To: Director Christopher J. Maghi

From: Kimberly M. Kreiss

Subject: The Minimum Legal Drinking Age

Summary

This document provides analysis around whether the minimum legal drinking age in the United States makes any difference in the likelihood of consuming alcohol. This analysis was prepared in preparation for next month's conference. I use data from the National Health Interview Survey to analyze if there is a jump in the likelihood of drinking after turning 21. This data also allows me to provide analysis to see if these effects differ by demographic groups, by level of drinking, and if the relationships are non-linear. Though it's important to keep in mind that survey data is often subject to measurement error–people often misreport outcomes like amount of alcohol consumer–this dataset provides important information for considering alcohol use.

My results suggest that the legal minimum drinking age has some effect on the likelihood of consuming alcohol, but the causal effect of these results should be interpreted with caution due to limitations of the dataset and research design. These effects are most pronounced among men. I go into these facets of the analysis in detail below.

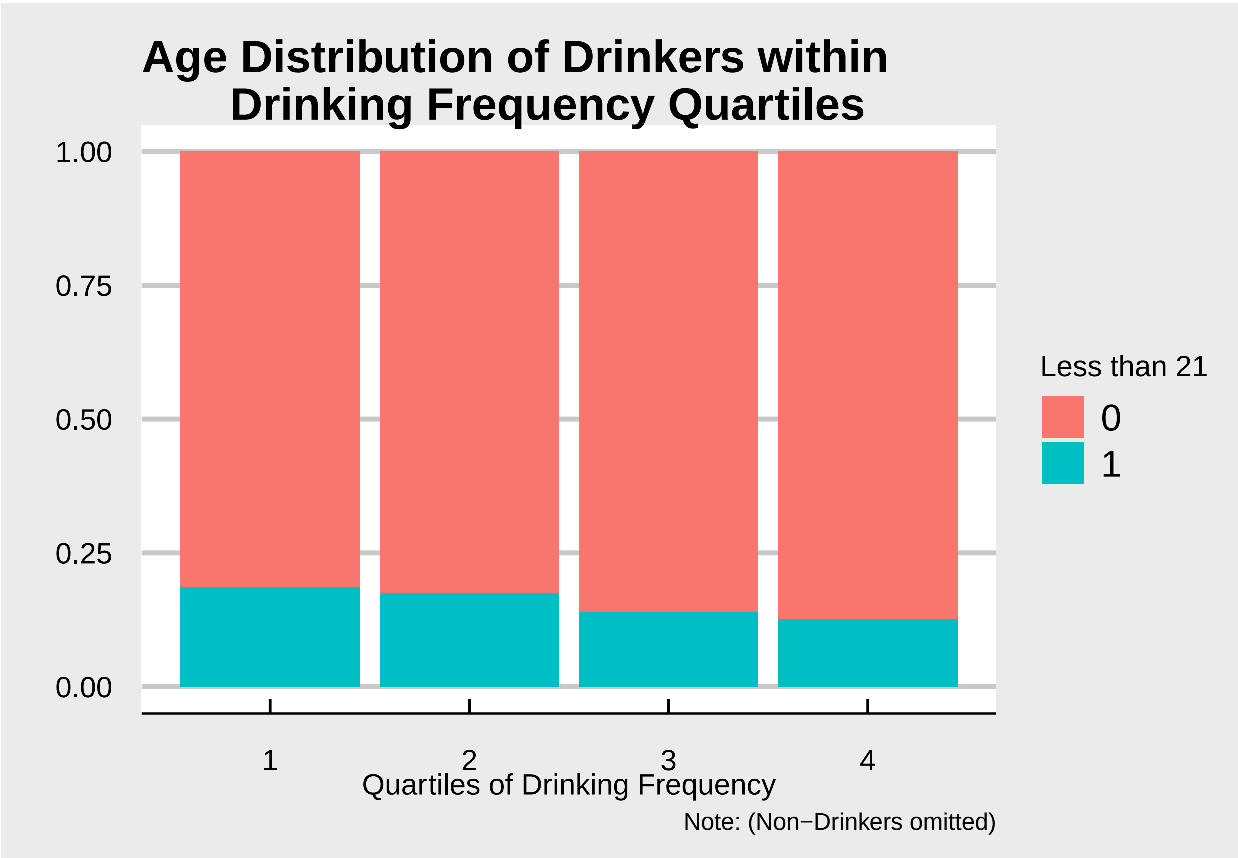
Descriptive Statistics

The data come from the National Health Interview Survey and provide information on the respondents age, sex, race,education level, employment status, marriage status, and drinking habits. It reports information about whether the respondent self-reports drinking alcohol as well as the percentage of days that a respondent drinks alcohol. The data show that most people (62.5%) report drinking alcohol–Table 1 in the appendix provides more information about the composition of the sample. It is important to note that there is a sizable share of the sample that reports not drinking at all–important outliers in this analysis. Table 2 provides more information about drinking statistics, namely information about years from the minimum legal drinking age as well as drinking frequency.

