THE EFFECTS OF ALCOHOL USE ON SCHOOL ENROLLMENT

Wesley A. Austin, University of Louisiana at Lafayette

ABSTRACT

Considerable controversy surrounds the effects youth alcohol use has on educational outcomes. This article addresses the question of whether youth drinking leads, in causal ways, to lower school enrollment, or is the widely reported negative correlation between drinking and this educational outcome caused by common unobservable factors? An instrumental variable model is estimated to study the effects of several drinking measures on the probability school enrollment for a sample of high school and college age individuals. Extensive testing is conducted to verify instrument strength and exogeneity. Results indicate that alcohol use reduces school enrollment among those of high school and college age and results are consistent across instrument specifications.

INTRODUCTION

In many health-related and social science fields, there has been considerable concern about the various harmful effects of alcohol use. Recent evidence indicates drinking, coupled with smoking, reduces income (Auld 2005). Another related consequence of alcohol use is the potential reduction in human capital accumulation by drinkers. This issue is particularly acute during adolescence and early adulthood, in which decisions regarding high school completion and college attendance are first considered, and academic performance realizations that affect longer-term educational and economic outcomes are initially observed. Excessive drinking has been associated with this age group despite its illegality until the age of 21. For instance, data from the 2006 and 2007 National Survey on Drug Use and Health (NSDUH) found approximately 18 percent of youths ages 15 – 18 (high school age) and approximately 43 percent of young adults ages 18 – 25 (college age) engaged in binge drinking, i.e. the consumption of at least five alcoholic beverages in one sitting, in the past month.

Several reasons might lead heavy drinking to impair human capital formation. Intoxication potentially interferes with class attendance and learning, and time spent in activities where drinking occurs could substitute away from time allocated to studying. This hurts academic performance in the short term, which might diminish the ability or incentive to continue schooling over the longer term. Risks stemming from intoxication, such as injury from accidents or fights, pregnancy and disease from unsafe sex, conflicts with parents or law enforcement, and a tarnished reputation with school authorities can also limit the capability of a student to remain in school (Cook and Moore 1993). Alternatively, social interactions associated with drinking might improve academic achievement by providing a means of relieving stress (Williams et al. 2003).

Much evidence has established a negative relationship between the regularity and intensity of drinking and human capital measures such as school completion. But distinguishing whether these relationships are causal, such that increased alcohol consumption directly reduces, for example, probable school enrollment, or merely correlational, with changes in other confounding variables simultaneously leading to drinking and lower enrollment rates, is critical.

Thus, for economists and policy makers, obtaining an accurate estimate of the magnitude of the causal effect that alcohol use has on educational outcomes should be a top priority. This task is a natural one to tackle by using econometric techniques such as instrumental variables (IV) regression – a method specifically designed to estimate the causal impact of a variable that does not otherwise vary independently with other unobserved determinants of the outcome being examined.

Why is the potential impact of alcohol use on school enrollment relevant for the discipline of economics? Human capital accumulation bears directly and heavily on earning potential and it is widely accepted that strong and statistically significant relationships link individual health and human capital formation. Moreover, variables such as school completion and enrollment are commonly examined education outcomes among broader literatures on human capital accumulation, given that they are easily measured and have a clear marginal impact on future wages that economists have long focused on estimating.

LITERATURE OVERVIEW

Only recently has the relationship between alcohol use and human capital accumulation been addressed by economists, and research on the topic had been fairly limited, with measures of drinking and schooling as well as conclusions varying across studies. Comparatively early research produces evidence of a negative relationship, but either makes no attempt to econometrically deal with the potential endogeneity of drinking in education equations, or does so in a way that has since been criticized as unsatisfactory, so it is unclear whether this negative correlation indeed represents declines in educational outcomes that are caused by drinking.

Cook and Moore (1993), estimate IV models in which the effect of current alcohol use on post-secondary schooling was identified by the state excise tax on beer and an indicator for whether the student could legally drink based on the state's MLDA. Results from three separate specifications show that heavy drinking in 12th grade decreased subsequent schooling. Dee and Evans (2003) call into question the causal effect interpretation of these results. They argue that the use of cross-state alcohol policy variation to identify the effects of drinking on other outcomes is potentially problematic because such variation might be correlated with unobservable attributes that affect both alcohol use and educational attainment.

Mullahy and Sindelar (1994), use ordinary least squares (OLS) regressions, and find that the onset of alcoholism symptoms by age 22 is associated with a five percent reduction in completed schooling. Yamada et al. (1996) use single equation probit models that do not account for the possibility that alcohol use is endogenous. Results show that the probability of high school graduation is 6.5 percent lower for students who consumed alcohol on at least two occasions in the previous week. In addition, drinking is inversely related to beer taxes, liquor prices, MLDAs and marijuana decriminalization, meaning that each is positively associated with high school graduation rates through its covariance with alcohol use.

Koch and Ribar (2001) examine the relationship between age of drinking onset and educational attainment by age. Estimates from IV models that specify sibling onset age as the instrument for respondent onset age imply that delaying alcohol initiation by a year increases subsequent schooling by 0.22 years. However, they argue that this represents an upper bound for the effect size based on the sign of the bias if the assumptions needed for consistency are not met, and indeed OLS and family fixed effects models produce estimates that are three to four times smaller for males, and still smaller and sometimes insignificant for females.

More recent evidence comes from Chatterji and DeSimone (2005), who estimate the effect of binge and frequent drinking by adolescents on subsequent high school dropout using an IV model with an indicator of any past month alcohol use as the identifying instrument. In contrast to the last two studies cited above, the authors find that OLS yields conservative estimates of the causal impact of heavy drinking on dropping out, such that binge or frequent drinking among 15–16 year old students lowers the probability of having graduated or being enrolled in high school four years later by at least 11 percent. The results of overidentification tests using two measures of maternal youthful alcohol use as additional instruments provide support for their empirical strategy. Also, Oreopoulos (2006) finds that the gains from policies requiring compulsory schooling up to a certain age are quite large, regardless of whether "these laws impact on a majority or minority of those exposed."

