Differences in rates of alcohol, marijuana, and cigarette use among rural and urban high school students

Eric Daigle - 94056389
Vancouver School of Economics
University of British Columbia

ECON 370: Introduction to Applied Economics

Dr. Jonathan Graves

December 8, 2021

1 Introduction and Background

Drug and alcohol use during high school is associated with a myriad of negative impacts on students including increased dropout rate (Townsend et al., 2007), more frequent involvement in bullying (Radliff et al., 2012), and increased likelihood of engaging in risky sexual activity (McAloney, 2015). Research has also identified links between high school substance use and negative outcomes later in life, with a 2006 study by Ringel et al finding that subjects who used drugs in high school had poorer career outcomes at age 29 (nearly a decade after graduation). Given these findings, it is unsurprising that governments and school districts have spent significant resources on preventing student substance use: total expenditures on school drug prevention programs in the United States in 2005 totaled over $2 billion (The National Center on Addiction and Substance Abuse at Columbia University, 2009). These expenditures, however, seem to be of limited effectiveness – a 2001 CASA report described these primarily classroom-based interventions as "of inherently limited value" (p. 5), and a further meta-analysis (West & O'Neal, 2004) concluded that Project DARE, the largest and most expensive school anti-drug program deployed in the United States, was completely ineffective. While a variety of explanations for these apparent failures have been proposed, one possibility is that programs have targeted the wrong schools – while substance issues are stereotypically associated with schools in poor urban neighbourhoods, recent research has identified rural addiction as a neglected issue in public health (Pettigrew et al., 2012). In this paper, I examine frequency of alcohol, tobacco, and cannabis use among students attending rural and urban high schools using data from the 2018-19 Canadian Student Tobacco, Alcohol, and Drugs Survey. Using a linear regression model developed in R, I investigate the three-part hypothesis that attending a rural school causes higher alcohol use (Part 1), cannabis use (Part 2), and cigarette use (Part 3).

2 Data

The dataset used to investigate this hypothesis is the 2018-19 Canadian Student Tobacco, Alcohol, and Drugs Survey, a Health Canada survey which collects data from Canadian high school students about their substance use. Dropping responses with invalid or missing variables yields a sample size of 5322 observations. My analysis considers six variables in total: the outcome (dependent) dataset variables ALC_040, CAN_040, and SS_060 describe how frequently an observed student used alcohol, cannabis, or cigarettes respectively in the 30 days before taking the questionnaire. All are ordered categorical variables ranking substance use frequency on a scale from a lowest of 1 to highest of 7 (ALC_040 and CAN_040) or 8 (SS_060), and are treated as continuous within this analysis. The treatment (independent) variable $RURAL$ is a transformed dummy set to 1 when the dataset variable $DVURBAN$ indicated a student attended a rural school and 0 when an urban school was indicated. The control (independent) variable IMM is a transformed dummy set to 1 when the dataset variable $DVRES$ indicated a student had lived 10 or fewer years in Canada and 0 otherwise. The control variable $LOWINC$ is a transformed dummy set to 1 when the dataset variable $DVHHINC2$ indicated a student's school was located in an area with median family income at least one standard deviation below the dataset average, and 0 otherwise. A table summarizing key statistics of the six analyzed variables is presented below.

Table 1: Statistical Summary of Relevant Variables

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
ALC_040	5,322	2.327	1.117	1	2	3	6
CAN_040	5,322	2.753	1.803	1	1	4	6
SS_060	5,322	3.111	2.497	1	1	5	8
RURAL	5,322	0.323	0.468	0	0	1	1
IMM	5,322	0.057	0.232	0	0	0	1
LOWINC	5,322	0.495	0.500	0	0	1	1

3 Methodology

I investigate my hypothesis using three parallel linear regression models in R. Model 1 predicts frequency of alcohol use ALC_040 from the treatment variable $RURAL$ and the control variables $LOWINC$ and IMM. Models 2 and 3 predict frequency of cannabis use CAN_040 and frequency of cigarette use SS_060 from these same treatment and control variables. Regression was chosen as the most suitable analysis technique due to the structure of the data - there is no obvious time scale or other progression over which to build a difference-in-difference model, and the few non-substance-related variables present in the dataset make finding good instruments unlikely.

The 3 models share the following specification:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_{1i} + \beta_3 X_{2i} + \epsilon_i$$

Where Y_i is the frequency of alcohol use ALC_040 in Model 1, the frequency of cannabis use CAN_040 in Model 2, and the frequency of cigarette use SS_040 in Model 3. Each model shares the same dependent variables: D_i is the dummy treatment variable $RURAL$, so β_1 is the effect on Y_i of living in a rural area. X_{1i} and X_{2i} are the dummy control variables $LOWINC$ and IMM respectively, so β_2 and β_3 are respectively the effects on Y_i of attending school in a low-income area and being born outside Canada. β_0 is the intercept and ϵ_i is the residual term.

