hard drugs until around age 17 followed by a decrease in the use of hard drugs until around age 27. The most important relationship that RD illuminated is that there is a clear increase in alcohol consumption at age 21, and a clear decrease in drug use around that same age. This is not enough evidence to claim that the two are substitutes, but in Deza's paper she is able to draw the conclusion that they are in fact substitutes, a result which remained robust when changing the age bandwidths.

[11] Carpenter C, Dobkin C. 2009. "The effect of alcohol consumption on mortality: regression discontinuity evidence from the minimum drinking age." *American Economic Journal: Applied Economics* 1(1): 164–182.

Our second result is that our research does not support the idea that alcohol is a "gateway drug." A gateway drug is a drug where the use of that less dangerous substance will lead to the use of harder and more dangerous drugs over time. Our research and results do not support the idea that alcohol is a gateway for hard drugs. For that to be the case, we would have to have observed an increase in alcohol consumption, followed by an increase in hard drug consumption and a simultaneous decrease in alcohol consumption. Looking at the two RD models, we first see an increase in both alcohol consumption and hard drug use, but as alcohol keeps increasing and maintains a steady level of consumption, hard drug use actually begins to decline after about age 17.

Our final conclusion was that both alcohol consumption and drug use clearly change with age. We discuss some implications of this conclusion in the above sections, but it is important to consider the implications of this realization. When looking at the RD models and graphs, we see clear and steep increases in alcohol consumption before participants turn 21. This could be a result of the non-response bias previously discussed. This could also be because younger participants are worried about admitting an illegal action on the survey, a fear which potentially decreases over the years. However, this could also be because there is in fact some connection between increasing age and the likelihood of drinking. However, the general plateau of alcohol consumption after 21 indicates that once it is legal, anyone who would drink alcohol is drinking alcohol, leading us to believe that any jump in our regression at the threshold is in fact due to the threshold and not some other aspect of the relationship between alcohol consumption and age.

# Acknowledgments

We would like to thank our Professor Eric Chyn for his help and guidance throughout this process. We would also like to thank our Teaching Assistant Jonathan Liu, who was always able to provide extra help and explanations on short notice.

# Appendix

**Figure 1: Overall Table of Means for NLSY97**

| VARIABLES | Overall | Women | Men | White | Non-White | Black | Hispanic |
|---|---|---|---|---|---|---|---|
| white | 0.506 | | | | | | |
| | (0.500) | | | | | | |
| black | 0.235 | | | | | | |
| | (0.424) | | | | | | |
| hispanic | 0.212 | | | | | | |
| | (0.409) | | | | | | |
| female | 0.488 | | | | | | |
| | (0.500) | | | | | | |
| ever alcohol | 0.946 | 0.940 | 0.952 | 0.972 | 0.927 | 0.927 | 0.952 |
| | (0.226) | (0.238) | (0.214) | (0.164) | (0.260) | (0.260) | (0.214) |
| ever hard drugs | 0.257 | 0.232 | 0.281 | 0.322 | 0.179 | 0.160 | 0.269 |
| | (0.437) | (0.422) | (0.449) | (0.467) | (0.383) | (0.367) | (0.444) |
| alcohol initiation age | 12.570 | 12.876 | 12.291 | 12.621 | 12.313 | 12.284 | 12.625 |
| | (2.687) | (2.529) | (2.795) | (2.444) | (3.116) | (3.150) | (2.756) |
| hard drug initiation age | 13.505 | 13.768 | 13.240 | 13.679 | 11.873 | 11.500 | 13.504 |
| | (2.867) | (2.472) | (3.198) | (2.473) | (5.150) | (5.540) | (2.597) |
| age (1997) | 14.354 | 14.367 | 14.341 | 14.287 | 14.319 | 14.333 | 14.347 |
| | (1.488) | (1.490) | (1.487) | (1.482) | (1.494) | (1.491) | (1.496) |
| Respondents | 8984 | 4385 | 4599 | 4548 | 2466 | 2107 | 1901 |