Table 1: Drinking Patterns

Names	Mean	Standard Deviation	Median	Min	Max
Years From 21	3.4	3.4	3.6	-3.3	9.9
Percent Days Drank	9.9	17.3	2.7	0.0	100.0

Moreover, Table 1 shows that the average percent of days that a respondent drank was about 10%. Still, there is a large standard deviation around this number, suggesting a fairly wide spread in the share of days drinking. Moreover, the median is just 2.7%, suggesting that the data are left skewed. Next, the figure below shows the share of drinkers within each frequency quartile that are less than 21. Underage drinkers are represented among all quartiles, but have a slightly hire representation among those in the first quartile. This doesn't suggest that underage drinkers are more likely to drink heavily than non-underage drinkers.



I turn next to regression analysis and make use of a linear probability model to estimate how age affects the likelihood of drinking alcohol. I start with a simple regression which is shown in column 1 in the regression table below. I start by regressing a binary variable that indicates whether the respondent is below age 21 on a binary outcome variable for if the respondent consumes alcohol. Indeed, this regression suggests a large negative effect on the likelihood of drinking if a respondent is under the drinking age. However, it's plausible that there is a non-linear relationship between age under 21 and whether the respondent drinks. For example, it's likely that a person who is 10 years old will have a much lower likelihood of drinking than someone who is 20.5 years old. This may be driving the results. As a result, the next two columns in the table show different specifications, which have age (in years) as the dependent variable and a quadratic to allow for the relationship with age to be non-linear. Indeed, when adding these in, it appears that an additional year in age increases the likelihood that a respondent drinks. So far, these results are consistent with the minimum legal drinking age having a negative effect on the likelihood that a person drinks. The quadratic term shows that the relationship is non-linear: in this specification, the effect of age is diminishing. That is, a one-year increase for older people has a less strong effect than a one-year increase for younger people on the likelihood of drinking. I estimated this with a cubic term, and the results suggest overfitting, suggesting that a quadratic is the best to use. The output is included in the appendix. It is important to note, however, that the R-squared's and adjusted R-squared's for these regressions are all very tiny, suggesting that these independent variables do not explain a significant amount of the variation of the dependent variable.

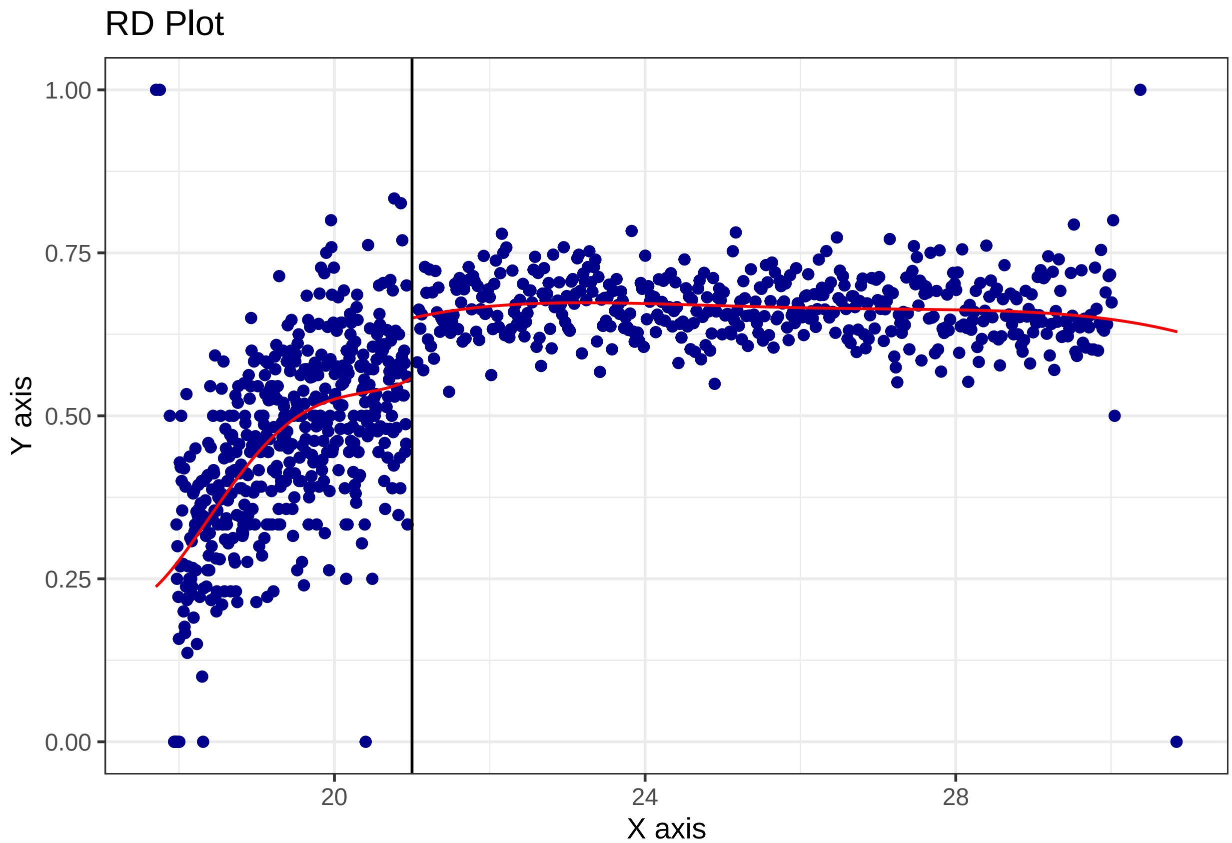
Table 2: Effect of Age on Probability of Drinking Alcohol

