The relationship between drug use and labor supply for young men

By: Gary A. Zarkin, Thomas A. Mroz, Jeremy W. Bray, Michael T. French

Zarkin, G. A., Mroz, T., Bray, J. W., & French, M. T. (1998). The relationship between drug use and labor supply for young men. *Labour Economics*, *5*(4), 385–409.

Abstract:

This paper examines the relationship between young men's hours worked and their use of marijuana, alcohol, cigarettes, cocaine, and other drugs using cross-section data from the 1991 National Household Survey on Drug Abuse (NHSDA), a nationally representative survey of the U.S. noninstitutionalized population age 12 and over. Our results indicate that substance use has little effect on the number of hours worked by young men in the past month, with the exception that young men who smoked 1 to 3 marijuana joints in the last month worked 42 *more* hours than nonusers. To assess the robustness of our 1991 results, we re-estimated identical models using data from the 1992 NHSDA, an independent cross-section that was collected using the same methodology as the 1991 survey. Comparing the 1991 and 1992 results, the 1992 data also show that substance use has little relationship overall to the number of hours worked. However, in contrast to the 1991 results, the 1992 results show that smoking 1 to 3 marijuana joints in the last month is associated with 41 *fewer* hours worked than nonusers. This paper is the first study to assess the robustness of drug use–labor supply results on adjacent cross sections. Our results demonstrate the value of re-estimating the drug use–labor supply relationship.

Keywords: drug use | labor supply | young men

Article:

1. Introduction

Substance use and abuse can create serious social costs. Rice et al. (1991)estimate that drug and alcohol abuse imposed a US$44.1 billion cost on society in 1985, and a significant proportion of these costs—US$14 billion—can be attributed to lost productivity. This loss in national productivity can occur because of reductions in wages, reductions in the propensity to work, a decrease in hours worked conditional on working, or all three.

In spite of the plausibility of a negative relationship between substance abuse and labor market outcomes, the literature indicates a mixed relationship between drug use and labor market outcomes. Most of the research on the labor market effects of drug use focuses on the drug use–wage relationship. As we discuss in more detail in French et al. (1996), the expected negative relationship between drug use and wages has not always been found. For example, several studies have shown that drug use is not significantly related to earnings (e.g., Kandel and Davies, 1990; Buchmueller and Zuvekas, 1994) or is associated with higher wages (Kaestner, 1991Kaestner, 1994; Gill and Michaels, 1992; Register and Williams, 1992; Zarkin et al., 1998).

A number of researchers have also estimated the relationship between drug use and the propensity to work. Overall, no consistent pattern emerges to describe this relationship. For example, Gill and Michaels (1992)found that predicted drug use is associated with a lower probability of employment, but that predicted hard drug use (e.g., cocaine, heroin, inhalants) is, surprisingly, not significantly related to employment. Register and Williams (1992) found that predicted marijuana use in the past 30 days is negatively associated with the probability of being employed, but having used marijuana on the job any time in the last year is positively associated with currently being employed; results for cocaine show a similar sign pattern, but the coefficients are not significant. Buchmueller and Zuvekas (1994) found that moderate drug use and drug abuse are not significantly related to young men's employment status.

A relatively understudied area is the relationship between labor supply and drug use.Register and Williams (Undated)used the National Longitudinal Survey of Youth (NLSY) to examine the impact of marijuana use on the labor supply of young men. They found significant *positive* effects of marijuana use on hours worked for all men and especially for unmarried men. Kaestner (1993), who also used the NLSY, presents both cross-section and panel data estimates of the relationship between annual hours of work and current and lifetime drug use. Although Kaestner focuses on the panel data results, he found substantial negative labor supply effects of current marijuana use for married and single males in the 1988 cross section. Surprisingly, he found substantially smaller and insignificant effects in 1984. Similar results hold for cocaine use. Married and single men in 1988 who had used cocaine in the past 30 days worked substantially fewer hours per year than nonusers. However, current cocaine use was not significantly related to hours worked for men in 1984 or women in either 1984 or 1988. In contrast, DeSimone (1996)examined the effects of marijuana and cocaine consumption on hours worked with data from the 1984, 1988, and 1992 panels of the NLSY. DeSimone utilizes regional cocaine prices and indicators of state marijuana decriminalization as instruments for cocaine and marijuana use. He finds a substantial negative relationship between drug use and both employment and hours worked during the past year for men but not for women.

This brief overview of the substance abuse–labor market outcomes literature suggests that no consistent relationship exists between drug use and labor market outcomes. Because no estimated coefficients can be dismissed as being `unreasonable', researchers must be particularly careful in their empirical work. To focus our analysis and allow us to address carefully the empirical challenges, we have consciously chosen to examine only one aspect of the substance abuse–labor supply relationship that has substantial policy relevance: the relationship between young men's drug use and hours worked (conditional on working positive hours). We focus on hours worked because modeling discrete outcomes (e.g., employment status) with endogenous explanatory variables generally requires an explicit specification of the joint distribution of all of the disturbances in the model. Except in the case of continuous explanatory variables with linear normal disturbances in all equations, there are no simple approaches to address the endogeneity issue (see, for example, Mroz and Guilkey, 1992). Consequently, we defer an examination of the relationship between substance use and labor force participation to future work. Lastly, our choice of young men is motivated by the fact that they are the heaviest users of drugs, and our results will provide important insights into the drug use–labor supply relationship for individuals during the early years of their employment careers.

Our paper provides a number of empirical and methodological contributions that will enhance the understanding of the relationship between drug use and labor supply. First, we use independent cross-section data from the 1991 and 1992 National Household Surveys on Drug Abuse (NHSDA), a nationally representative survey of the U.S. noninstitutionalized population aged 12 and older. In contrast, much of the literature on the relationship between substance use and labor market outcomes has used the NLSY. An advantage of the NLSY is that it contains panel data on youth, and youth are relatively high substance users (Zarkin et al., 1995). However, the NLSY is limited by its relatively poor measures of drug use. The NLSY measures the *frequency* of current and lifetime drug use and not the *quantity* of use per occasion (Kaestner, 1993). A distinct advantage of the NHSDA compared to the NLSY is that it includes both the quantity and frequency of current drug and alcohol use variables. By multiplying the quantity of use per occasion by the frequency of use, we can estimate the total consumption of particular substances.

The richness of the NHSDA also allows us to include in our hours regressions the quantity of marijuana and alcohol use, as well as whether the individual smoked cigarettes, used cocaine, or used other drugs. Thus, another strength of our approach is that our illicit drug use estimates also control for the potentially confounding effects of other substance use.

Another contribution of our analysis is that we first estimated our hours of work model using the 1991 NHSDA data and then re-estimated our final model using 1992 data. By examining adjacent independent cross sections of a nationally representative household survey, we can evaluate the robustness of our empirical models with cross-section data—which are much more common than panel data for studying these relationships—and in a setting in which we expect to see stability of the estimated parameters. We are aware of no other paper that assesses the robustness of drug use–labor market outcomes in adjacent cross sections. In addition, this separation between the specification search and final estimation allows classical statistical interpretation of our 1992 results and eliminates the bias potentially associated with pretesting alternative specifications.

Although the majority of the coefficient estimates are similar in sign and significance level between adjacent cross-section data sets, we did find that the coefficient on light marijuana use was statistically significant in both years, but surprisingly with opposite signs. This conflicting result raises a potentially troubling question about how much weight should be placed on the results of a single cross section, or, for that matter, on the results using short-duration panel data. The striking difference between the light marijuana use coefficient in adjacent years suggests there might be considerable pre-test bias in the 1991 estimate. Given the disparity of estimates in the literature and the large cross-section variability of our results, researchers in this area may want to use cross-validation techniques to minimize the potential for pre-test bias.