Standard deviations in parenthesis

## Figure 2: OLS and Fixed Effects Table of Means

| OLS and Fixed Effects | | | |
|---|---|---|---|
| **VARIABLE** | **Overall** | **(1)** | **(2)** |
| white | 0.5062 | 0.7364 | 0.6095 |
| | (0.5000) | (0.4407) | (0.4879) |
| black | 0.2345 | 0.0742 | 0.1933 |
| | (0.4237) | (0.2621) | (0.3949) |
| hispanic | 0.2116 | 0.2109 | 0.2034 |
| | (0.4084) | (0.4080) | (0.4025) |
| female | 0.4881 | 0.4542 | 0.4919 |
| | (0.4999) | (0.4980) | (0.4999) |
| dad_hgc_pos | 12.5643 | 12.9500 | 12.8162 |
| | (3.2117) | (3.2342) | (3.1609) |
| mom_hgc_pos | 12.4381 | 13.0017 | 12.7435 |
| | (2.9133) | (2.8603) | (2.8851) |
| north_east | 0.1424 | 0.1753 | 0.1707 |
| | (0.3495) | (0.3803) | (0.3763) |
| north_central | 0.1895 | 0.2043 | 0.2325 |
| | (0.3919) | (0.4032) | (0.4224) |
| south | 0.3359 | 0.3118 | 0.3531 |
| | (0.4723) | (0.4633) | (0.4779) |
| west | 0.1930 | 0.3040 | 0.2331 |
| | (0.3946) | (0.4601) | (0.4228) |
| Observations | 98824 | 4381 | 55562 |

Standard deviations in parenthesis

(1) - Universe : Respondents not lost due to non-response for alcohol participation (alc_participation_month) and drug participation (drugs_dli)

(2) - Universe : Respondents not lost due to non-response for alcohol participation (alc_participation_month)

**Figure 3: Regression Discontinuity Table of Means**

| | Regression Discontinuity | | | |
|---|---|---|---|---|
| **VARIABLE** | **Overall** | **(1)** | **(2)** | **(3)** |
| white | 0.5062 | 0.5466 | 0.5458 | 0.5460 |
| | (0.5000) | (0.4978) | (0.4979) | (0.4979) |
| black | 0.2345 | 0.2517 | 0.2515 | 0.2522 |
| | (0.4237) | (0.4340) | (0.4339) | (0.4343) |
| hispanic | 0.2116 | 0.2119 | 0.2121 | 0.2118 |
| | (0.4084) | (0.4086) | (0.4088) | (0.4086) |
| female | 0.4881 | 0.4983 | 0.4975 | 0.4982 |
| | (0.4999) | (0.5000) | (0.5000) | (0.5000) |
| dad_hgc_pos | 12.5643 | 12.5887 | 12.5825 | 12.5880 |
| | (3.2117) | (3.2082) | (3.2125) | (3.2075) |
| mom_hgc_pos | 12.4381 | 12.4743 | 12.4713 | 12.4732 |
| | (2.9133) | (2.9347) | (2.9355) | (2.9340) |
| north_east | 0.1424 | 0.1647 | 0.1648 | 0.1648 |
| | (0.3495) | (0.3709) | (0.3710) | (0.3710) |
| north_central | 0.1895 | 0.2196 | 0.2192 | 0.2195 |
| | (0.3919) | (0.4139) | (0.4137) | (0.4139) |
| south | 0.3359 | 0.3870 | 0.3871 | 0.3871 |
| | (0.4723) | (0.4871) | (0.4871) | (0.4871) |
| west | 0.1930 | 0.2237 | 0.2237 | 0.2235 |
| | (0.3946) | (0.4167) | (0.4167) | (0.4166) |
| Observations | 98824 | 83742 | 84478 | 83984 |

Standard deviations in parenthesis

(1) - Universe : Respondents not lost due to non-response for alcohol frequency (alc_freq) and drug frequency (drugs_freq)

(2) - Universe : Respondents not lost due to non-response for alcohol frequency (alc_freq)

(3) - Universe : Respondents not lost due to non-response for drug frequency (drugs_freq)