Attending school in a low-income area and being recent immigrants were chosen as covariates as failing to control for them likely would have yielded different potential outcomes between the rural and urban groups. Students in the rural sample are less likely to live in a low-income area and less likely to be recent immigrants, and both these variables likely have a relationship with substance use outcomes - previous research has identified significantly differing patterns of substance use across different income brackets (Patrick et al., 2012), and significantly lower substance use among immigrants (Salas-Wright et al., 2018). While the dataset contains other variables, such as gender and grade level, which have relationships with substance use, these are not included in the models as they are present in similar proportions in the rural and urban groups and thus unlikely to cause bias. More detailed information on demographic makeup across the urban and rural groups is presented in the tables below.

Table 2: Covariate Means of Rural Sample

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
LOWINC	1,719	0.058	0.234	0	0	0	1
IMM	1,719	0.046	0.209	0	0	0	1
SEX	1,719	1.549	0.498	1	1	2	2
GRADE	1,719	10.446	1.378	7	10	12	12

Table 3: Covariate Means of Urban Sample

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
LOWINC	3,603	0.224	0.417	0	0	0	1
IMM	3,603	0.062	0.242	0	0	0	1
SEX	3,603	1.517	0.500	1	1	2	2
GRADE	3,603	10.407	1.302	7	10	11	12

To be able to interpret these models as causal, three assumptions must hold: the treated and untreated (rural and urban) groups must have the same potential alcohol, cannabis, and tobacco use outcomes after controlling for covariates X_{1i} (being in a low-income school district) and X_{2i} (being born outside Canada). The treatment effect must be constant across all values of X_{1i} and X_{2i}, and the models must not omit any variables related to their outcome variable. I discuss these assumptions further later in the paper.

4 Results

The table below gives an overview of coefficients and statistics for Models 1, 2, and 3 in the respectively numbered columns.

Table 4:

	Dependent variable:		
	ALC_040	CAN_040	SS_060
	(1)	(2)	(3)
RURAL	0.024	-0.106^{**}	0.216^{***}
	(0.033)	(0.054)	(0.075)
LOWINC	-0.087^{**}	-0.271^{***}	0.380^{***}
	(0.042)	(0.067)	(0.093)
IMM	-0.034	-0.185^{*}	-0.070
	(0.066)	(0.106)	(0.147)
Constant	2.336^{***}	2.844^{***}	2.981^{***}
	(0.021)	(0.034)	(0.047)
Observations	5,322	5,322	5,322
R^2	0.001	0.004	0.004
Adjusted R^2	0.001	0.003	0.003
Residual Std. Error (df = 5318)	1.116	1.800	2.493
F Statistic (df = 3; 5318)	2.038	6.933^{***}	7.128^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01		

These results show statistically significant correlations between attending a rural school and cigarette and cannabis use. Attending a rural rather than an urban school is associated with an 0.216 ($p < 0.01$) unit increase in cigarette use level, and a 0.106 ($p < 0.05$) decrease in cannabis use level. No statistically significant correlation is found between attending a rural school and alcohol use level. Turning attention to control variables, statistically significant correlations are found between low income and use of all 3 substances: living in an area with below-average median household income is associated with an 0.087 unit decrease ($p < 0.05$) and an 0.271 unit decrease ($p < 0.01$) in alcohol and cannabis use respectively, while cigarette use is found to increase by 0.380 units ($p < 0.01$). The only statistically significant relationship involving recent immigration is an 0.185 unit decrease ($p < 0.1$) in cannabis use.

5 Discussion and Extension

Plugging these regression results (the outcome variables, the treatment effect β_1, the covariate effects X_{1i}, X_{2i}, and the constant terms) and our variables into the model specifications defined above, we are left with the following equations for Models 1, 2, and 3 respectively:

$$ALC\_040 = 2.336 + 0.024RURAL - 0.087LOWINC - 0.034IMM + 1.116 \tag{1}$$

$$CAN\_040 = 2.844 - 0.106RURAL - 0.271LOWINC - 0.185IMM + 1.800 \tag{2}$$

$$SS\_060 = 2.981 + 0.216RURAL + 0.380LOWINC - 0.070IMM + 2.493 \tag{3}$$

These models seem to indicate differing results with regard to our multi-part hypothesis. Part 2 seems to be rejected, with Model 2 indicating that attending a rural school decreases cannabis use rates by 0.106 units. Part 3 is supported, with Model 3 showing that attending a rural school increases cigarette use frequency by 0.216 units. As shown above, Model 1 finds no statistically significant relationship between alcohol use and attending a rural school, leaving us unable to make any conclusions about Part 1. Moving beyond our hypothesis, we also find interesting relationships between our covariates and outcome variables: attending school in a low-income area significantly decreases alcohol and cannabis use frequency, which sharply increasing cigarette use (an 0.38 unit jump).