	Dependent variable:		
	Probability of Drinking		
	(1)	(2)	(3)
lt_21	-0.196^{***}		
	(0.005)		
age		0.017^{***}	0.274^{***}
		(0.001)	(0.009)
age_sq			-0.005^{***}
			(0.0002)
Constant	0.665^{***}	0.200^{***}	-2.845^{***}
	(0.002)	(0.014)	(0.105)
Observations	61,263	61,263	61,263
R^2	0.027	0.015	0.028
Adjusted R^2	0.027	0.015	0.028
Residual Std. Error	0.478 (df = 61261)	0.481 (df = 61261)	0.477 (df = 61260)
F Statistic	$1,678.377^{***}$ (df = 1; 61261)	927.281^{***} (df = 1; 61261)	894.372^{***} (df = 2; 61260)
Note:			*p<0.1; **p<0.05; ***p<0.01

For my final specification, I chose a model where the dependent variables include age, an indicator variable for being less than the legal drinking age, and interaction terms for age and the minimum legal drinking age. I also include polynomials to allow for non-linearity. The results from this specification are presented below in table 2. This specification allows us to control for the effect of age and to allow the effect of age to be non-linear. Moreover, it allows us to control for the effect of the minimum legal drinking age, and the interaction terms allows us to capture that there is a separate relationship for the effect of age on the likelihood of drinking before age 21 and it is non-linear. Indeed, I believe that this specification is the best option, and as such use it going forward. Indeed, this specification suggests that there is a small effect of age on the likelihood of drinking and that it is non-linear. In particular, the regression tells us that the effect of an increase in age for an older person on the probability of drinking is smaller than the effect for a younger person. Moreover, the specification suggests that the minimum legal drinking age has a negative effect on the likelihood of drinking, but has less of an effect for those just before the cutoff. As an example, I can now provide some estimates of how the likelihood of drinking changes as a person ages from 19.5 to 20 and from 20 to 20.5. The former shows a change of .0346 and the latter shows a change of .0153.

Before adding in control variables and considering other specifications, I make use of a regression discontinuity design. A regression discontinuity design allows us to see if there is a big jump in the likelihood of drinking in the period immediately after someone turns 21 and the period immediately before. The graph below shows the results of the regression discontinuity analysis. The visual shows results that are consistent with the previous analysis. Indeed, it seems that there is a jump in the likelihood of drinking after age 21, with each year after that having little effect on the probability of drinking alcohol. Indeed, this effect is estimated to be roughly an 8% increase in the likelihood.

```
## [1] "Mass points detected in the running variable."
```



Finally, I include a small number of demographic controls. Indeed, when including these controls, two important implications emerge. First, being male has a large, positive and statistically significant effect on the likelihood of drinking alcohol, as does being employed. Being married has a large, negative, and statistically negative effect on the likelihood of drinking. Finally, being a student, surprisingly, has little effect on the likelihood of drinking alcohol. Estimates from this specification are available in the appendix.