## Figure 4: OLS and Fixed Effects Balance Test

| | Universe: Respondents who answered yes to consuming VARIABLE | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| VARIABLES | female | age | white | black | hispanic | dad_hgc_pos | mom_hgc_pos | north_east | north_central | south |
| alc_participati | | | | | | | | | | |
| on_month | -0.0131 | 0.0928 | 0.0054 | -0.0032 | -0.0017 | 0.0261 | 0.0278 | -0.0002 | 0.0001 | -0.0003 |
| | (0.0005) | (0.0026) | (0.0006) | (0.0005) | (0.0005) | (0.0043) | (0.0035) | (0.0004) | (0.0005) | (0.0006) |
| Constant | 0.5630 | 20.7300 | 0.5800 | 0.2110 | 0.2120 | 12.6700 | 12.5900 | 0.1720 | 0.2320 | 0.3540 |
| | (0.0066) | (0.0255) | (0.0065) | (0.0053) | (0.0054) | (0.0462) | (0.0399) | (0.0048) | (0.0055) | (0.0061) |
| Observations | 55562 | 55562 | 55562 | 55562 | 55562 | 45518 | 51877 | 55562 | 55562 | 55562 |
| R-squared | 0.0280 | 0.0260 | 0.0050 | 0.0030 | 0.0010 | 0.0030 | 0.0040 | 0.0000 | 0.0000 | 0.0000 |
| F-stat | 572.1000 | 1287.0000 | 90.8700 | 47.4400 | 12.9800 | 36.8800 | 64.4000 | 0.3050 | 0.0240 | 0.2070 |
| p-value | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0003 | 0.0000 | 0.0000 | 0.5810 | 0.8770 | 0.6490 |
| drugs_dli | -0.0002 | -0.0013 | -0.0001 | 0.0000 | -0.0001 | -0.0037 | -0.0017 | -0.0001 | 0.0000 | 0.0002 |
| | (0.0001) | (0.0005) | (0.0001) | (0.0000) | (0.0001) | (0.0007) | (0.0005) | (0.0001) | (0.0001) | (0.0001) |
| Constant | 0.4610 | 20.3200 | 0.7280 | 0.0852 | 0.2170 | 13.0400 | 13.0000 | 0.1770 | 0.2040 | 0.3100 |
| | (0.0142) | (0.0634) | (0.0117) | (0.0062) | (0.0114) | (0.1050) | (0.0852) | (0.0108) | (0.0107) | (0.0125) |
| Observations | 4738 | 4738 | 4738 | 4738 | 4738 | 3880 | 4426 | 4738 | 4738 | 4738 |
| R-squared | 0.0020 | 0.0020 | 0.0000 | 0.0000 | 0.0000 | 0.0130 | 0.0040 | 0.0000 | 0.0000 | 0.0010 |
| F-stat | 6.5910 | 6.6780 | 1.2080 | 0.8220 | 1.7360 | 26.2500 | 12.2400 | 0.7330 | 0.0509 | 3.6560 |
| p-value | 0.0103 | 0.0098 | 0.2720 | 0.3650 | 0.1880 | 0.0000 | 0.0005 | 0.3920 | 0.8220 | 0.0560 |
| | Universe : Respondents who answered yes to consuming Alcohol and VARIABLE | | | | | | | | | |
| drugs_dli | -0.0002 | -0.0016 | -0.0001 | 0.0001 | -0.0001 | -0.0036 | -0.0019 | -0.0001 | 0.0000 | 0.0002 |
| | (0.0001) | (0.0005) | (0.0001) | (0.0001) | (0.0001) | (0.0007) | (0.0005) | (0.0001) | (0.0001) | (0.0001) |
| Constant | 0.4630 | 20.4600 | 0.7410 | 0.0710 | 0.2150 | 13.1000 | 13.0800 | 0.1780 | 0.2060 | 0.3040 |
| | (0.0147) | (0.0642) | (0.0120) | (0.0059) | (0.0118) | (0.1080) | (0.0876) | (0.0112) | (0.0113) | (0.0130) |
| Observations | 4381 | 4381 | 4381 | 4381 | 4381 | 3623 | 4102 | 4381 | 4381 | 4381 |
| R-squared | 0.0020 | 0.0030 | 0.0000 | 0.0010 | 0.0010 | 0.0130 | 0.0040 | 0.0000 | 0.0000 | 0.0020 |
| F-stat | 4.4580 | 10.0700 | 1.2890 | 2.2040 | 2.2650 | 27.4000 | 15.0100 | 0.6050 | 0.3040 | 4.7080 |
| p-value | 0.0349 | 0.0015 | 0.2560 | 0.1380 | 0.1320 | 0.0000 | 0.0001 | 0.4370 | 0.5820 | 0.0301 |