2. Data

We used independent cross-section data from the 1991 and 1992 NHSDA public use files. The NHSDA instrument collects data on the prevalence of current and lifetime use of tobacco, alcohol, and illicit drugs, as well as basic demographic and employment data. Because of the sensitive nature of the survey topic, self-administered answer sheets are used for the drug use questions to increase the confidentiality and anonymity of the respondent's answers. This format is designed to minimize underreporting of drug use, which is a potential limitation of self-reported surveys (Hoyt and Chaloupka, 1993). In a 1990 field test of various survey instruments, Turner et al. (1992)found that the self-administered format of the NHSDA decreases the underreporting of drug use compared to an interviewer-administered format.

In both 1991 and 1992 the NHSDA used a five-stage area probability sample design (Substance Abuse and Mental Health Services Administration, 1992 and Substance Abuse and Mental Health Services Administration, 1994). The two surveys refer to the same population, the U.S. household population; were collected by the same firm, Research Triangle Institute; and use the same methodology. Sampling weights were computed based on the probability of selection at each stage, and the weights were used in all analyses. Because of the stratified sampling strategy employed by the NHSDA, the observations are clustered within the primary sampling unit. Our estimation procedures (discussed below) account for the clustered sample design.

Our analysis file includes young men between the ages of 18 and 24 who were not on active military duty and not currently enrolled in school. In addition, we included only those respondents who had nonmissing values for all variables used in our estimation procedures. The final sample sizes used in our analyses are 893 individuals in 1991 and 1019 individuals in 1992.

The dependent variable in our analysis is the self-reported hours worked at all jobs in the past month by employed young males. While we acknowledge that studying the relationship between drug use and labor market participation is an important question deserving careful attention, we focus on the relationship between drug use and hours worked for young men who work. Thus, our dependent variable is strictly positive.[1] Our core demographic variables include state indicator variables, race, education, marital status, nonlabor income, number of children living in the household, self-reported health status, and age. As we describe below, we also include interactions between age and race in our estimating equations. Table 1 reports the means of the demographic variables (excluding the state indicator variables), in addition to other variables.

---

[1] Our sample size would be approximately 33% larger if we included individuals with zero hours of work. Marijuana use is slightly higher in the $h = 0$ group compared to the $h > 0$ group, but all other substance use is approximately the same across the two groups.

Table 1
Variable definitions and descriptive statistics

Variable	1991 ($n = 893$)	1992 ($n = 1019$)
Past month substance use		
Any past month marijuana use	0.133 (0.340)	0.125 (0.331)
Any past month cocaine use	0.035 (0.183)	0.024 (0.152)
Any past month heroin, hallucinogens, or inhalants use	0.040 (0.197)	0.032 (0.177)
Any past month alcohol use	0.691 (0.462)	0.661 (0.474)
Any past month cigarette use	0.408 (0.492)	0.360 (0.480)
Lifetime substance use		
Any lifetime marijuana use	0.587 (0.493)	0.570 (0.495)
Any lifetime cocaine use	0.243 (0.429)	0.181 (0.386)
Any lifetime heroin, hallucinogens, or inhalants use	0.220 (0.414)	0.209 (0.407)
Any daily cigarette use	0.421 (0.494)	0.396 (0.489)
Age respondent tried alcohol (alcohol users only)	14.860 (2.816)	14.880 (2.742)
Age respondent first drank alcohol monthly (monthly alcohol users only)	16.955 (2.512)	17.280 (2.276)
Never drank alcohol	0.068 (0.251)	0.093 (0.291)
Never smoked cigarettes	0.217 (0.412)	0.236 (0.425)
Labor supply		
The number of hours worked in the past month by the respondent at all jobs, conditional on the respondent's working in the past month	156.999 (60.227)	158.748 (57.704)
Demographics		
White	0.729 (0.445)	0.708 (0.455)
Black	0.110 (0.313)	0.124 (0.330)
Hispanic	0.134 (0.341)	0.143 (0.351)
Native American, Asian, or other race	0.027 (0.163)	0.024 (0.154)
Age	21.661 (1.844)	21.836 (1.876)
Fewer than 12 yr of education	0.241 (0.428)	0.231 (0.421)
Exactly 12 yr of education	0.509 (0.500)	0.507 (0.500)
13 to 15 yr of education	0.166 (0.372)	0.157 (0.364)
16 or more years of education	0.084 (0.278)	0.105 (0.307)
MSA with a population of more than 1 million	0.439 (0.497)	0.446 (0.497)
MSA with a population of 250,000 to 1 million	0.253 (0.435)	0.235 (0.424)
MSA with a population of less than 250,000	0.084 (0.278)	0.067 (0.250)
Non-MSA urban area	0.068 (0.252)	0.106 (0.308)
Rural area	0.155 (0.362)	0.146 (0.353)
Family nonlabor income	20.950 (82.326)	10.873 (60.442)
Married	0.243 (0.429)	0.314 (0.464)
Number of biological or adopted children living in the household	0.208 (0.515)	0.304 (0.647)
Self-assessed fair to poor health in the past year	0.064 (0.246)	0.058 (0.235)
Risk variables		
Perceived risk[a] of		
smoking one or more packs of cigarettes per day	3.350 (0.792)	3.333 (0.746)
trying marijuana once or twice	2.215 (1.115)	2.283 (1.102)

Table 1 (continued)

Variable	1991 ($n = 893$)	1992 ($n = 1019$)
Risk variables		
using marijuana occasionally	2.768 (0.996)	2.801 (0.964)
using marijuana regularly	3.502 (0.810)	3.503 (0.748)
trying heroin once or twice	3.544 (0.759)	3.490 (0.811)
using heroin regularly	3.922 (0.395)	3.903 (0.408)
trying cocaine once or twice	3.290 (0.947)	3.339 (0.885)
using cocaine occasionally	3.684 (0.654)	3.685 (0.627)
using cocaine regularly	3.890 (0.402)	3.893 (0.431)
taking one or two drinks nearly every day	2.580 (1.005)	2.611 (0.951)
taking four to five drinks nearly every day	3.330 (0.824)	3.419 (0.768)
taking five or more drinks once or twice a week	3.069 (0.956)	3.152 (0.920)
Perceived difficulty[b] of		
getting marijuana	4.261 (1.049)	4.152 (1.126)
getting cocaine	3.466 (1.439)	3.417 (1.421)
getting heroin	2.459 (1.331)	2.529 (1.396)

[a]Responses range from 1 for no risk to 4 for great risk.
[b]Responses range from 1 for probably impossible to 5 for very easy.
Note: Standard deviation of the variable in parentheses.

We defined five categories of substance use: marijuana use, cocaine use, cigarette use, alcohol use, and use of other drugs consisting of heroin, hallucinogens, and inhalants. We also distinguished between past month use[2] and ever-in-lifetime use. For past month alcohol and marijuana use, we used questions from the NHSDA on the frequency and quantity of past month use. We estimated the total quantity of alcohol or marijuana used in the past month by multiplying the reported number of times that each of these substances was used in the past month by the average consumption per occasion (joints for marijuana, drinks for alcohol) for that individual. To reduce the impact of observations with very large consumption, we created discrete categories of use in the past month (among those who used alcohol or marijuana) that represent approximate quartiles of use.[3] For marijuana, the four categories are 1 to 3 joints, 4 to 15 joints, 16 to 59 joints, and 60 or more joints. For alcohol, the four categories of past month use are 1 to 7 drinks, 8 to 23 drinks, 24 to 59 drinks, and 60 or more drinks. Respondents who used no alcohol or marijuana in the past month were assigned to separate categories. All other past month substance use variables are simple use/no-use indicators.