DATA

The National Survey on Drug Use and Health (NSDUH), sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA), is administered to approximately 55,000 civilian, non-institutionalized individuals age 12 and over, chosen so that the application of sample weights produces a nationally representative sample, with approximately equal numbers of respondents from the 12–17, 18–25 and 26 and over age groups. Data from the NSDUH allow for both breadth and depth of coverage on the topic. Breadth comes from the ability to study aspects of educational outcomes using data from an elaborate questionnaire covering a wide array of youth experiences. Depth is provided by numerous variables on demographics, family income, family composition and relocation.

An equally important facet of the NSDUH data is that they are conducive for the use of the IV regression methodology to estimate the causal effect of alcohol use on human capital.

Abundant information is collected on experiences related to alcohol consumption, including measures of religiosity and the perceived risks involved in alcohol/ drug use. An assortment of variables are observed, therefore, that have the potential to serve as instruments for the proposed model, in the sense that they are very likely to be highly correlated with alcohol use but would not have any obvious reason to be otherwise associated with educational outcomes.

A potentially problematic attribute of the data is non-random measurement error emanating from the self-reported nature of responses. Although IV will eliminate bias from random measurement error, it cannot salvage data plagued by systematic measurement error. However, studies on the quality of self-reported academic variables and drinking data suggest that such reporting bias should be minimal. Cassady (2001) finds that self-reported GPA values are "remarkably similar to official records" and therefore are "highly reliable" and "sufficiently adequate for research use." Grant et al. (1988), Midanik (1988) and Reinisch et al. (1991) conclude that youth drinking self-reports are reliable, based on the consistency of responses to alcohol use questions from repeated interviews. Harrison and Hughes (1997) find that survey methods not requiring subjects to verbally answer questions, as in the NSDUH, increase the accuracy of substance use self-reports.

RESEARCH METHOD AND EMPIRICAL SPECIFICATION

In determining causation, the primary methodological question is whether drinking is properly specified as an exogenous variable with respect to educational outcomes or should instead be treated as endogenous. Consider the following equations, in which drinking (D) is a function of exogenous factors and an educational variable such as school enrollment (E) is a function of some (but not all) of the same exogenous determinants as well as D,

(1) $$D = \alpha_0 + Z\alpha_1 + X\alpha_2 + \omega,$$

(2) $$E = \beta_0 + \beta_1 D + X\beta_2 + \varepsilon.$$

In the above equations, which apply to individual NSDUH respondents (with the corresponding observation-level subscript suppressed), vectors X and Z represent sets of exogenous variables that affect both drinking and enrollment (X), and drinking but not enrollment (Z), ω and ε are error terms that encompass all factors influencing the corresponding dependent variable that are not explicitly controlled for on the right hand side of the equations, and the α's and β's are parameters to be estimated. Econometrically, alcohol use is exogenous in equation 2 if it is uncorrelated with the error term ε. This condition holds, by definition, if none of the unobserved schooling determinants are related to drinking. If so, there is no need to estimate equation 1; a single equation regression method such as OLS will produce consistent estimates of the causal effect of drinking, β_1.

However, two sources of endogeneity could possibly lead to a nonzero correlation between alcohol use (D) and the error term in (2). One is unobserved heterogeneity, which would occur if any unmeasured educational outcome (e.g. enrollment) determinants that are subsumed in the error term ε are correlated with alcohol use; the resulting estimate of β_1 in (2) would suffer from omitted variable bias, which cannot be eliminated directly because the omitted variables are not recorded in the data. Disruptive events such as parental separation or divorce might simultaneously be responsible for greater alcohol consumption and lower school enrollment rates.

Such events are not observed and thus are not held constant in the regression. The negative correlation between drinking and school enrollment that they induce becomes embedded into the alcohol use coefficient, which is thus biased negatively as an estimate of the causal drinking effect. Conversely, unmeasured ability or socioeconomic background could create a positive bias in the estimated drinking effect if higher ability individuals are better able to function normally after alcohol consumption, or students who have more money to spend on alcohol also enjoy greater academic success and are more likely to be enrolled in school.

The other potential source of endogeneity is reverse causation. If alcohol use and educational outcomes like enrollment are simultaneously determined, the outcome will not only be a function of drinking, as specified in equation 2, but also will be a contributing factor to the decision regarding whether and how much alcohol to consume. In terms of equation 2, shocks to the error term ε that, by definition, influence educational outcomes will ultimately extend to drinking through the feedback effect of educational outcomes on alcohol consumption, thus creating a correlation between alcohol use and ε that renders the estimate of the causal drinking effect β_1 inconsistent. To investigate the possibility that alcohol use is endogenous as an explanatory factor for school enrollment, this analysis utilizes the method of instrumental variables (IV).

To use IV, there must be at least one, preferably two or more, variables (i.e. instruments or IVs) that affect alcohol use but have no direct impact on enrollment. In the case of exactly one instrument Z, the IV method works by estimating the causal drinking effect β_1 as the ratio of the sample correlation between the instrument and school enrollment to the sample correlation between the instrument and alcohol use, i.e.

$$(3) \qquad \beta_1 = \text{corr}[Z, E] / \text{corr}[Z, D],$$

where the quantity is estimated from the data and the correlations are estimated while holding constant the vector X of explanatory factors. Because the instrument is exogenous and related to enrollment only through drinking, the sample correlation between the instrument and enrollment is purely a product of that between drinking and enrollment. Thus, the sample correlation between the instrument and enrollment merely needs to be standardized by that between the instrument and drinking in order to be used as an estimate for the causal effect of drinking on school enrollment. In the case of two or more instruments, $\hat{D}$, the linear projection of Z onto D, takes the place of Z in equation 3.