## Figure 5: Regression Discontinuity Balance Test

| | Universe: Respondents who answered yes to consuming VARIABLE | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| VARIABLES | female | age | white | black | hispanic | dad_hgc_pos | mom_hgc_pos | north_east | north_central | south |
| alc_freq | -0.0170 | 1.5730 | 0.1920 | -0.1740 | -0.0226 | 0.7510 | 0.8300 | 0.0211 | 0.0426 | -0.0968 |
| | (0.0077) | (0.0364) | (0.0074) | (0.0070) | (0.0062) | (0.0569) | (0.0474) | (0.0053) | (0.0060) | (0.0073) |
| Constant | 0.5090 | 19.7400 | 0.4170 | 0.3680 | 0.2270 | 12.0700 | 11.9100 | 0.1510 | 0.1910 | 0.4520 |
| | (0.0075) | (0.0321) | (0.0073) | (0.0074) | (0.0063) | (0.0563) | (0.0466) | (0.0051) | (0.0056) | (0.0073) |
| Observations | 84478 | 84478 | 84478 | 84478 | 84478 | 67186 | 78190 | 84478 | 84478 | 84478 |
| R-squared | 0.0000 | 0.0460 | 0.0330 | 0.0360 | 0.0010 | 0.0120 | 0.0180 | 0.0010 | 0.0020 | 0.0090 |
| F-stat | 4.9130 | 1870.0000 | 682.3000 | 622.4000 | 13.1300 | 173.8000 | 306.6000 | 16.1300 | 50.4000 | 174.6000 |
| p-value | 0.0267 | 0.0000 | 0.0000 | 0.0000 | 0.0003 | 0.0000 | 0.0000 | 0.0001 | 0.0000 | 0.0000 |
| drugs_freq | -0.0477 | -0.5220 | 0.1840 | -0.1710 | 0.0024 | 0.2920 | 0.4610 | 0.0098 | -0.0197 | -0.0717 |
| | (0.0134) | (0.0610) | (0.0113) | (0.0069) | (0.0108) | (0.1020) | (0.0802) | (0.0101) | (0.0099) | (0.0119) |
| Constant | 0.5010 | 20.8300 | 0.5350 | 0.2620 | 0.2120 | 12.5700 | 12.4500 | 0.1640 | 0.2210 | 0.3910 |
| | (0.0056) | (0.0172) | (0.0055) | (0.0049) | (0.0045) | (0.0399) | (0.0340) | (0.0040) | (0.0044) | (0.0052) |
| Observations | 83984 | 83984 | 83984 | 83984 | 83984 | 66793 | 77762 | 83984 | 83984 | 83984 |
| R-squared | 0.0000 | 0.0010 | 0.0070 | 0.0080 | 0.0000 | 0.0000 | 0.0010 | 0.0000 | 0.0000 | 0.0010 |
| F-stat | 12.7500 | 73.3000 | 264.4000 | 626.2000 | 0.0493 | 8.1720 | 33.0100 | 0.9520 | 3.9400 | 36.2000 |
| p-value | 0.0004 | 0.0000 | 0.0000 | 0.0000 | 0.8240 | 0.0043 | 0.0000 | 0.3290 | 0.0472 | 0.0000 |
| | Universe : Respondents who answered yes to consuming Alcohol and VARIABLE | | | | | | | | | |
| drugs_freq | -0.0479 | -0.5180 | 0.1850 | -0.1710 | 0.0024 | 0.2930 | 0.4610 | 0.0098 | -0.0195 | -0.0721 |
| | (0.0134) | (0.0610) | (0.0113) | (0.0068) | (0.0108) | (0.1020) | (0.0803) | (0.0101) | (0.0099) | (0.0119) |
| Constant | 0.5010 | 20.8300 | 0.5360 | 0.2620 | 0.2120 | 12.5700 | 12.4500 | 0.1640 | 0.2210 | 0.3910 |
| | (0.0056) | (0.0172) | (0.0055) | (0.0049) | (0.0045) | (0.0399) | (0.0340) | (0.0040) | (0.0044) | (0.0052) |
| Observations | 83742 | 83742 | 83742 | 83742 | 83742 | 66632 | 77544 | 83742 | 83742 | 83742 |
| R-squared | 0.0000 | 0.0010 | 0.0070 | 0.0080 | 0.0000 | 0.0000 | 0.0010 | 0.0000 | 0.0000 | 0.0010 |
| F-stat | 12.8200 | 72.2100 | 265.5000 | 629.9000 | 0.0481 | 8.2750 | 32.9600 | 0.9410 | 3.8620 | 36.5300 |
| p-value | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.8260 | 0.0040 | 0.0000 | 0.3320 | 0.0494 | 0.0000 |

## Figure 6: OLS Participation Regression

| VARIABLES | (1) drugs_dli | (2) drugs_dli | (3) drugs_dli | (4) drugs_dli |
|---|---|---|---|---|
| alc_participation_month | -0.00456 | 0.392 | 0.34 | 0.416 |
| | (0.2370) | (0.2470) | (0.2470) | (0.2750) |
| female | | | -8.578** | -5.658 |
| | | | (3.9220) | (4.1560) |
| white | | | -4.079 | -2.419 |
| | | | (5.6840) | (6.5170) |
| black | | | 8.381 | 1.338 |
| | | | (9.1210) | (10.8300) |
| hispanic | | | -7.44 | -14.28** |
| | | | (4.7140) | (6.1480) |
| dad_hgc_pos | | | | -3.829*** |
| | | | | (1.0420) |
| mom_hgc_pos | | | | 0.502 |
| | | | | (0.9850) |
| north_east | | | | -2.581 |
| | | | | (5.9100) |
| north_central | | | | -0.549 |
| | | | | (5.7480) |
| south | | | | 8.056 |
| | | | | (5.3160) |
| Year Effects | N | Y | Y | Y |
| Cluster (id) | Y | Y | Y | Y |
| Constant | 43.25*** | 33.89*** | 41.95*** | 97.38*** |
| | (2.7480) | (6.0720) | (8.1120) | (16.2100) |
| Observations | 4,381 | 4,381 | 4,381 | 3,539 |
| R-squared | 0 | 0.035 | 0.039 | 0.059 |
| F-stat | 0.000369 | 2.514 | 1.892 | 2.297 |
| p-value | 0.985 | 0.113 | 0.169 | 0.13 |