Finally, I analyze how the minimum legal drinking age affects the probability of drinking alcohol for different types of drinkers. I consider three groups: a low drinking group, for people who drink 1 day or less of the week; a moderate drinking group, for people who drink 2 to 4 days of the week; and a heavy drinking group, for those who drink 5 days or more a week. Overall, I find the highest effect of age on those who are low drinkers followed by moderate drinkers. The specification results are included in the appendix. Indeed, this makes sense: heavy drinkers are likely problem drinkers and people who won't be deterred by any type of policy intervention.

Overall, my results suggest that the minimum legal drinking age has some effect in lowering the likelihood that someone consumes alcohol. Still, caution should be employed when interpreting these results, especially as causal effects. First, there are some limitations to the data, including that people may under report whether they consume alcohol. Moreover, the control variables are limited and the R-squared's and adjusted R-squared's are low for all the specifications, suggesting that there is substnatial omitted variable bias. Alcohol consumption is correlated with many regional and cultural variations as well that we are unable to control for in our data. Moreover, an issue with an regression discontinuity design is that it inherently sorts people into two groups. In this case, people are divided into groups of those who can and cannot legally consume alcohol. This policy has been in place for some time; indeed it's plausible that drinking alcohol after 21 is just an accepted social norm, whereas drinking it before 21 is not. Thus, it is not clear that lowering the legal drinking age to a number like 20 would have any meaningful effect on the effect of the minimum legal drinking age.

Appendix

Table 3: Sample Demographics

Demographic	Percent
hs_diploma	80.6
hispanic	24.3
white	55.1
black	15.2
employed	71.2
married	29.4
male	44.5
drinks_alcohol	62.5
student	10.0

Table 4:

	Dependent variable:
	Probability of Drinking with Cubic Terms
age	1.697^{***}
	(0.104)
age_sq	-0.065^{***}
	(0.004)
age_cu	0.001^{***}
	(0.0001)
Constant	-13.996^{***}
	(0.817)
Observations	61,263
R^2	0.031
Adjusted R^2	0.031
Residual Std. Error	0.476 (df = 61259)
F Statistic	661.294^{***} (df = 3; 61259)
Note:	*p<0.1; **p<0.05; ***p<0.01

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Thu, Dec 15, 2022 - 13:57:38

Table 5:

	Dependent variable:
	Probability of Drinking with Demographic Controls
age	0.250***
	(0.009)
age_sq	−0.005***
	(0.0002)
hs_diploma	0.083***
	(0.005)
hispanic	−0.148***
	(0.005)
black	−0.216***
	(0.005)
employed	0.115***
	(0.005)
student	0.026***
	(0.008)
male	0.142***
	(0.004)
married	−0.121***
	(0.004)
Constant	−2.698***
	(0.102)
Observations	61,263
R^2	0.131
Adjusted R^2	0.131
Residual Std. Error	0.451 (df = 61253)
F Statistic	1,029.146*** (df = 9; 61253)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6:

	Dependent variable:		
	Probability of Drinking by drinker Frequency	mod_drinker	low_drinker
	(1)	(2)	(3)
age	0.008*	0.050***	0.116***
	(0.004)	(0.010)	(0.015)
age_sq	−0.0001	−0.001***	−0.002***
	(0.0001)	(0.0002)	(0.0003)
hs_diploma	−0.003**	0.039***	0.066***
	(0.002)	(0.004)	(0.005)
hispanic	−0.013***	−0.069***	−0.056***
	(0.002)	(0.003)	(0.005)
black	−0.008***	−0.075***	−0.104***
	(0.002)	(0.004)	(0.006)
employed	−0.0001	0.020***	0.110***
	(0.002)	(0.004)	(0.005)
student	−0.010***	0.024***	0.023***
	(0.003)	(0.006)	(0.008)
male	0.027***	0.103***	−0.029***
	(0.001)	(0.003)	(0.004)
married	−0.012***	−0.085***	−0.013***
	(0.001)	(0.003)	(0.005)
under21	−0.005	−0.059***	0.003
	(0.003)	(0.007)	(0.011)
Constant	−0.088	−0.502***	−1.107***
	(0.057)	(0.125)	(0.184)
Observations	61,263	61,263	61,263
R^2	0.013	0.063	0.032
Adjusted R^2	0.013	0.063	0.032
Residual Std. Error (df = 61252)	0.151	0.335	0.492
F Statistic (df = 10; 61252)	79.692***	410.578***	203.230***
Note:			*p<0.1; **p<0.05; ***p<0.01