Equation 3 makes transparent the two important conditions that the instrument vector Z must satisfy in order for IV to produce consistent estimates of the causal drinking effect β_1: First, the instruments must be highly correlated with alcohol use but not correlated with school enrollment through any other mechanism besides drinking. If the correlation between the instruments and drinking is not statistically significant, the denominator in (3) is statistically equal to zero, thus rendering the expression for β_1 indeterminate. The strength of this correlation is judged from the F-statistic for the joint significance of α_1 in equation 1. Minimally, α_1 should be significant at the 1 percent level; beyond this, Staiger and Stock (1997) advise a more stringent requirement that the associated F-statistic be at least 10.

Second, if a direct correlation between the instruments and school enrollment exists outside of the pathway from the instruments to drinking to enrollment, the numerator in (3) includes variation that is not part of the relationship between drinking and enrollment, and consequently the expression is no longer a consistent estimate of the causal effect of drinking. The reason multiple instruments are preferred is this overidentifies equation 2, which allows for specification tests to determine the empirical validity of excluding the instrument set Z from (2).

Under the null hypothesis that the instruments are not separately correlated with school enrollment, the sample size multiplied by the R-squared from a regression of the residual in (2), $\hat{\varepsilon}$, on all the exogenous variables (i.e. a constant, X and Z) is distributed as chi-square with degrees of freedom equal to one less than the number of instruments. Typically, the estimator represented by equation 3 is generated by a two-stage least squares (2SLS) procedure. The first stage estimates equation 1 above using OLS. From the estimated parameters, predicted values of alcohol use, $\hat{D}$, are constructed for each respondent using their corresponding values of the explanatory variables X and instruments Z. The second stage estimates equation 2 using the fitted values $\hat{D}$ in place of observed drinking D.

2SLS yields consistent estimates even when alcohol use and/or education variables are represented by a binary indicator. However, for binary drinking measures, e.g. an indicator of any past month binge drinking, an approach suggested by Wooldridge (2003) to improve efficiency is utilized. It is similar to 2SLS with two modifications. First, before running 2SLS, a preliminary probit regression for equation 1 is estimated. Second, the ensuing 2SLS procedure uses the predicted probabilities of drinking from the probit regression as instruments in place of Z. The resulting estimates are likely to be similar in magnitude to those that would be generated by the analogous 2SLS regression, but standard errors will be slightly smaller.

One other methodological point merits attention. Although IV estimates are consistent if the instrument strength and exogeneity conditions outlined above are satisfied, they are inefficient relative to OLS if it turns out that alcohol use is truly exogenous with respect to school enrollment, in which case the OLS estimates can be interpreted as causal effects. Thus, it is desirable to econometrically test the null hypothesis that drinking is exogenous in the enrollment equation. This is done using a Hausman (1978) test, which proffers that, if drinking and the error term are uncorrelated, IV and OLS estimates should differ only by sampling error. If the null hypothesis of exogeneity is rejected, OLS estimates are inconsistent and hence conclusions should be based on IV estimates; failure to reject the null means that OLS estimates are preferable because of their smaller standard errors.

SCHOOL ENROLLMENT

Current school enrollment is a binary variable indicating whether the respondent is currently enrolled in middle or high school (including those who are home schooled) or a college/ university. Approximately 99 percent of youth ages 15 and under report attending school, and individuals ages 26 and above who have not graduated from college are particularly likely to have experienced previous gaps in school enrollment, not currently be enrolled and not return to school in the future. The enrollment analysis is conducted utilizing a sample of high-school age students (15-18 years old) and college age students (19-25 years old). For the high school age sample, age 15 is the omitted category in the regressions thus mitigating the effects of compulsory attendance laws which typically require school attendance up to age 16.

DRINKING VARIABLES

Among the varied measures utilized are: the number of days the respondent drank in the past year (which is coded as '0' for nondrinkers and those that consumed no drinks in the previous year) and the number of drinks consumed in the previous month (which is coded as '0' for nondrinkers and those that consumed no drinks in the previous month). Binge drinking is defined as consuming five or more drinks on the same occasion on at least one day in the past thirty days. Although the timing of the number of drinks and binge drinking variables is not an ideal match for the enrolment measure, in the sense that past month consumption cannot literally affect behavior that preceded the past month, this work will follow that of previous studies in assuming that previous month drinking patterns proxy those occurring in the recent period prior to the previous month.

The impact on enrollment from alcohol abuse or dependence in the past year is also examined. This is accomplished by an indicator in the NSDUH of whether respondents exhibited symptoms of alcohol abuse or dependence in the past year. This is retrospectively coded by SAMHSA based on responses to questions corresponding to criteria outlined in the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders*, the clinical standard for establishing drug abuse and dependence.

EXOGENOUS VARIABLES

Several variables from the NSDUH data are considered exogenous (i.e. explanatory) in the model: family income is measured in seven categories: $10,000 or less; $10,000-$19,999; $20,000-$29,999; $30,000-$39,999; $40,000-$49,999; $50,000-$74,999; and $75,000 or greater, with $10,000 or less as the omitted category. Population density is represented by indicators for two categories: an MSA with one million persons or greater and an MSA of less than one million persons, with non-MSA areas as the omitted category. A binary measure is included for whether the respondent has ever been arrested. For race, indicators are specified for African Americans, Native Americans, Asians, non-white Hispanics and multiracial, with Caucasians as the omitted category in the regressions. Family size is measured using two variables: the number of members if the household has one to five members and an indicator for those with over five members. A binary measure of gender is included as well.

Age indicators for the high school age sample are 16, 17, or 18 years old and 19, 20, 21, 22 or 23, 24 or 25 years old for the college age sample. Indicators for the last grade completed is 9th, 10th or 11th grade (with 12th as the omitted grade) for the high school age sample and freshman or sophomore/ junior (with senior as the omitted category) for the college age sample.