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

**Figure 7: OLS Alcohol & Drug Frequency Regression**

| VARIABLES | (1) alc_freq | (2) drugs_freq |
|---|---|---|
| over_21 | 0.122*** | -0.00819* |
| | (0.0082) | (0.0043) |
| female | 0.0023 | -0.00794** |
| | (0.0070) | (0.0033) |
| white | 0.0325*** | 0.0162*** |
| | (0.0105) | (0.0048) |
| black | -0.168*** | -0.0347*** |
| | (0.0134) | (0.0050) |
| hispanic | -0.0338*** | -0.00865 |
| | (0.0113) | (0.0055) |
| dad_hgc_pos | 0.00209 | -0.00150** |
| | (0.0015) | (0.0007) |
| mom_hgc_pos | 0.0153*** | 0.00390*** |
| | (0.0016) | (0.0007) |
| north_east | 0.0072 | -0.0154** |
| | (0.0108) | (0.0061) |
| north_central | 0.00327 | -0.0310*** |
| | (0.0103) | (0.0053) |
| south | -0.0433*** | -0.0232*** |
| | (0.0097) | (0.0050) |
| Constant | 0.479*** | 0.0428*** |
| | (0.0243) | (0.0118) |
| Year Effects | Y | Y |
| Observations | 65,184 | 64,800 |
| R-squared | 0.099 | 0.014 |
| F-stat | 223.9 | 3.685 |
| p-value | 0 | 0.0549 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Figure 8: Fixed Effects Regression**

| VARIABLES | (1) drugs_dli | (2) drugs_dli | (3) drugs_dli |
|---|---|---|---|
| alc_participation_month | -0.313 | 0.139 | 0.419 |
| | (0.4290) | (0.4390) | (0.2740) |
| female | | | -5.701 |
| | | | (4.1550) |
| white | | | -2.384 |
| | | | (6.5160) |
| black | | | 1.177 |
| | | | (10.8500) |
| hispanic | | | -14.13** |
| | | | (6.1460) |
| dad_hgc_pos | | | -3.828*** |
| | | | (1.0420) |
| mom_hgc_pos | | | 0.504 |
| | | | (0.9840) |
| Constant | 45.89*** | 41.46*** | 83.84*** |
| | (3.6650) | (9.7680) | (14.5000) |
| Year Effects | N | Y | Y |
| Fixed Effects | id | id | region |
| Cluster (id) | Y | Y | Y |
| Observations | 4,381 | 4,381 | 3,539 |
| R-squared | 0.59 | 0.618 | 0.06 |
| F-stat | 0.533 | 0.101 | 2.333 |
| p-value | 0.465 | 0.751 | 0.127 |