As previously explained, interpreting these results causally (claiming that the differences in substance use frequency are caused by, rather than simply associated with, rural school attendance) requires a series of assumptions to be met. I believe these models meet the conditional independence assumption: as covariates which occur at significantly different rates (recent immigration and low-income location) between the urban and rural groups were controlled for, it is unlikely that there is a relationship between rural vs. urban status and substance use potential outcomes. The treatment effect heterogeneity assumption also likely holds, as it is hard to imagine that *the effect on one's substance use of living in a rural vs. urban area* (rather than one's substance use level itelf) is affected by immigration status or income level. It is more challenging to argue that no relevant variables were omitted from the models: this dataset contains information on very few topics other than substance use, and it is likely that variables with a relationship to substance use frequency are missing from the dataset and thus the model. In particular, previous research has identified boredom as a risk factor for increased substance use among adolescents (Biolcati et al., 2017) - as this variable is not included in the models, its effect may be being attributed to other variables which were included, introducing bias and representing a significant limitation to our analysis.

Further analysis without these limitations could be performed using a more extensive dataset. In particular, it would be useful to have a variable representing student responses to a question asking how often they experienced boredom. We could then run another set of models with this extra regression term, taking the form:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_{1i} + \beta_3 X_{2i} + \beta_4 X_{3i} + \epsilon_i$$

Where X_{3i} is boredom level, β_4 is the effect of boredom level on alcohol, cannabis, or tobacco use, and the other variables are identical to those in our current models. We could then run these regression models in R as was done in this paper, again yielding estimates of the effects of rural school attendance on alcohol, cannabis, and tobacco use - these models are more likely to be truly causal than those in this report, as we have removed omitted variables bias.

6 Conclusion

Analyzing the effect of rural school attendance on substance use, we find sharply differing results across alcohol, cannabis, and cigarette use: attending a rural school decreases alcohol and cannabis use, but greatly increases cigarette use. However, this causal interpretation is put in question by potential omitted variables bias in our regression model estimates. Further research should investigate this question using a more detailed dataset, allowing more variables to be included in regressions and allowing more robust causal inferences.

References

Biolcati, R., Mancini, G., & Trombini, E. (2017). Proneness to boredom and risk behaviors during adolescents' free time. *121*(2), 303–323. https://doi.org/10.1177/0033294117724447

McAloney, K. (2015). Clustering of sex and substance use behaviors in adolescence. *Substance Use & Misuse*, *50*(11), 1406–1411. https://doi.org/10.3109/10826084.2015.1014059

Patrick, M. E., Wightman, P., Schoeni, R. F., & Schulenberg, J. (2012). Socioeconomic status and substance use among young adults: A comparison across constructs and drugs. *Journal of Studies on Alcohol and Drugs*, *73*(5), 772–782. https://doi.org/10.15288/jsad.2012.73.772

Pettigrew, J., Miller-Day, M., Krieger, J., & Hecht, M. L. (2012). The rural context of illicit substance offers: A study of appalachian rural adolescents. *Journal of Adolescent Research*, *27*(4), 523–550. https://doi.org/10.1177/0743558411432639

Radliff, K. M., Wheaton, J. E., Robinson, K., & Morris, J. (2012). Illuminating the relationship between bullying and substance use among middle and high school youth. *Addictive Behaviours*, *37*(4), 569–572. https://doi.org/10.1016/j.addbeh.2012.01.001

Ringel, J. S., Ellickson, P. L., & Collins, R. L. (2007). High school drug use predicts job-related outcomes at age 29. *Addictive Behaviours*, *32*(3), 576–589. https://doi.org/10.1016/j.addbeh.2006.05.019

Salas-Wright, C. P., Vaughn, M. G., Goings, T. T. C., Córdova, D., & Schwartz, S. J. (2018). Substance use disorders among immigrants in the united states: A research update. *Addictive Behaviors*, *76*, 169–173. https://doi.org/10.1016/j.addbeh.2017.08.014

The National Center on Addiction and Substance Abuse at Columbia University. (2001). *Malignant neglect: Substance abuse and America's schools*. https://eric.ed.gov/?id=ED460333

The National Center on Addiction and Substance Abuse at Columbia University. (2009). *Shoveling up II: The impact of substance abuse on federal, state, and local budgets*. https://drugfree.org/reports/the-impact-of-substance-abuse-on-state-budgets-2009-center-on-addiction/

Townsend, L., Flisher, A. J., & King, G. (2007). A systematic review of the relationship between high school dropout and substance use. *Clinical Child and Family Psychology Review*, *10*, 295–417. https://doi.org/10.1007/s10567-007-0023-7

West, S. L., & O'Neal, K. K. (2004). Project D.A.R.E outcome effectiveness revisited. *American Journal of Public Health*, *94*(6), 1027–1029. https://doi.org/10.2105/ajph.94.6.1027