INSTRUMENTAL VARIABLES

Several NSDUH variables conceivably influence drinking without having direct effects on school enrollment and are thus candidates to serve as instrumental variables. The specific variables utilized for the high school age sample are: perceived risk of bodily harm from alcohol use; whether religious beliefs are important and whether religious beliefs influence decisions. The specific variables utilized for the college age sample are: perceived risk of bodily harm from alcohol use; perceived risk of bodily harm from marijuana use and whether religious beliefs influence decisions.

For alcohol risk, a binary measure indicates if the respondent feels there are great/ moderate risks or slight/ no risks of harm, physically or otherwise, from consuming four to five drinks once or twice a week. For marijuana risk, a binary measure indicates if the respondent feels there are great/ moderate risks or slight/ no risks of harm, physically or otherwise, from using marijuana once or twice a week. Given that these variables only pertain to consuming illegal substances, it is presumed that there is no direct influence on school enrollment.

For both religion variables, a binary variable is created and coded as ‘0’ if religion is not important or does not influence decisions and ‘1’ otherwise. Religiosity has been linked to drinking behaviors (Kenkel and Ribar, 1994) but some evidence has established exogeneity with respect to educational outcomes (Wolaver, 2002). All instrumental variables undergo extensive testing in the following section.

EMPIRICAL FINDINGS

The causal effect drinking has on the probability of school enrollment is estimated using the three instrumental variables listed above. The main results of the IV analysis are also compared with parameter estimates obtained using OLS methodology. While discussion that follows concentrates on the effects of alcohol consumption and specification tests, appendix 1, for the binge drinking measure, shows the IV coefficients and marginal effect standard errors of all exogenous variables on the probability of enrollment for the high school age sample. Appendix 2 does the same for the college age sample.

Tables 1 and 2 present select summary statistics. The mean number of days drinks were consumed in the past year is about 18 (high school age) and 50 (college age) while the mean number of drinks consumed in the past month is 5.7 (high school age) and 15.5 (college age). Mean alcohol abuse/ dependence is 0.08 (high school age) and 0.14 (college age). Mean school enrollment is 0.44 for those of college age, and as expected, very high (0.93) for the high school age sample. Mean reported family income for college age sample is lower across the board as individuals of this age have moved out of the parental household. About 90 percent of respondents in both samples live in an MSA, roughly equally split between MSAs with populations greater than and less than one million. African Americans comprise about 14 percent of both samples while non-white Hispanics account for about 16 percent of the high school sample and 19 percent of the college sample.

Table 1. Descriptive Statistics (high school age sample) (n=19,022)

Variable	Mean	Standard Deviation
Number of days drank-past year	17.823	45.594
Number of drinks in previous month	5.703	32.916
Binge drinking in the past 30 days	0.119	0.324
Abuse/ Dependence on alcohol classification	0.080	0.272
Respondent perceives risk of harm from drinking	0.762	0.426
Religious beliefs are important in life	0.720	0.449
Religion influences your decisions	0.633	0.482
Probability of school enrollment	0.931	0.253
Family income ($10,000-$19,999)	0.108	0.310
Family income ($20,000-$29,999)	0.116	0.320
Family income ($30,000-$39,999)	0.105	0.307
Family income ($40,000-$49,999)	0.106	0.308
Family income ($50,000-$74,999)	0.190	0.392
Family income ($75,000 or more)	0.287	0.452
MSA segment with 1+ million persons	0.417	0.493

Table 1. Descriptive Statistics (high school age sample) (n=19,022)		
Variable	**Mean**	**Standard Deviation**
MSA segment of less than 1 million	0.486	0.500
Age of student (15 years old)	0.282	0.450
Age of student (16 years old)	0.278	0.448
Age of student (17 years old)	0.272	0.445
Age of student (18 years old)	0.255	0.436
Last grade in (9th grade)	0.015	0.123
Last grade in (10th grade)	0.135	0.342
Last grade in (11th grade)	0.306	0.461
Last grade in (12th grade)	0.300	0.458
Ever been arrested	0.096	0.498
Race (African American)	0.146	0.354
Race (Native American)	0.016	0.124
Race (Asian)	0.033	0.179
Race (non-white Hispanic)	0.165	0.371
Number in family	3.191	1.543
Number in family (>5)	0.139	0.346

Table 2. Descriptive Statistics (college age sample) (n=20,666)		
Variable	**Mean**	**Standard Deviation**
Number of days drank-past year	49.773	76.094
Number of drinks in previous month	15.536	50.292
Binge drinking in the past 30 days	0.300	0.458
Abuse/ Dependence on alcohol classification	0.148	0.355
Respondent perceives risk of harm from drinking	0.891	0.310
Religion influences your decisions	0.627	0.483
Respondent perceives risk of harm from marijuana	0.790	3.506
Probability of school enrollment	0.441	0.496
Family income ($10,000-$19,999)	0.156	0.362
Family income ($20,000-$29,999)	0.139	0.346
Family income ($30,000-$39,999)	0.116	0.321
Family income ($40,000-$49,999)	0.111	0.314
Family income ($50,000-$74,999)	0.140	0.347
Family income ($75,000 or more)	0.161	0.367

Table 2. Descriptive Statistics (college age sample) (n=20,666)		
Variable	**Mean**	**Standard Deviation**
MSA segment with 1+ million persons	0.399	0.489
MSA segment of less than 1 million	0.516	0.499
Age of student (19 years old)	0.157	0.364
Age of student (20 years old)	0.140	0.347
Age of student (21 years old)	0.126	0.332
Age of student (22 or 23 years old)	0.205	0.403
Age of student (24 or 25 years old)	0.189	0.392
Freshman	0.148	0.355
Sophomore/ Junior	0.191	0.393
Ever been arrested	0.193	0.395
Race (African American)	0.142	0.349
Race (Native American)	0.017	0.129
Race (Asian)	0.031	0.174
Race (non-white Hispanic)	0.192	0.394
Number in family	2.950	1.388
Number in family (>5)	0.104	0.305

FIRST STAGE REGRESSION RESULTS

Table 3 presents the probit results for the drinking measures on the instruments for the high school age sample. Of those who perceive that there is moderate to great risk of harm from consuming alcohol, the number of days drinking occurred in the past year is lowered by about 23 days. The number of drinks consumed in the past month is reduced by 11, while the likelihood of binge drinking in the last 30 days falls by 0.13 percentage points. The likelihood of being categorized as abusive/ dependent on alcohol falls by 0.09 points.