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

**Figure 9: Regression Discontinuity Table**

| VARIABLES | (1) drug_freq | (2) alc_freq |
|---|---|---|
| over_21 = Di | -0.0667*** | 0.0517*** |
| | (0.0010) | (0.0002) |
| age_pos | 6.343*** | -0.0394* |
| | (0.1320) | (0.0227) |
| age2 | -0.390*** | 0.0193*** |
| | (0.0101) | (0.0016) |
| age3 | 0.0104*** | -0.000967*** |
| | (0.0003) | (0.0001) |
| age4 | -0.000103*** | 1.42e-05*** |
| | (0.0000) | (0.0000) |
| age x Di | -3.58E-05 | 1.16e-05*** |
| | (0.0000) | (0.0000) |
| age2 x Di | 2.83e-05** | -7.00E-07 |
| | (0.0000) | (0.0000) |
| age3 x Di | -2.53e-06** | 7.47E-08 |
| | (0.0000) | (0.0000) |
| age4 x Di | 5.62e-08* | -1.90E-09 |
| | (0.0000) | (0.0000) |
| female | -0.000227 | 1.31E-05 |
| | (0.0003) | (0.0000) |
| white | -0.000806* | -0.000133** |
| | (0.0004) | (0.0001) |
| black | -0.000498 | -5.78E-05 |
| | (0.0005) | (0.0001) |
| hispanic | 0.000113 | -3.27E-05 |
| | (0.0005) | (0.0001) |
| dad_hgc_pos | -4.66E-05 | 3.18E-06 |
| | (0.0001) | (0.0000) |
| mom_hgc_pos | 3.69E-05 | 5.39E-06 |
| | (0.0001) | (0.0000) |
| north_east | 0.000132 | 1.73E-05 |
| | (0.0005) | (0.0001) |
| north_central | 0.000382 | 5.38E-05 |
| | (0.0005) | (0.0001) |
| south | 0.000789* | -9.08E-07 |
| | (0.0004) | (0.0001) |
| Year Effects | Y | Y |
| Cluster (id) | Y | Y |
| Constant | -36.26*** | -0.793*** |
| | (0.6380) | (0.1170) |
| Observations | 64,646 | 64,646 |
| Clusters | 6780 | 6780 |
| R-squared | 0.932 | 0.996 |
| F-stat | 4783 | 67658 |
| p-value | 0 | 0 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Figure 10: Alcohol Frequency RD**