Table 1 provides the weighted descriptive statistics for the substance use variables in 1991 and 1992. The prevalence results indicate that in 1991 approximately 13% of young men aged 18 to 24 who worked and were not enrolled in school used marijuana in the past month, 69% used alcohol in the past month, and 41% smoked cigarettes in the past month. Looking at the drug use variables, marijuana was used much more frequently in the last month than either cocaine (3.5%) or heroin and other drugs (4%). The prevalence estimates in 1992 are similar to those in 1991.

We also include measures of lifetime substance use. To estimate lifetime use of marijuana, cocaine, and other drugs, we used categorical response questions from the NHSDA that asked

[2] The NHSDA instrument asks about use 'in the past 30 days', which we refer to as 'past month use'.
[3] To implement this, we identified the unweighted quartiles of past month use in 1991 and assigned each user to his relevant category. Because the categories were based on 1991 use data, the categories were only approximate quartiles based on 1992 use data.

respondents how many times they had used a substance in their lives. Although the NHSDA provided eight possible lifetime use categories for each substance (including one for no lifetime use), the low prevalence rates of some response categories required us to collapse the eight categories into five. The five categories of lifetime use are never used, used 1 or 2 times, used 3 to 10 times, used 11 to 99 times, and used 100 or more times. The lifetime prevalence results in Table 1 indicate that in 1991 approximately 59% of young men reported using marijuana in their lifetime, 24% used cocaine, and 22% reported using at least one of the other drugs (heroin, hallucinogens, or inhalants).

The NHSDA did not ask similar questions for lifetime cigarette or alcohol use. For cigarettes, we used two questions that asked respondents the number of years they smoked daily[4] and the number of cigarettes they smoked per day while smoking daily. We multiplied the answers to these two questions to obtain an estimate of the total number of cigarettes smoked during their lifetime while smoking daily. Using this estimate, we created discrete categories of lifetime cigarette use that approximated quartiles of use, and we assigned smokers to one of these categories. The categories are defined as 1 to 419 packs; 420 to 1113 packs; 1114 to 2226 packs; and 2227 packs or more. For alcohol, we included two variables that proxy for lifetime use—the age an individual first drank and the age an individual first drank monthly or more often. These variables equal zero if the respondent never drank or never drank monthly, respectively. For both cigarette and alcohol use we included indicator variables equal to one if the person had no lifetime use. Again, Table 1 provides definitions and descriptive statistics for these variables. The prevalence results for 1991 indicate that 42% of young men smoked cigarettes daily and 22% reported never smoking at all; only 7% reported never drinking alcohol.

As described below, we require instrumental variables to perform exogeneity tests of the drug use variables. The instruments we chose consisted of respondents' assessments of the risk associated with using various substances (e.g., the risk associated with using marijuana regularly) and their difficulty in obtaining certain illicit substances (e.g., cocaine, heroin, and marijuana). The NHSDA instrument asked respondents to rank the risk from 1 for no risk to 4 for great risk and the difficulty of obtaining a substance from 1 for probably impossible to 5 for very easy. Table 1 reports the mean values of our risk and difficulty of obtaining substance variables.

3. Empirical model and estimation issues

Most studies of the relationship between substance use and labor supply examine a semi-reduced form labor supply function like Eq. (1):

$$H = \beta_0 + \beta_1 D + \beta_2 X + e \qquad (1)$$

where H is a measure of hours of work, D is the consumption of drugs or alcohol, X is a set of exogenous control variables, and e is an error term containing, among other things, unobserved tastes and preferences. (See, for example, Register and Williams (Undated); Kaestner (1994); Mullahy and Sindelar (1995)). To be consistent with the existing literature, we estimate

[4] Eligibility for this question required that individuals smoked daily for at least 1 yr.

labor supply functions like Eq. (1). Before doing so, we discuss the interpretation of the coefficient, β_1.

Because Eq. (1)includes *D* as a regressor—a potentially endogenous variable— Eq. (1)represents a conditional labor supply function. Although specifications such as Eq. (1)are quite common in the literature, little theoretical justification is given for using such specifications of the conditional labor supply function. In an earlier version of this paper, we discuss how Eq. (1)can be interpreted as the marginal rate of substitution between hours of work and drug consumption. From this perspective, β_1 can be interpreted as the ratio of the income effects for leisure and drug use. Its value depends on the sign and magnitude of the income elasticities of leisure and substance use.

The labor supply literature suggests that the marginal rate of substitution between hours worked and drug consumption depends on demographic variables such as age, race, marital status, number of children living in the household, location (urban/rural), and health status. Since we do not control directly for the wage rate, we include variables that affect the wage, including education and higher-order age effects. We also control for possible impacts of previous drug consumption on both wages and the marginal rate of substitution between hours worked and drug use. Although our theoretical specification does not include such effects, we incorporate them for two reasons. First, examining whether future labor supply models need to include past drug use is a useful contribution. Second, if individuals are myopic and do not recognize the long-term impact of current drug use, then past drug use variables would be insignificant and the empirical model would coincide with the theoretical specification.

Guided by the theoretical framework described above, we carried out an extensive specification search with the 1991 NHSDA data set to identify the appropriate set of explanatory variables in Eq. (1). In this specification search, we never consciously selected models that produced a particular sign or size of the substance use coefficients. We did, however, monitor explicitly the impact of different specifications on the substance use coefficients. As a rule of thumb, we used significance tests as a basis for including particular linear and interaction effects (e.g., state effects and race/age interactions). When including additional regressors appreciably affected the substance use coefficients, we opted to leave them in the model. We also experimented with various ways to capture the relationship between substance use and labor supply—from simple use/nonuse indicator variables to continuous measures of current use. The categorical specification for drug use described earlier appears to provide an adequate description of the relationship.

Despite these specification searches, the final empirical specification is quite standard. Besides the categorical substance use measures, we included race/ethnicity, age, education, number of children, marital status, health status, and indicator variables for Metropolitan Statistical Area (MSA) size and states (to capture both labor market conditions and variations in the prices of legal and illegal substances). This relatively standard labor supply specification makes the significant and conflicting estimates for light marijuana use across adjacent years an important finding.

Because unobserved tastes for leisure may influence both hours worked and drug use, a potential correlation exists between past month substance use and the error term. To evaluate this possibility, we performed tests for the exogeneity of all past month substance use variables. To conduct this test, we required a set of exogenous variables (instruments) that are excluded from the hours regression but explain past month substance use. Our instruments included respondents' assessments of the risk associated with using the various substances and their assessment of the difficulty in obtaining certain illicit substances.[5] Following the suggestion of Bound et al. (1993), we examined the significance of these instruments in the first-stage regressions and found that they were generally significant.[6] Although lifetime use variables are less likely than past month substance use to be correlated with the error term in the past month hours of work equation, lifetime use variables may also be endogenous and should be tested for exogeneity. The data set does not contain enough information to test the 20 lifetime use variables for exogeneity, so we treated them as exogenous variables.