Importance of religious beliefs reduces all alcohol use measures. For those that report that religion is important in life, the number of days drinking occurred in the past year is lowered by approximately one day. The number of drinks consumed in the past month is reduced by 0.30, while the probability of binge drinking in the last 30 days falls by 0.02 percentage points. The likelihood of being categorized as abusive/ dependent on alcohol falls by 0.007 points.

When religiosity impacts decisions, the effects on the drinking measures are more pronounced. The number of days drinking occurred in the past year is lowered by nine days. The number of drinks consumed in the past month is reduced by about two, while the probability of binge drinking in the last 30 days falls by 0.45 points. The likelihood of being categorized as abusive/ dependent on alcohol falls by 0.04 points. The χ^2 coefficients and associated p-values indicate that the instruments are jointly significant for all the drinking measures.

Table 3. First stage regression estimates for the probability of enrollment (high school age) (n=19,022)				
exogeneous variables	number of days drank in past year	number of drinks in past month	Binge drinking	Abuse/ Dependence on alcohol
Risk of bodily harm from drinking	-22.895	-10.946	-0.130	-0.089
	(1.012)	(0.766)	(0.007)	(0.006)
Religious beliefs are important in life	-0.891	-0.030	-0.016	-0.007
	(0.912)	(0.691)	(0.006)	(0.006)
Religion influences your decisions	-8.676	-2.830	-0.045	-0.036
	(0.854)	(0.646)	(0.006)	(0.005)
F stat/ chi2-coefficient of joint significance	249.05	82.12	418.29	272.28
P-value of significance level	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Table 4. First stage regression estimates for the probability of enrollment (college age) (n=20,666)				
exogeneous variables	number of days drank in past year	number of drinks in past month	Binge drinking	Abuse/ Dependence on alcohol
Risk of bodily harm from drinking	-42.628	-18.468	-0.201	-0.105
	(1.579)	(1.067)	(0.009)	(0.007)
Risk of bodily harm from using marijuana	-0.816	-0.280	-0.003	-0.002
	(0.138)	(0.093)	(0.008)	(0.001)
Religion influences your decisions	-15.077	-4.690	-0.086	-0.039
	(1.018)	(0.688)	(0.006)	(0.005)
F stat/ chi2-coefficient of joint significance	352.67	125.76	665.92	241.11
P-value of significance level	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Table 4 presents the probit results for the instruments for the college age group. For this age group, if moderate to great risk of harm from consuming alcohol is perceived, the number of days in which drinking occurred in the past year is lowered by 42 days. The number of drinks consumed in the past month is reduced by roughly18, while the probability of binge drinking in the last 30 days falls by 0.20 percentage points. The likelihood of being categorized as abusive/ dependent on alcohol decreases by 0.11 points.

If moderate to great risk of harm from using marijuana is perceived, the number of days in which drinking occurred in the past year is lowered by one day. The number of drinks consumed in the past month is reduced by 0.28, while the probability of binge drinking in the last 30 days falls by 0.003 percentage points. The likelihood of being categorized as abusive/ dependent on alcohol falls by 0.002 points. When religiosity impacts decisions, the number of days in which drinking occurred in the past year is reduced by 15 and the number of drinks consumed in the past month is reduced by four. The probability of binge drinking in the last 30 days falls by 0.09 percentage points while the likelihood of being categorized as abusive/ dependent on alcohol falls by 0.04 points. The F statistics and χ^2 p-values signify support for the hypothesis of joint instrument significance for all the drinking measures.

THE EFFECTS OF DRINKING ON THE PROBABILITY OF SCHOOL ENROLLMENT (HIGH SCHOOL AGE)

As shown in table 5, drinking has significant, negative effects on the probability of being enrolled. For each daily increase in past year drinking, the probability of being enrolled is subsequently lowered by 0.001. For each additional drink increase in the past month, the probability of enrollment is also lowered by 0.003. If, for instance, the respondent reports drinking 52 days in the previous year, the likelihood of enrollment is diminished by approximately 0.052 points compared to not drinking at all. If the student reports consuming 30 drinks in the previous month, the probability of enrollment decreases by 0.09 points.

Table 5. IV estimates of drinking on the probability of enrollment (high school age) All three instruments (n=19,022)

Alcohol variables	IV	OLS
number of days drank-past year	-0.001*	-0.0002*
Marginal Effect Standard Error	(0.0002)	(0.0000)
P-value of overidentification test	0.828	
Hausman statistic (p-value)	-5.243 (0.000)	
number of drinks in past month	-0.003*	-0.0003*
Marginal Effect Standard Error	(0.0006)	(0.0001)
P-value of overidentification test	0.303	
Hausman statistic (p-value)	-4.483 (0.000)	
binge drinking	-0.230*	-0.0042*
Marginal Effect Standard Error	(0.040)	(0.0054)
P-value of overidentification test	0.649	
Hausman statistic (p-value)	-5.772 (0.000)	
abuse/ dependence on alcohol	-0.329*	0.0017*
Marginal Effect Standard Error	(0.060)	(0.0060)
P-value of overidentification test	0.825	
Hausman statistic (p-value)	-5.624 (0.000)	
*Statistically significant at 1%		

Binge drinking further reduces the probability of enrollment by 0.23 points. For students who have engaged in binge drinking, the probability of school enrollment declines by approximately 24 percent compared to not binging. For those classified as abusive/ dependent with respect to alcohol, the probability of enrollment decreases by 0.32 points and this categorization reduces the probability of school enrollment by 35 percent. For all drinking indicators, the overidentification tests have associated p-values that offer strong evidence in support of the assumption of instrument exogeneity at the 10 percent level. The p-values associated with the Hausman coefficient signify that there are statistically significant differences between the OLS and IV parameter estimates for all the drinking measures.