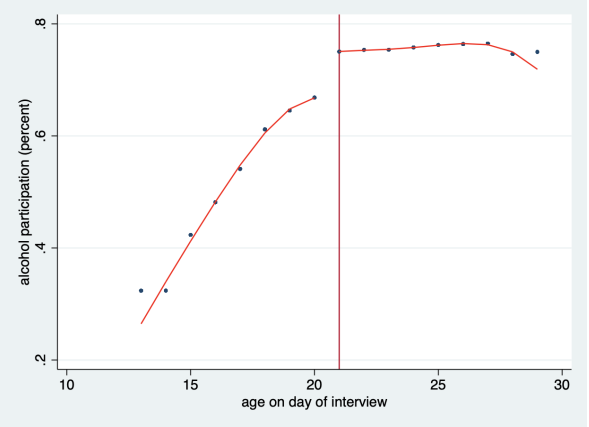



**Figure 11: Hard Drug Frequency RD**

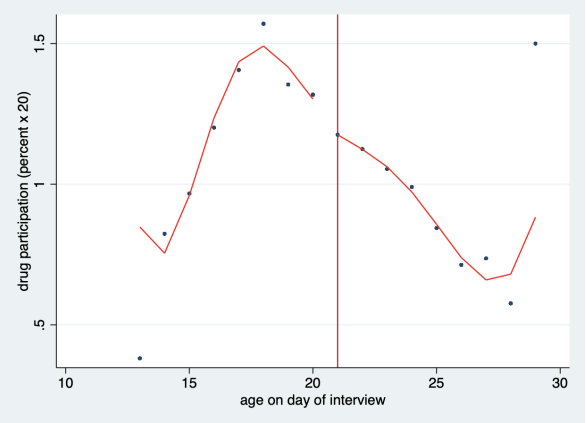



**Figure 12: Non-RD Hard Drug Frequency Model**

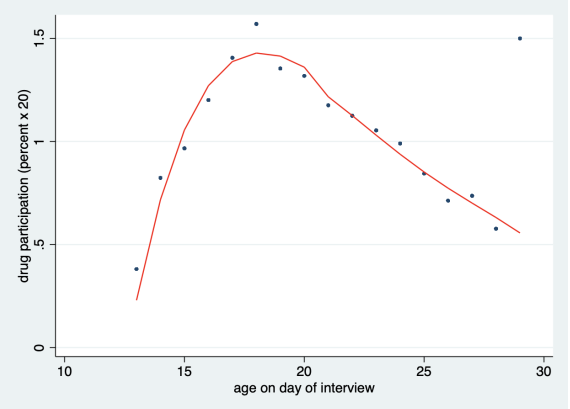
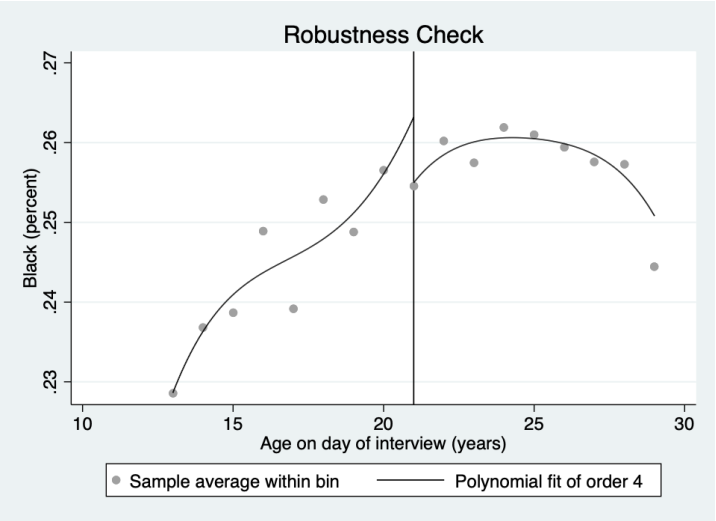
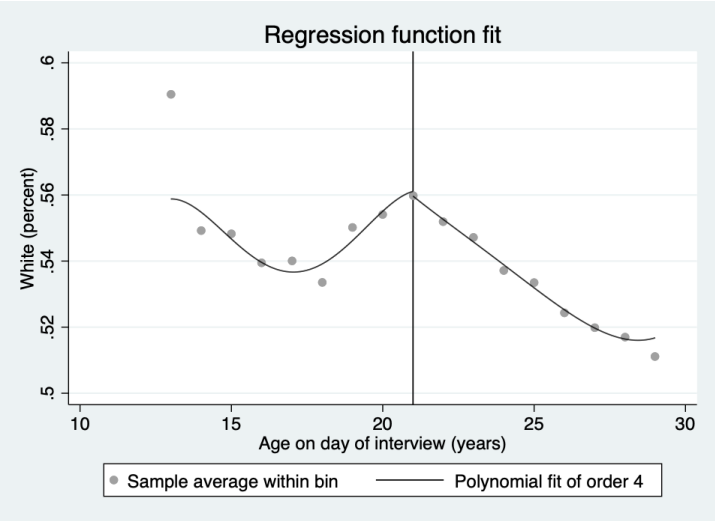


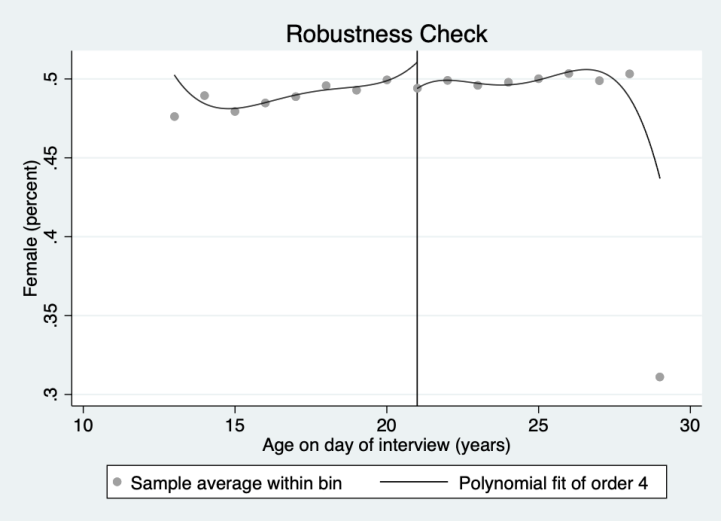
**Figure 13: Regression Discontinuity Robustness**

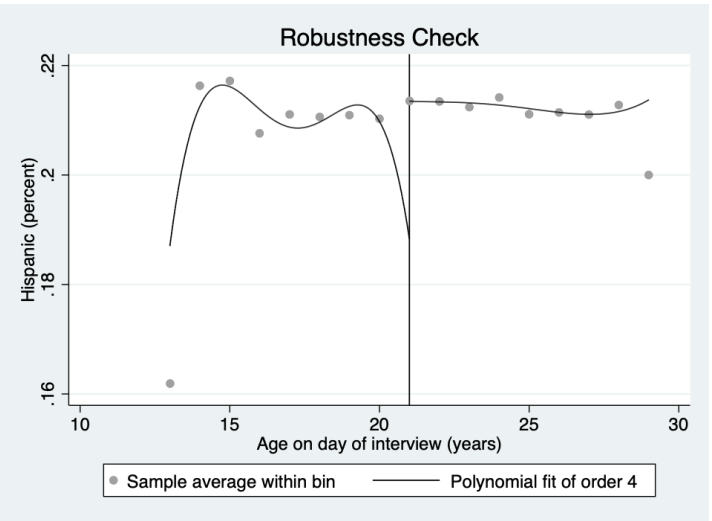
| VARIABLES | (1) drugs_freq | | | (2) alc_freq | | |
|---|---|---|---|---|---|---|
| | Age 20 | Age 21 | Age 22 | Age 20 | Age 21 | Age 22 |
| Di | -0.114*** | -0.0667*** | 0.0836*** | -0.00412*** | 0.0517*** | -0.00739*** |
| | (0.0010) | (0.0010) | (0.0015) | (0.0002) | (0.0002) | (0.0002) |
| Constant | -27.62*** | -36.26*** | -37.99*** | 2.225*** | -0.793*** | 1.651*** |
| | (0.6790) | (0.6380) | (0.6710) | (0.0833) | (0.1170) | (0.0666) |
| Controls | Y | Y | Y | Y | Y | Y |
| Year Effects | Y | Y | Y | Y | Y | Y |
| Cluster (id) | Y | Y | Y | Y | Y | Y |
| age-poly interaction | Y | Y | Y | Y | Y | Y |
| Observations | 64,646 | 64,646 | 64,646 | 64,646 | 64,646 | 64646 |
| R-squared | 0.937 | 0.932 | 0.934 | 0.988 | 0.996 | 0.988 |
| F-stat | 11940 | 4783 | 2939 | 275.7 | 67658 | 1993 |
| p-value | 0 | 0 | 0 | 0 | 0 | 0 |