We tested for the exogeneity of each past month substance abuse using a Hausman test (Davidson and MacKinnon, 1993, p. 237). The exogeneity test consists of testing the joint significance of the residuals from all the substance use equations. We also tested the overidentifying restrictions of the model (i.e., that the excluded instruments are orthogonal to the error in the hours regression) (Davidson and MacKinnon, 1993, p. 235).

The statistical models used in this paper require that we address pre-estimation issues, clustered data (i.e., nonindependent observations within groups), and potentially nonnormal and heteroscedastic disturbances when constructing standard errors and test statistics. Consequently, in this paper we report standard errors and test statistics that are based on bootstrap estimates.[7] We used STATA's bootstrapping routine to draw 2000 random samples. Because our data are clustered, we sampled clusters and not observations. For each sample, we drew clusters with replacement until we had a sample with the same number of clusters as our original dataset. Thus, because the clusters vary in size, our samples also vary somewhat in size. We then estimated our models on each sample (STATA, 1994) to generate an empirical distribution of estimated parameters.

---

[5] Because we include state indicator variables in the first-stage and second-stage regressions, we cannot include state-level drug and alcohol prices as instruments for drug and alcohol use. Importantly, unlike state-level drug and alcohol prices, state indicator variables control for all cross-state variation in the variables of interest.

[6] The risk variables were jointly significant at the 0.10 level or better in 7 out of the 11 first-stage, past month, substance use regressions. We also implemented a procedure suggested by Shea (1996). The Shea procedure provides insights, although no definitive test, into whether our system is identified in the case with multiple endogenous variables. Our first-stage R^2 s are smaller when we account for multiple endogenous variables, but Shea does not provide a cut-off value for R^2 s below which identification is rejected. Based on the generally good results of our *F*-tests of the excluded risk variables, we proceeded with exogeneity testing and the testing of the overidentification restrictions of the model.

[7] In some preliminary analyses, we used Taylor series approximations for constructing test statistics (e.g., Whiter/Huber standard errors). Several studies have found that first-order Taylor series expansions for computing test statistics in complicated models can often yield tests with incorrect sizes (see, for example, Horowitz, 1994). In several instances, these test statistics yielded results somewhat . different from those obtained by bootstrap procedures. Given the known possibility of poor performance by the first-order asymptotic expansions, we interpret the differences as an indication that our sample sizes may not be large enough for the series approximations to yield reliable test statistics.

Table 2
Tests of exogeneity and overidentification for past month substance use: 1991 data

Hours equation	Identifying restrictions	χ^2 for test of exogeneity	χ^2 for test of over identification
1. Eq. (1)	Risk variables and lifetime substance use variables	4.14 ($df = 11$) $p = 0.97$	37.61 ($df = 24$) $p = 0.04$
2. Eq. (1) plus lifetime substance use variables	Risk variables	3.64 ($df = 11$) $p = 0.98$	4.86 ($df = 4$) $p = 0.30$

Note: All results based on bootstrapped instrumental variable estimates (2000 replications) that account for both pre-estimation and clustering.

Table 2 shows the results of both our endogeneity and overidentification testing using bootstrapping on the 1991 data (complete results from all tests and specifications are available upon request). We tested two specifications to assess the results with alternative identifying assumptions. The first row of Table 2 shows the results of using the risk variables and the lifetime substance use variables as the identifying restrictions (i.e., we exclude the risk and lifetime use variables from Eq. (1)). Using this specification, we cannot reject the null hypothesis that past month substance use is exogenous; however, we do reject (at the 0.04 level) that our overidentifying restrictions hold. Row 2 shows the results of adding the lifetime use variables to Eq. (1)and using only the risk variables as the identifying variables. Again, we cannot reject the null hypothesis of exogeneity, but now we do not reject the overidentifying restrictions. This result suggests that lifetime use variables should be included in the hours equation.

In summary, our analysis suggests that past month substance use variables can be treated as exogenous; that lifetime use variables should be included in Eq. (1); and that OLS is appropriate for estimating Eq. (1).

4. Results

Table 3 presents the substance use coefficients for our weighted OLS regressions of Eq. (1)using 1991 data with bootstrapped standard errors. Because of our particular interest in the relationship between illicit drug use and hours worked and the relatively high prevalence of past month marijuana use, we began with a parsimonious specification that includes only one substance use variable—an indicator variable that is equal to one if the individual used marijuana in the past month and is zero otherwise (model 1). In model 2, we replace the simple use/no-use marijuana indicator with four discrete marijuana use coefficients. We include the entire set of past month substance use variables in model 3 and add lifetime use variables in model 4. The full set of demographic variables (reported in Table A-1 in Appendix A) and the state indicator variables are included in all specifications.

Table 3
Substance use coefficients: 1991 data

Variable	Model 1 (n = 893)	Model 2 (n = 893)	Model 3 (n = 893)	Model 4 (n = 893)
Any past month marijuana use	1.815 (8.685)	–	–	–
1 to 3 joints	–	32.607[b] (15.650)	33.878[a] (20.321)	41.702[b] (20.145)
4 to 15 joints	–	−22.548[a] (13.191)	−18.867 (12.133)	−22.297[a] (13.431)
16 to 59 joints	–	−8.621 (14.468)	−0.431 (14.437)	1.674 (19.028)
60 or more joints	–	6.165 (10.147)	10.483 (12.172)	9.443 (15.486)
Any past month cocaine use	–	–	5.694 (14.209)	24.565 (16.817)
Any past month heroin, hallucinogens, or inhalants use	–	–	−28.453[a] (16.189)	−33.255[a] (17.966)
Any past month alcohol use				
1 to 7 alcoholic drinks	–	–	0.503 (8.847)	3.105 (9.061)
8 to 23 alcoholic drinks	–	–	−7.503 (8.270)	−3.919 (10.192)
24 to 59 alcoholic drinks	–	–	5.247 (8.749)	10.954 (10.726)
60 or more alcoholic drinks	–	–	5.270 (9.904)	8.994 (11.922)
Any past month cigarette use	–	–	−1.045 (6.414)	4.875 (9.659)
Adjusted R^2	0.129	0.140	0.143	0.181
χ^2_4 for joint significance of past month marijuana use variables	–	9.019[a] (p = 0.061)	7.270 (p = 0.122)	9.553[b] (p = 0.049)
χ^2_4 for joint significance of past month alcohol use variables	–	–	2.088 (p = 0.720)	3.012 (p = 0.556)
χ^2_4 for joint significance of lifetime marijuana use variables	–	–	–	4.385 (p = 0.356)
χ^2_4 for joint significance of lifetime cocaine use variables	–	–	–	8.203 (p = 0.084)
χ^2_4 for joint significance of lifetime other drug use variables	–	–	–	1.603 (p = 0.808)
χ^2_4 for joint significance of lifetime cigarette use variables	–	–	–	3.316 (p = 0.651)
χ^2_4 for joint significance of lifetime alcohol use variables	–	–	–	1.382 (p = 0.710)

Note: Standard errors in parentheses for coefficient estimates; p values in parentheses for χ^2s. All results based on bootstrapped OLS estimates (2000 replications) that account for both pre-estimation and clustering. All models include the demographic variables listed in Table 1. See Table A-1 in Appendix A for all estimated coefficients (excluding the state indicator variables) with their bootstrapped standard errors.
[a]Significant at the 0.10 level.
[b]Significant at the 0.05 level.
[c]Significant at the 0.01 level.