Overall, in the high school sample, there is a strong indication that drinking, possibly by raising the opportunity cost of high school education, impairing cognitive functioning, etc., reduces enrollment in high school. And, considering the additional resources the student devotes toward drinking if the student binge drinks or is abusive/ dependent on alcohol, there is compelling evidence that the probability of high school enrollment is largely and negatively impacted.

INSTRUMENT ROBUSTNESS AND THE PROBABILITY OF ENROLLMENT (HIGH SCHOOL AGE)

To determine if there is any sensitivity in the main results attributable to changes in the instrument set, regressions are performed with varying pairs of instruments with results presented in table 6. The instrument that is omitted from the IV combination is utilized as an explanatory variable and its coefficient and standard error is reported.

For all drinking variables, the effect on enrollment using IV pairs is remarkably similar to those in the main regression where all three instruments are employed. For all drinking variables the overidentification test results support exogeneity for all IV pairs. Hausman tests indicate there are statistically significant differences between IV and OLS estimates in all specifications and the additional instrument not used to identify drinking is never significant in the enrollment equation.

THE EFFECTS OF DRINKING ON THE PROBABILITY OF SCHOOL ENROLLMENT (COLLEGE AGE)

As shown in table 7, drinking has significant, negative effects on the probability of being enrolled for the college age group. For each daily increase in past year drinking, the probability of being enrolled is subsequently lowered by 0.001. For each additional drink increase in the past month, the probability of enrollment is also lowered by 0.002. If, for instance, the respondent reports drinking 52 days in the previous year, the likelihood of enrollment is diminished by approximately 0.052 points compared to not drinking at all. If the student reports consuming 30 drinks in the previous month, the probability of enrollment decreases by 0.06 points.

Binge drinking and abuse/ dependence on alcohol further reduce the probability of enrollment by 0.19 points. For students who have engaged in binge drinking, the probability of school enrollment declines by approximately 43 percent compared to not binging. For those classified as abusive/ dependent with respect to alcohol, the probability of enrollment decreases by 0.37 points. Categorization as abusive/ dependent reduces the probability of school enrollment by 83 percent.

Table 6. IV estimates of drinking on the probability of enrollment using IV pairs (high school age) (n=19,022)

Alcohol variables	religion important and alcohol risk	religious decisions and alcohol risk	religion important and religious decisions
number of days drank-past year	-0.001*	-0.001*	-0.002*
Marginal Effect Standard Error	(0.0003)	(0.0003)	(0.0004)
P-value of overidentification test	0.942	0.828	0.931
Hausman statistic (p-value)	-3.958 (0.000)	-4.759 (0.000)	-3.360 (0.000)
Coefficient (Standard Error) of omitted IV	0.002 (0.005)	-0.0002 (0.004)	-0.005 (0.012)
number of drinks in past month	-0.003*	-0.003*	-0.005*
Marginal Effect Standard Error	(0.0007)	(0.0006)	(0.0016)
P-value of overidentification test	0.992	0.429	0.995
Hausman statistic (p-value)	-3.627 (0.000)	-4.128 (0.000)	-3.024 (0.000)
Coefficient (Standard Error) of omitted IV	0.006 (0.004)	0.004 (0.004)	-0.025 (0.020)
binge drinking	-0.220*	-0.239*	-0.240*
Marginal Effect Standard Error	(0.051)	(0.047)	(0.067)
P-value of overidentification test	0.702	0.739	0.662
Hausman statistic (p-value)	-4.354 (0.000)	-5.197 (0.000)	-3.577 (0.000)
Coefficient (Standard Error) of omitted IV	0.002 (0.005)	-0.002 (0.005)	-0.002 (0.011)
abuse/ dependence on alcohol	-0.323*	-0.341*	-0.333*
Marginal Effect Standard Error	(0.078)	(0.069)	(0.095)
P-value of overidentification test	0.834	0.906	0.826
Hausman statistic (p-value)	-4.238 (0.000)	-5.092 (0.000)	-3.602 (0.000)
Coefficient (Standard Error) of omitted IV	0.001 (0.005)	-0.002 (0.005)	-0.001 (0.011)
*Statistically significant at 1%			

Table 7. IV estimates of drinking on the probability of enrollment (college age)
All three instruments
(n=20,666)

Alcohol variables	IV	OLS
number of days drank-past year	-0.001*	-0.0001*
Marginal Effect Standard Error	(0.0002)	(0.0000)
P-value of overidentification test	0.162	
Hausman statistic (p-value)	-5.043 (0.000)	
number of drinks in past month	-0.002*	-0.0002*
Marginal Effect Standard Error	(0.0004)	(0.0001)
P-value of overidentification test	0.082	
Hausman statistic (p-value)	-4.528 (0.000)	
binge drinking	-0.191*	-0.0112*
Marginal Effect Standard Error	(0.0359)	(0.0070)
P-value of overidentification test	0.263	
Hausman statistic (p-value)	-5.963 (0.000)	
abuse/ dependence on alcohol	-0.376*	0.0127*
Marginal Effect Standard Error	(0.0756)	(0.0080)
P-value of overidentification test	0.225	
Hausman statistic (p-value)	-5.258 (0.000)	
*Statistically significant at 1%		

For number of days drinking occurred in the past year, binging and abuse/ dependence on alcohol, the overidentification tests have associated p-values that afford strong evidence in support of the assumption of instrument exogeneity at the 10 percent level. Even for the past month drinking variable, instrument exogeneity is not rejected at the 5 percent level. The p-values associated with the Hausman coefficient signify that OLS and IV estimates statistically differ for all the drinking measures.