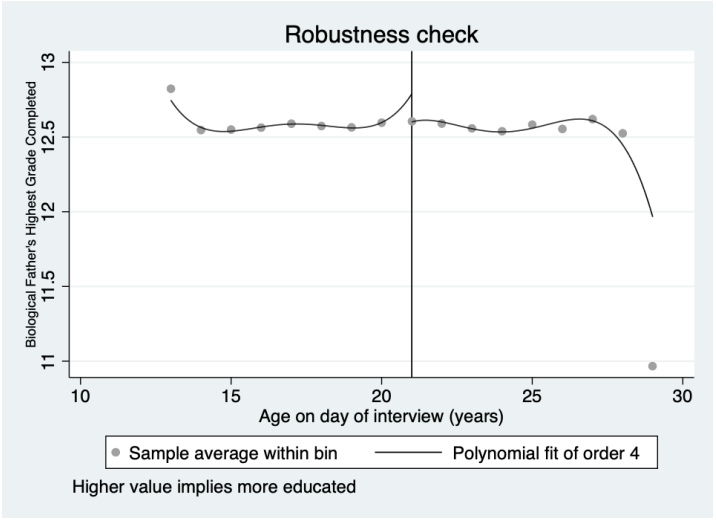
**Figures 14 - 19 : Smooth Transition of Demographics & Parents' Education**



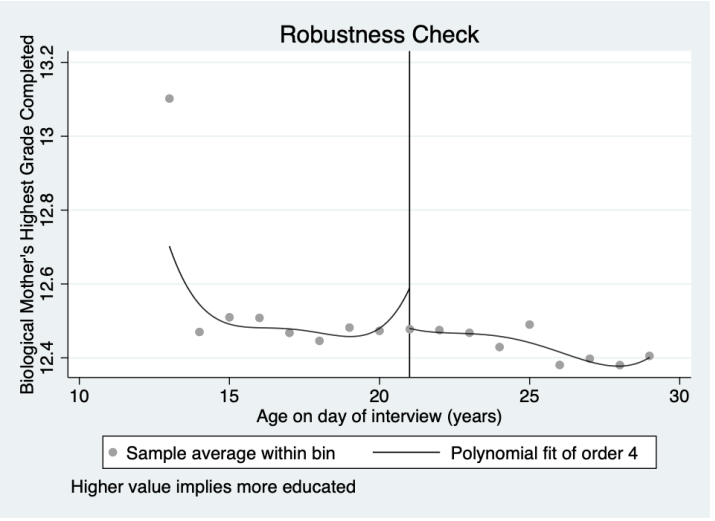



**Figure 20 : OLS regression alc_participation_month on over_21**

| VARIABLES | (1)<br>alc_participation_month |
|---|---|
| over_21 | 1.567*** |
| | (0.135) |
| female | -2.216*** |
| | (0.109) |
| white | 0.473*** |
| | (0.150) |
| black | -0.505*** |
| | (0.193) |
| hispanic | -0.159 |
| | (0.163) |
| dad_hgc_pos | 0.0246 |
| | (0.0222) |
| mom_hgc_pos | 0.107*** |
| | (0.0233) |
| north_east | 0.00328 |
| | (0.166) |
| north_central | -0.255* |
| | (0.154) |
| south | -0.0474 |
| | (0.146) |
| Constant | 1.113*** |
| | (0.333) |
| Year Effects | Y |
| Cluster (id) | Y |
| Observations | 44,370 |
| R-squared | 0.079 |
| F-stat | 134 |
| p-value | 0 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1