Looking first at model 1, the marijuana indicator variable suggests that marijuana users work 2 more hours per month than nonmarijuana users, but the coefficient is not significant. However, when we replace the use/no-use indicator with four categorical marijuana use variables (based on quantities of use), the model 2 results show that light marijuana use is positively and significantly related to hours worked. In particular, young men who smoked between 1 and 3 joints in the past month worked 33 more hours in the past month than young men who smoked no marijuana.[8] Young men who smoked 4 to 15 joints in the past month, however, worked 23 h less than young men who smoked no marijuana; an effect that is only significant at the 0.10 level. The other two marijuana coefficients are insignificant and alternate in sign, suggesting that higher levels of marijuana use are unrelated to labor supply. The χ_4^2 test on the joint significance of the marijuana use variables at the bottom of model 2 shows that, as a group, the marijuana use variables are significant at the 0.06 level.[9]

The remaining past month substance use variables are included in model 3. This specification allows us to control for the potential confounding (i.e., co-morbid) effects of other substances on the estimated marijuana coefficients. For example, individuals may be using both marijuana and alcohol (or other drugs). If we did not control for these other substances, the estimated marijuana coefficients may capture some of the effect of these other substances. However, the estimated marijuana use coefficients in model 3 are approximately the same magnitude and significance as in model 2. The coefficient on smoking 16 to 59 joints in the past month does change from −8.621 in model 2 to −0.431 in model 3, but both estimates are still within 1 standard error of each other. The model 3 results suggest that the more parsimonious model (model 2) estimates are not confounded by a failure to control for other substances.

Looking at substances other than marijuana, we see that indicators for past month cocaine use and cigarette smoking are insignificant. The coefficient on other drug use is significant at the 0.10 level and suggests that young men who used heroin, hallucinogens, or inhalants in the past month worked 28 h less than those who did not. All but one of the estimated coefficients on alcohol consumption are positive, but none are significant. The χ_4^2 test on the joint significance of the alcohol variables reported at the bottom of model 3 indicates that as a group the alcohol variables are insignificant. Overall, the model 3 results indicate that most types of substance use have little effect on the number of hours worked in the last month.

Model 4 adds the lifetime use variables to the model. The coefficient on smoking 1 to 3 marijuana joints is slightly larger than in models 2 and 3 and is significant at the 0.05 level. As in model 1, the coefficient on smoking 4 to 15 joints in the past month is significant at the 0.10 level and indicates that young men who smoked between 4 and 15 joints in the past month worked 22 h less than those who did not smoke marijuana. Theχ_4^2 test on the joint significance of the four marijuana use coefficients shows that they are jointly significant at the 0.05 level. The coefficients on cigarette smoking, alcohol use, and cocaine use remain insignificant, and the coefficient on other drug use in the past month is approximately the same as in model 3.

Only two of the lifetime use variables are individually significant (see Table A-1 in Appendix A). The coefficient on using cocaine 1 or 2 times is significant at the 0.10 level and indicates that young men who used cocaine 1 or 2 times in their lives worked 20 h more in the past month than

---

[8] We also re-estimated this model using a Heckman correction. We found that the inverse mill's ratio was insignificant and the point estimates on 1 to 3 joints of marijuana use increased slightly.

[9] We also tested a more parsimonious version of this specification in which we first replaced the past month marijuana use intervals with the midpoints of each interval, and then tested whether the midpoints lie on the same line. We decisively rejected the linearity assumption.

those who did not use cocaine. The coefficient on using cocaine 3 to 10 times is significant at the 0.05 level and indicates that young men who used cocaine at that level in their lives worked 33 h less in the past month than those who did not use cocaine.

As mentioned earlier, we used the 1992 NHSDA data to examine the robustness of our 1991 results. Table A-2 in Appendix Apresents our 1992 results. Comparing the specifications that include the lifetime variables (model 4), we see a few noteworthy differences between the substance use estimates from adjacent years. First, smoking between 1 and 3 marijuana joints is significant at the 0.05 level and is associated with approximately 41 *fewer* hours of work in 1992—the exact opposite of the results we found in 1991. Second, the past month marijuana use variables are jointly significant at the 0.05 level in 1991 but are not significant even at the 0.10 level in 1992. Third, having used other drugs in the past month is associated with an insignificant decrease of 3 h in 1992, but a significant decrease (p<0.10) of 33 h worked in the past month in 1991.

Because the NHSDA surveys are nationally representative of the U.S. household population, we are surprised that the results from adjacent years, especially the signs of the light-use marijuana variables, are not more alike. The differences between the estimates for the 2 yr suggest several different interpretations. First, changes in the relationship between hours worked and drug use may have occurred from 1991 to 1992. Although this change is possible, it seems fairly unlikely over such a short time interval. However, with approximately 30 individuals in the 1991 and 1992 NHSDA samples who smoke between 1 and 3 joints, we may have sampled individuals by chance whose labor supply–drug use relationship is quite different in adjacent years. Second, our model may not be rich enough to capture important features of the relationship between hours of work and substance use. However, our specification included a large number of covariates, such as age–race interactions and state indicators to account for these factors. After extensive evaluation, we find no compelling evidence of misspecification.

A third interpretation is that the specification searches we undertook with the 1991 data used the inherent randomness in the data set to help find a `reasonable' specification. The resulting specification for the 1991 data might be due more to our specification searches, in conjunction with both the inherent randomness of the sample and the small number of substance users, than with the true underlying behavioral relationship. Although we took particular care in our specification search not to judge the reasonableness of our labor supply models based on the drug use coefficients, the differences between 1991 and 1992 suggest that pretesting model specifications may have yielded biased estimates for 1991. This lesson is important because it demonstrates how easily pretesting may lead analysts to uncover potentially misleading results. Furthermore, without valid standard errors for the 1991 estimates that correct for pretesting, assessing whether significant differences exist in the behavioral responses across years is difficult.

Despite the irregularities between adjacent years in a few of the key coefficient estimates, the results still show that substance use is generally not significantly related to labor supply for young men. For example, less than 20% of the substance use variables in model 4 are significant at the 0.10 level in either 1991 or 1992. Nevertheless, our results do suggest that pretest bias may be present in our estimates from 1991, and we should probably have less confidence in these results. Compared to the 1991 findings, the 1992 estimates are more reliable due to our cross-validation exercise, and we can interpret these coefficients in a classical fashion.

5. Summary and conclusion

The purpose of this paper is to examine the relationship between young men's drug use and hours worked. Using data from the 1991 NHSDA, we first conducted a specification search that included exogeneity testing of past month substance use variables (marijuana, alcohol, and other drug use). Overall, our results from the 1991 data indicate that drug use has little relationship with the number of hours worked by young men in the past month. In our most comprehensive specification, we found no significant relationship between past month labor supply and past month cigarette, alcohol, and cocaine use, but our results suggest that young men who smoke between 1 and 3 marijuana joints in the past month worked approximately 42 h more in the past month than nonusers (504 h yr^{-1}).

To assess the robustness of our 1991 results, we re-estimated our models using data from the 1992 NHSDA. It is important to emphasize that our labor supply models were estimated just once with 1992 data. This estimation strategy effectively eliminates problems of pretest biases that may be inherent in analyses that use the same data set for both the specification search and estimation. As a consequence, we feel that our approach and estimates provide accurate and classically interpretable assessments of the labor market consequences of substance use.