The estimated effects for binge drinking and abuse/ dependence are quite large, possibly indicating that for college age individuals, resources (monetary and otherwise) spent on drinking undercut the probability of post high school education, especially considering that there are greater costs (especially monetary) associated with obtaining education at that age. In addition, if the college age person has a history of drinking, especially at abuse and dependence levels, pre-college academic achievement might have been much lower thus precluding post high school enrollment in colleges, universities and other institutions.

INSTRUMENT ROBUSTNESS AND THE PROBABILITY OF SCHOOL ENROLLMENT (COLLEGE AGE)

To determine if there is any sensitivity in the main results attributable to changes in the instrument set, regressions are performed with varying pairs of instruments with results presented in table 8. Again, the instrument that is omitted from the IV combination is utilized as an explanatory variable and its coefficient and standard error is reported.

Table 8. IV estimates of drinking on the probability of enrollment using IV pairs (college age) (n=20,666)

Alcohol variables	religious decisions and alcohol risk	religious decisions and marijuana risk	alochol risk and marijuana risk
number of days drank-past year	-0.001*	-0.001*	-0.001*
Marginal Effect Standard Error	(0.0002)	(0.0003)	(0.0002)
P-value of overidentification test	0.456	0.215	0.353
Hausman statistic (p-value)	-5.211 (0.000)	-3.081 (0.000)	-3.574 (0.000)
Coefficient (Standard Error) of omitted IV	0.001 (0.001)	-0.013 (0.018)	-0.001 (0.007)
number of drinks in past month	-0.002*	-0.004*	-0.002*
Marginal Effect Standard Error	(0.0004)	(0.0010)	(0.0005)
P-value of overidentification test	0.177	0.213	0.447
Hausman statistic (p-value)	-4.627 (0.000)	-2.865 (0.000)	-3.448 (0.000)
Coefficient (Standard Error) of omitted IV	0.001 (0.001)	0.030 (0.025)	-0.003 (0.007)
binge drinking	-0.202*	-0.213*	-0.165*
Marginal Effect Standard Error	(0.036)	(0.064)	(0.043)
P-value of overidentification test	0.718	0.289	0.350
Hausman statistic (p-value)	-6.102 (0.000)	-3.605 (0.000)	-4.287 (0.000)
Coefficient (Standard Error) of omitted IV	0.001 (0.001)	-0.006 (0.016)	-0.002 (0.007)
abuse/ dependence on alcohol	-0.396*	-0.458*	-0.320*
Marginal Effect Standard Error	(0.078)	(0.148)	(0.086)
P-value of overidentification test	0.550	0.295	0.401
Hausman statistic (p-value)	-5.357 (0.000)	-3.216 (0.000)	-3.911 (0.000)
Coefficient (Standard Error) of omitted IV	0.001 (0.001)	-0.012 (0.020)	-0.002 (0.007)
*Statistically significant at 1%			

For all drinking variables, the effect on enrollment is remarkably similar to those in the main regression. For all drinking variables the overidentification test results support the exogeneity hypothesis for all IV pairs. Hausman tests indicate there are statistically significant differences between IV and OLS estimates in all specifications and the additional instrument not used to identify drinking is never significant in the enrollment equation.

Overall, the robustness evaluation for both samples offers strong evidence to support the hypothesis that instruments are exogeneous. Throughout the analyses, OLS parameter estimates consistently underestimate the magnitude of the negative effects in the main specification for enrollment. This could be ascribed to the prospect that higher ability (i.e. higher achieving) students perform better academically even when they drink. And these higher achievers are more likely to be enrolled in school. In addition, higher income students (who spend more on alcohol and therefore drink more) also command more resources that can be channeled toward education, such as test preparation for the SAT, and simply have more money to pay for college, and, once in college, funds to pay for tutoring services, etc. This in turn could serve to keep enrollment elevated.

CONCLUDING REMARKS

This paper contributes to the literature by examining the effects of youth drinking on the probability of school enrollment while accounting for unobserved endogeneity. The literature has established a negative link between drinking and educational variables, but many of these studies do not account for the possibility that the negative correlation between these factors may be the result of unobserved variables that cause simultaneous increases in drinking and reductions in educational variables. And, for studies that have incorporated unobserved endogeneity, instrumental variable procedures have been subject to criticism.

This study finds strong evidence that the probability of school enrollment is lowered when students use alcohol more frequently and intensely. Binge drinking and abuse of alcohol have the most detrimental impact on enrollment. Throughout the analysis, overidentification tests generally confirm instrument exogeneity and thus show that adolescent alcohol consumption should be treated as endogenous. OLS regressions consistently underestimate the effects of alcohol use on enrollment.

Although there is no direct analysis of the effectiveness of laws and other programs designed to curtail youth drinking, the conclusions in this paper support the premise that reducing adolescent alcohol use enhances human capital accumulation. Minimum legal drinking ages, high school anti-drug programs and other policies aimed at lowering youth drinking may well be justified on human capital grounds. Although the instrumental variables prove to be very effective and useful, further research should include continued exploration for reliable instruments to ensure that the relationship between drinking and academic outcomes is properly identified. A further examination of the effectiveness of public policies that purport to reduce youth drinking would also prove valuable.

REFERENCES

Auld, M. (2005). Smoking, Drinking, and Income. *Journal of Human Resources, 15*(2), 505-518.