Comparing the 1991 and 1992 results, the 1992 data also show that substance use has little relationship overall to the number of hours worked. However, in contrast to the 1991 results, the 1992 results show that smoking between 1 and 3 marijuana joints in the last month is associated with 41 *fewer* hours worked. Perhaps surprisingly, our results fail to demonstrate a significant decrease in hours worked for heavy marijuana or heavy alcohol use.

Although finding the same overall results in adjacent years would be reassuring, a disturbing feature of our findings is that the estimated relationship between light marijuana use and hours worked is so dramatically different between adjacent years. Theory indicates that even a misspecified model would yield similar results across adjacent years if the relationship were stable (White, 1982). The apparent lack of robustness of the light marijuana use result poses a problem to researchers and policy-makers. Based on the 1991 data we would conclude that light marijuana use is positively related to hours worked—a result that accords with the view that drug use increases productivity. However, based on the 1992 results, one might conclude that light marijuana use is associated with *decreased* work effort. The conflicting results leave us in a quandary as to the relationship between light marijuana use and labor supply.

We are fortunate to have adjacent cross sections of the same survey. For the most part, researchers have access to only one cross section; thus, they are not able to re-estimate their model on data that have been collected using the same methodology in adjacent years as we are able to do here. Our paper demonstrates the value of such re-estimation in the context of substance use and labor supply and asks whether one should believe results based on a single cross section. Furthermore, if the drug use–hours relationship is indeed not robust, then the limited panel data that are available in the NLSY and other panel data sets will not necessarily yield more stable or believable estimates.

Based on our results and the mixed results of other researchers on this topic, there is little compelling evidence of a robust labor supply–drug use relationship. However, because individual substance use parameters may occasionally be `significant', consumers and producers of these empirical models should exercise a great deal of caution before concluding that the estimated effects are stable or important. Additional analysis of the NHSDA and other data sets is needed to understand whether the relationship between substance use and other labor market outcomes such as wages and the propensity to work are similarly unstable.

Acknowledgements

This research was supported by a grant from the National Institute on Drug Abuse, DA07821-02. We thank Frank Chaloupka for providing us with state alcohol tax and substance use regulations data and Robert Kaestner, Naci Mocan, and an anonymous referee for comments on an earlier draft.

Appendix A

Table A-1. All OLS coefficients: 1991 data.

Variable	Model 1 ($n = 893$)	Model 2 ($n = 893$)	Model 3 ($n = 893$)	Model 4 ($n = 893$)
Intercept	−646.185	−632.924	−600.100	−673.492
	(564.805)	(540.794)	(547.539)	(539.542)
Demographics				
Black	2036.878[b]	1986.867[a]	1968.394[a]	2014.747[a]
	(1035.048)	(1012.206)	(1030.964)	(1072.519)
Hispanic	694.418	556.965	478.195	508.003
	(924.169)	(906.457)	(935.928)	(928.904)
Native American, Asian, or other race	2687.006	2394.703	3460.563	2598.640
	(4905.427)	(5033.284)	(5361.859)	(4825.272)
Age	68.302	67.301	64.379	71.727
	(52.893)	(50.589)	(51.613)	(50.966)
Age^2	−1.496	−1.477	−1.418	−1.583
	(1.227)	(1.173)	(1.198)	(1.183)
Black × Age	−197.994[b]	−193.625[b]	−191.524[b]	−195.273[a]
	(97.666)	(95.538)	(97.502)	(101.831)
Black × Age^2	4.725[b]	4.634[b]	4.576[b]	4.642[a]
	(2.287)	(2.239)	(2.288)	(2.399)
Hispanic × Age	−65.814	−53.141	−45.982	−47.812
	(87.337)	(85.668)	(88.565)	(87.895)
Hispanic × Age^2	1.500	1.209	1.044	1.063
	(2.046)	(2.007)	(2.077)	(2.062)
(Native American, Asian, or other race) × Age	−265.248	−238.162	−334.941	−256.302
	(443.708)	(455.191)	(485.901)	(434.165)
(Native American, Asian, or other race) × Age^2	6.352	5.732	7.921	6.128
	(10.009)	(10.267)	(10.980)	(9.747)
Fewer than 12 yr of education	11.201	11.533	9.426	7.672
	(12.601)	(12.753)	(14.029)	(14.641)
Exactly 12 yr	18.618	17.980	16.727	15.828

Table A-1 (continued)

Variable	Model 1 ($n = 893$)	Model 2 ($n = 893$)	Model 3 ($n = 893$)	Model 4 ($n = 893$)
Demographics				
of education	(11.917)	(11.993)	(12.870)	(13.461)
13 to 15 yr	26.451[b]	25.500[b]	24.629[b]	28.580[b]
of education	(11.308)	(11.558)	(11.690)	(12.236)
MSA with a population of more than 1 million	−11.040	−11.121	−6.504	−4.801
	(11.452)	(11.687)	(12.514)	(14.433)
MSA with a population of 250,000 to 1 million	−2.214	−0.486	3.401	1.044
	(13.214)	(13.319)	(14.090)	(15.555)
MSA with a population of fewer than 250,000	2.944	4.726	5.964	3.282
	(15.570)	(16.450)	(16.418)	(17.687)
Non-MSA urban area	12.208	11.907	16.611	21.974
	(11.786)	(11.904)	(13.080)	(14.856)
Family nonlabor income	0.020	0.014	0.009	0.015
	(0.028)	(0.029)	(0.029)	(0.030)
Married	9.738	10.706	11.004	13.273
	(8.880)	(8.918)	(8.974)	(9.968)
Number of biological or adopted children living in the household	7.396	6.275	6.860	6.210
	(10.320)	(10.429)	(10.648)	(11.417)
Fair to poor health in the past year	−5.699	−3.564	−0.687	−0.709
	(14.886)	(14.119)	(13.599)	(12.501)
Any past month marijuana use	1.815	–	–	–
	(8.685)			
1 to 3 joints	–	32.607[b]	33.878[a]	41.702[b]
		(15.650)	(20.321)	(20.145)
4 to 15 joints	–	−22.548[a]	−18.867	−22.297[a]
		(13.191)	(12.133)	(13.431)
16 to 59 joints	–	−8.621	−0.431	1.674
		(14.468)	(14.437)	(19.028)
60 or more joints	–	6.165	10.483	9.443
		(10.147)	(12.172)	(15.486)
Any past month cocaine use	–	–	5.694	24.565
			(14.209)	(16.817)
Any past month heroin, hallucinogens, or inhalants use	–	–	−28.453[a]	−33.255[a]
			(16.189)	(17.966)
Any past month alcohol use				
1 to 7 alcoholic drinks	–	–	0.503	3.105
			(8.847)	(9.061)
8 to 23 alcoholic	–	–	−7.503	−3.919

Table A-1 (continued)

Variable	Model 1 ($n = 893$)	Model 2 ($n = 893$)	Model 3 ($n = 893$)	Model 4 ($n = 893$)
Any past month alcohol use				
drinks			(8.270)	(10.192)
24 to 59 alcoholic drinks	–	–	5.247 (8.749)	10.954 (10.726)
60 or more alcoholic drinks	–	–	5.270 (9.904)	8.994 (11.922)
Any past month cigarette use	–	–	−1.045 (6.414)	4.875 (9.659)
Any lifetime marijuana use				
1 or 2 times	–	–	–	3.324 (9.807)
3 to 10 times	–	–	–	8.839 (10.353)
11 to 99 times	–	–	–	−9.896 (9.898)
100 or more times	–	–	–	15.576 (16.313)
Any lifetime cocaine use				
1 or 2 times	–	–	–	20.379[a] (11.980)
3 to 10 times	–	–	–	−32.608[b] (15.131)
11 to 99 times	–	–	–	−27.781 (18.162)
100 or more times	–	–	–	−30.110 (24.050)
Any lifetime heroin, hallucinogens, or inhalants use				
1 or 2 times	–	–	–	−10.976 (12.583)
3 to 10 times	–	–	–	−9.320 (14.199)
11 to 99 times	–	–	–	1.831 (18.157)
100 or more times	–	–	–	−17.751 (26.629)
Any daily cigarette use				
1 to 419 total packs smoked while smoking daily	–	–	–	−7.033 (14.350)
420 to 1113 total	–	–	–	5.795

Table A-1 (continued)

Variable	Model 1 ($n = 893$)	Model 2 ($n = 893$)	Model 3 ($n = 893$)	Model 4 ($n = 893$)
Any daily cigarette use				
packs smoked while smoking daily				(12.417)
1114 to 2226 total packs smoked while smoking daily	–	–	–	−1.127 (15.532)
2227 or more total packs smoked while smoking daily	–	–	–	3.560 (14.561)
Never smoked cigarettes	–	–	–	12.278 (8.529)
Lifetime alcohol use				
Age tried alcohol	–	–	–	−1.204 (1.048)
Age first drank alcohol monthly	–	–	–	−0.058 (0.424)
Never drank alcohol	–	–	–	−24.376 (22.246)
Adjusted R^2	0.129	0.140	0.143	0.181
χ^2_4 for joint significance of past month marijuana use variables	–	9.019[a] ($p = 0.061$)	7.270 ($p = 0.122$)	9.553[b] ($p = 0.049$)
χ^2_4 for joint significance of past month alcohol use variables	–	–	2.088 ($p = 0.720$)	3.012 ($p = 0.556$)
χ^2_4 for joint significance of lifetime marijuana use variables	–	–	–	4.385 ($p = 0.356$)
χ^2_4 for joint significance of lifetime cocaine use variables	–	–	–	8.203 ($p = 0.084$)
χ^2_4 for joint significance of lifetime other drug use variables	–	–	–	1.603 ($p = 0.808$)
χ^2_4 for joint significance of lifetime cigarette use variables	–	–	–	3.316 ($p = 0.651$)
χ^2_4 for joint significance of lifetime alcohol use variables	–	–	–	1.382 ($p = 0.710$)

Note: Standard errors in parentheses for coefficient estimates, p values in parentheses for χ^2s. All results based on bootstrapped OLS estimates (2000 replications) that account for both pre-estimation and clustering.

[a]Significant at the 0.10 level.

[b]Significant at the 0.05 level.

[c]Significant at the 0.01 level.

Table A-2. All OLS coefficients: 1992 data.

Variable	Model 1 ($n=1019$)	Model 2 ($n=1019$)	Model 3 ($n=1019$)	Model 4 ($n=1019$)
Intercept	−583.520 (634.773)	−607.351 (653.248)	−498.217 (642.064)	−557.847 (618.576)
Demographics				
Black	−657.968 (1123.592)	−658.278 (1143.641)	−800.700 (1165.537)	−710.419 (1223.043)
Hispanic	54.742 (900.554)	92.086 (910.901)	77.550 (920.108)	166.470 (917.263)
Native American, Asian, or other race	900.932 (2506.838)	1092.511 (2425.044)	1017.580 (2537.618)	1957.785 (2559.431)
Age	67.557 (58.573)	70.509 (60.308)	59.486 (59.398)	64.568 (56.749)
Age^2	−1.488 (1.349)	−1.565 (1.389)	−1.313 (1.370)	−1.427 (1.313)
Black × Age	60.136 (104.648)	60.827 (106.491)	75.029 (108.428)	67.993 (113.712)
Black × Age^2	−1.400 (2.420)	−1.431 (2.463)	−1.776 (2.506)	−1.647 (2.628)
Hispanic × Age	−2.601 (84.287)	−6.405 (85.187)	−5.219 (86.008)	−12.940 (85.966)
Hispanic × Age^2	−0.010 (1.960)	0.085 (1.979)	0.061 (1.997)	0.228 (2.000)
(Native American, Asian, or other race) × Age	−95.266 (238.353)	−114.047 (230.389)	−106.629 (241.141)	−195.412 (243.025)
(Native American, Asian, or other race) × Age^2	2.412 (5.606)	2.866 (5.416)	2.691 (5.671)	4.772 (5.710)
Fewer than 12 yr of education	−1.011 (13.049)	−0.780 (12.898)	2.734 (12.952)	6.122 (13.091)
Exactly 12 yr of education	1.395 (10.946)	0.420 (11.122)	3.608 (10.947)	3.822 (10.434)
13 to 15 yr of education	18.766[a] (11.056)	17.070 (10.958)	20.547[a] (10.820)	18.631[a] (10.975)
MSA with a population of more than 1 million	−15.381 (11.778)	−16.159 (11.727)	−18.026 (12.251)	−12.592 (12.316)
MSA with a population of 250,000 to 1 million	−20.207 (14.412)	−19.936 (14.490)	−23.027 (14.684)	−18.269 (14.500)
MSA with a population of fewer than 250,000	−23.826 (19.784)	−23.887 (18.980)	−24.190 (19.539)	−22.090 (18.964)
Non-MSA urban area	−5.852 (15.292)	−7.569 (15.871)	−9.844 (16.158)	−4.576 (16.382)
Family nonlabor	0.012	0.017	0.016	0.008

Table A-2 (continued)

Variable	Model 1 ($n = 1019$)	Model 2 ($n = 1019$)	Model 3 ($n = 1019$)	Model 4 ($n = 1019$)
Demographics				
income	(0.031)	(0.033)	(0.034)	(0.031)
Married	8.596	9.258	9.290	6.347
	(8.189)	(7.899)	(7.640)	(7.133)
Number of biological or adopted children living in the household	7.291	7.770	8.481[a]	10.151[b]
	(5.162)	(4.963)	(4.831)	(4.713)
Fair to poor health in the past year	−1.472	−3.281	−6.268	−10.741
	(8.283)	(8.018)	(8.870)	(8.765)
Any past month marijuana use	−11.353	–	–	–
	(7.993)			
1 to 3 joints	–	−40.007[b]	−43.998[c]	−40.692[b]
		(16.562)	(16.830)	(17.374)
4 to 15 joints	–	−9.639	−14.833	−12.750
		(14.235)	(13.344)	(15.600)
16 to 59 joints	–	11.152	6.226	10.366
		(11.617)	(12.671)	(14.048)
60 or more joints	–	−15.080	−22.006	−9.010
		(17.541)	(20.825)	(22.631)
Any past month cocaine use	–	–	−5.350	−15.095
			(17.481)	(20.083)
Any past month heroin, hallucinogens, or inhalants use	–	–	8.422	−3.011
			(17.317)	(17.766)
Any past month alcohol use				
1 to 7 alcoholic drinks	–	–	9.522	4.449
			(9.183)	(10.245)
8 to 23 alcoholic drinks	–	–	15.781[b]	8.606
			(7.012)	(9.124)
24 to 59 alcoholic drinks	–	–	11.528	4.741
			(8.241)	(11.619)
60 or more alcoholic drinks	–	–	18.808[b]	17.038
			(9.089)	(11.250)
Any past month cigarette use	–	–	−2.737	−0.887
			(5.678)	(8.365)
Any lifetime marijuana use				
1 or 2 times	–	–	–	1.819
				(7.318)
3 to 10 times	–	–	–	−12.148[a]
				(7.099)