Bryant, A., Schulenberg, J., O'Malley, P., Bachman, J., & L. Johnston (2003). How academic achievement, attitudes, and behaviors relate to the course of substance use during adolescence: A 6-year, multiwave longitudinal study. *Journal of Research on Adolescence*, 13(3), 361–397.

Chatterji, P. and J. DeSimone (2005). Adolescent drinking and high school dropout. Working paper 11337, National Bureau of Economic Research.

Cook, P. J., & M. J. Moore (1993). Drinking and schooling. *Journal of Health Economics*, 12(4), 411–429.

Dee, T. S., & W. N. Evans (2003). Teen drinking and educational attainment: Evidence from two-sample instrumental variables estimates. *Journal of Labor Economics*, 21(1), 178–209.

Evans, W., Oates, W., & R. Schwab (1992). Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy*, 100(51), 966–991.

Grant, B. F., Harford, T. C., & M. B. Grigson (1998). Stability of alcohol consumption among youth: a national longitudinal survey. *Journal of Studies on Alcohol*, 49(3), 253–260.

Harrison, L., & A. Hughes (1997). The validity of self-reported drug use: Improving the accuracy of survey estimates. *NIDA Research Monograph*, 167, 1–16.

Hausman, J. (1978). Specification tests in econometrics. *Econometrica*, 46(6), 1251-1271.

Hoyt, G., & F. Chaloupka (1994). Effect of survey conditions on self-reported substance use. *Contemporary Economic Policy*, 12(3), 109–121.

Koch, S. F., & D. C. Ribar (2001). A siblings analysis of the effects of alcohol consumption onset on educational attainment. *Contemporary Economic Policy*, 19(2), 162–174.

Midanik, L. (1998). Validity of self-reported alcohol use: a literature review and assessment. *British Journal of Addiction*, 83, 1019–1029.

Mullahy, J., & J. L. Sindelar (1994). Alcoholism and income: The role of indirect effects. *The Milbank Quarterly*, 72(2), 359–375.

Oreopoulos, P. (2006). Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter. *American Economic Review*, 96(1), 152–175.

Reinisch, E. J., Bell, R. M., & P. Ellickson (1991). How accurate are adolescent reports of drug use? Santa Monica, CA: Rand Corporation.

Staiger, D., & J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3), 557–586.

Williams, J., Powell, L. M., & H. Wechsler (2003). Does alcohol consumption reduce human capital accumulation? Evidence from the college alcohol study. *Applied Economics*, 35(10) 1227–1239.

Wolaver, A M. (2002). Effects of heavy drinking in college on study effort, grade point average, and major choice. *Contemporary Economic Policy*, 20(4), 415–428.

Wooldridge, J. M. (2003). *Introductory econometrics: a modern approach (second edition).* South-Western College Publishing.

Yamada, T., Kendix, M., & T. Yamada (1996). The impact of alcohol consumption and marijuana use on high school graduation. *Health Economics*, 5(1), 77–92.

Appendix 1. All IV estimates on the probability of enrollment for binge drinking (high school age) (n=19,022)	
Explanatory variables	IV coefficient (Marginal Effect SE)
Binge drinking	-0.229 (0.040)
Female	-0.005 (0.003)
Race (African American)	-0.003 (0.006)
Race (Native American)	-0.026 (0.017)
Race (Asian)	0.028 (0.007)
Race (non-white Hispanic)	-0.034 (0.005)
Age of student (16 years old)	-0.034 (0.005)
Age of student (17 years old)	-0.124 (0.007)
Age of student (18 years old)	-0.255 (0.009)
Last grade completed (9th grade)	0.001 (0.005)
Last grade completed (10th grade)	0.044 (0.007)
Last grade completed (11th grade)	0.141 (0.008)
Ever been arrested	-0.031 (0.010)
Number in family	-0.007 (0.002)
Number in family (>5)	-0.058 (0.015)
Family income ($10,000-$19,999)	-0.045 (0.011)
Family income ($20,000-$29,999)	-0.017 (0.109)
Family income ($30,000-$39,999)	-0.005 (0.010)
Family income ($40,000-$49,999)	0.011 (0.010)
Family income ($50,000-$74,999)	0.024 (0.009)
Family income ($75,000 or more)	0.032 (0.009)
MSA segment with 1+ million persons	-0.003 (0.006)
MSA segment of less than 1 million	-0.007 (0.006)
Year 2006 indicator	-0.027 (0.006)

Appendix 2. All IV estimates on the probability of enrollment for binge drinking (college sample) (n=20,666)	
Explanatory variables	IV coefficient (Marginal Effect SE)
Binge drinking	-0.191 (0.035)
Female	-0.027 (0.007)
Race (African American)	-0.009 (0.011)
Race (Native American)	-0.026 (0.022)
Race (Asian)	0.111 (0.016)
Race (non-white Hispanic)	-0.068 (0.008)
Age of student (19 years old)	-0.271 (0.007)
Age of student (20 years old)	-0.434 (0.010)
Age of student (21 years old)	-0.503 (0.011)
Age of student (22-23 years old)	-0.599 (0.010)
Age of student (24-25 years old)	-0.690 (0.009)
Last grade completed (Freshman)	0.350 (0.008)
Last grade completed (Sophomore/ Junior)	0.512 (0.008)
Ever been arrested	-0.030 (0.010)
Number in family	-0.012 (0.003)
Number in family (>5)	-0.103 (0.014)
Family income ($10,000-$19,999)	-0.115 (0.010)
Family income ($20,000-$29,999)	-0.133 (0.010)
Family income ($30,000-$39,999)	-0.122 (0.010)
Family income ($40,000-$49,999)	0.125 (0.011)
Family income ($50,000-$74,999)	0.086 (0.010)
Family income ($75,000 or more)	0.027 (0.010)
MSA segment with 1+ million persons	0.082 (0.011)
MSA segment of less than 1 million	0.060 (0.010)
Year 2006 indicator	-0.056 (0.010)