Table A-2 (continued)

Variable	Model 1 ($n = 1019$)	Model 2 ($n = 1019$)	Model 3 ($n = 1019$)	Model 4 ($n = 1019$)
Any lifetime marijuana use				
11 to 99 times	–	–	–	4.734 (10.489)
100 or more times	–	–	–	−17.670 (13.464)
Any lifetime cocaine use				
1 or 2 times	–	–	–	8.805 (8.323)
3 to 10 times	–	–	–	−14.331 (16.224)
11 to 99 times	–	–	–	−1.985 (15.210)
100 or more times	–	–	–	6.806 (25.643)
Any lifetime heroin, hallucinogens, or inhalants use				
1 or 2 times	–	–	–	0.203 (8.667)
3 to 10 times	–	–	–	2.092 (11.331)
11 to 99 times	–	–	–	34.548[b] (16.489)
100 or more times	–	–	–	25.042 (31.300)
Any daily cigarette use				
1 to 419 total packs smoked while smoking daily	–	–	–	−16.276 (9.944)
420 to 1113 total packs smoked while smoking daily	–	–	–	−5.435 (12.086)
1114 to 2226 total packs smoked while smoking daily	–	–	–	4.467 (14.382)
2227 or more total packs smoked while smoking daily	–	–	–	−11.886 (12.565)
Never smoked cigarettes	–	–	–	−9.334 (8.936)

Table A-2 (continued)

Variable	Model 1 ($n = 1019$)	Model 2 ($n = 1019$)	Model 3 ($n = 1019$)	Model 4 ($n = 1019$)
Lifetime alcohol use				
Age tried alcohol	–	–	–	0.474 (1.315)
Age first drank alcohol monthly	–	–	–	0.290 (0.497)
Never drank alcohol	–	–	–	−2.724 (25.608)
Adjusted R^2	0.116	0.125	0.131	0.147
χ^2_4 for joint significance of past month marijuana use variables	–	8.235[a] ($p = 0.083$)	10.159[b] ($p = 0.038$)	7.185 ($p = 0.126$)
χ^2_4 for joint significance of past month alcohol use variables	–	–	8.293[a] ($p = 0.081$)	3.081 ($p = 0.544$)
χ^2_4 for joint significance of lifetime marijuana use variables	–	–	–	7.343 ($p = 0.119$)
χ^2_4 for joint significance of lifetime cocaine use variables	–	–	–	2.466 ($p = 0.651$)
χ^2_4 for joint significance of lifetime other drug use variables	–	–	–	5.472 ($p = 0.242$)
χ^2_4 for joint significance of lifetime cigarette use variables	–	–	–	5.142 ($p = 0.399$)
χ^2_4 for joint significance of lifetime alcohol use variables	–	–	–	2.058 ($p = 0.560$)

Note: Standard errors in parentheses for coefficient estimates, p values in parentheses for χ^2s. All results based on bootstrapped OLS estimates (2000 replications) that account for both pre-estimation and clustering.
[a]Significant at the 0.10 level.
[b]Significant at the 0.05 level.
[c]Significant at the 0.01 level.

References

Bound, J., Jaeger, D.A., Baker, R., 1993. The cure can be worse than the disease: A cautionary tale regarding instrumental variables, Technical Paper No. 137, National Bureau of Economic Research, Cambridge, MA.

Buchmueller, T.C., Zuvekas, S.H., 1994. Drug use, drug abuse, and labor market outcomes, Unpublished manuscript, University of California-Irvine, Graduate School of Management.

Davidson, R., MacKinnon, J.G., 1993. Estimation and Inference in Econometrics, Oxford University Press, New York.

DeSimone, J., 1996, Illegal drug use and labor supply, Paper presented at the Microeconomics Workshop on Labor and Population, December 13.

French, M.T., Zarkin, G.A., Dunlap, L.J., 1996. The relationship between illicit drug use, absenteeism, and earnings at six worksites in the U.S., unpublished manuscript.
Gill, A.M., Michaels, R.J., 1992. Does drug use lower wages?. Industrial Labor Rel. Rev. 45, 419–434.
Horowitz, J.L., 1994. Bootstrap-based critical values for the information matrix test. J. Econometrics 61, 395–411.
Hoyt, G.M., Chaloupka, F.J., 1993. Self-reported substance use and survey conditions: An examination of the national longitudinal survey of youth, Paper presented at the Western Economic Association annual meeting, Lake Tahoe, NV, June 23.
Kaestner, R., 1991. The effect of illicit drug use on the wages of young adults. J. Labor Econ. 9, 381–412.
Kaestner, R., 1993. The effect of illicit drug use on the labor supply of young adults. J. Hum. Resources 29, 126–155.
Kaestner, R., 1994. New estimates of the effect of marijuana and cocaine use on wages. Industrial Labor Rel. Rev. 47, 454–470.
Kandel, D.B., Davies, M., 1990. Labor force experiences of a national sample of young adult men. Youth Soc. 21, 411–445.
Mroz, T.A., Guilkey, D., 1992. Discrete factor approximations for use in simultaneous equation models with both continuous and discrete endogenous variables, University of North Carolina, Chapel Hill, mimeo.
Mullahy, J., Sindelar, J.L., 1995. Employment, unemployment, and problem drinking, Working Paper Number 5123, NBER Working Paper Series, Cambridge, MA, National Bureau of Economic Research.
Register, C.A., Williams, D.R., 1992. Labor market effects of marijuana and cocaine use among young males. Industrial Labor Rel. Rev. 45, 435–448.
Register, C.A., Williams, D.R., undated. Marijuana use and the hours of work of young males, unpublished manuscript, University of Baltimore.
Rice, D.P., Kelman, S., Miller, L.S., 1991. Economic costs of drug abuse, in: Economic Costs, Cost-effectiveness, Financing, and Community-based Drug Treatment, Research Monograph Series 113, Rockville, MD, National Institute on Drug Abuse, pp. 10–32.
Shea, J., 1996. Instrument relevance in multivariate linear models: a simple measure, Technical Working Paper Number 193, NBER Working Paper Series, Cambridge, MA, National Bureau of Economic Research.
STATA, 1994, Stata statistical software: Release 4.0, Stata Corporation, College Station, TX.
Substance Abuse and Mental Health Services Administration SAMHSA , 1992, National household Ž . survey on drug abuse: Public release codebook 1991, U.S. Department of Health and Human Services, Rockville, MD.
Substance Abuse and Mental Health Services Administration SAMHSA , 1994, National household Ž . survey on drug abuse: Public release codebook 1992, U.S. Department of Health and Human Services, Rockville, MD.
Turner, C., Lessler, J., Devore, J., 1992. Effects of mode of administration and wording on reporting of drug use, in: Survey measurement of drug use: Methodological studies, National Institute on Drug Abuse, Rockville, MD.
White, H., 1982. Maximum likelihood estimation of misspecified models, Econometrica, 50.

Zarkin, G.A., Bray, J.W., French, M.T., 1995. An overview of substance use and employment: Findings from the 1991 national household survey on drug abuse, unpublished manuscript, Research Triangle Institute, Research Triangle Park, NC.

Zarkin, G.A., French, M.T., Mroz, Th.A., Bray, J.W., 1998. Alcohol use and wages: new results from the national household survey on drug abuse. Journal of Health Economics 17, 53–68.